Journal of Experimental Psychology

EDITED BY

JOHN B. WATSON, New York City
HOWARD C. WARREN, Princeton University (*Review*)
JAMES R. ANGELL, New York City (*Monographs*)
SHEPHERD I. FRANZ, Govt. Hosp. for Insane (*Bulletin*) and
MADISON BENTLEY, University of Illinois (*Index*)

ADVISORY EDITORS

H. A. CARR, University of Chicago; RAYMOND DODGE, Wesleyan University; KNIGHT DUNLAP, The Johns Hopkins University; H. S. LANGFELD, Harvard University; E. L. THORNDIKE, Columbia University.

Volume V, 1922

PUBLISHED BI-MONTHLY BY
PSYCHOLOGICAL REVIEW COMPANY
PRINCETON, N. J.

PRESS OF
THE NEW ERA PRINTING COMPANY
LANCASTER, PA.

INDEX FOR 1922.

Journal of
Experimental Psychology

VOL. V., NO. 1 FEBRUARY, 1922

A MEASURE OF THE ATTENTION

BY H. C. McCOMAS
Princeton University

In the experiment to be described, continuous discrimination time is used as a measure of attention. Inasmuch, however, as attention is only one factor among several, which go into discriminations and reactions, it may seem arbitrary to call the records of such experiments 'measures of attention.' Indeed, much the same sort of method has been used to find variations in efficiency for given tasks, differences in motor impulses and the influence of distractions. Obviously, in recognizing signals, and making movements to indicate such recognition, both receptors and effectors are engaged as well as central processes, and the variations in sense organs and muscles are as inevitably represented in the records as changes in the central processes. The justification for calling continuous discrimination work a measure of attention is simply that the attention is apparently the dominating factor in the time and accuracy of the reactions. This will become evident as the procedure of the experiment is described.

The present attempt to find a 'measure' of attention was actuated by a need in psychiatry for a finer indication of fluctuations than that usually employed, and also the need in applied psychology of some tests of efficiency in tasks which require a vigorous concentration of the attention.

The apparatus used is described in detail in the JOURNAL OF EXPERIMENTAL PSYCHOLOGY, Vol. II., No. 3. Briefly, it consisted in a large cardboard screen with a small ground glass window in the center; behind which were four electric bulbs colored red, green, blue and yellow. When the colored light from one of the bulbs illuminated the glass the subject, who sat before the screen with two fingers of each hand upon the reacting keys, instantly pressed the key which extinguished that colored light. This reaction not only extinguished that color but immediately presented another, to which the subject again reacted as quickly as possible. This continued for a period of ten minutes. The order in which the different colors appeared was so mixed that it was impossible to anticipate what color would appear. The arrangement of the keys controlling the colors was changed with each experiment, so no habit could be formed between certain keys and certain colors. A record of right reactions and wrong reactions was obtained by a system of markers which recorded upon a kymograph or a paper tape, and the time was indicated in five-second periods. From these records of right and wrong reactions for five-second periods during sittings of ten minutes' duration the material of this paper is prepared. The introspection of each subject was taken.

The apparatus was designed to eliminate the effects of fixation of the eyes and their fatigue, and also the uncertainties which occur in tachistoscopic work when the field is not all exposed instantaneously. With the small movements which were easily and naturally made the effects of fatigue and of learning motor coördinations were reduced to a minimum. If fatigue of either the eyes or the hands entered, it would have been cumulative and the records at the end of a ten-minute test would be characteristically different from the beginning, but this did not appear in the records. On the other hand, the time of reaction was conspicuously lengthened by anything which diverted the attention from the task.

A necessary feature of attention measurement is, of course, the interest the task arouses. If one becomes bored and indifferent obviously the record is vitiated. In these reactions to the colored lights there is an element of amusement. The game of 'putting out the lights' as quickly as possible 'with no wrong moves' proved entertaining to children and to elderly people.

Eleven subjects took part in the work, which was conducted during the second terms of two college years. In the first year *A* and *B* were graduate students with two years' laboratory experience, *C* was an instructor and *D* a professor in psychology. *D* acted as subject in both years; his first year's work is indicated by D^1 and the second year by D^2. In the second year *E* was an instructor and *F* and *G* graduate students of one and three years' laboratory experience respectively. *H* was a graduate student who was also teaching in the high school; this was his first year in laboratory work. *I* was a girl of fifteen, *J* and *K* were boys of fifteen from the high school.

Records were made during the first year upon a kymograph with the Jaquet marker registering time in one-second periods. During the second year a Foot-Pierson telegraph recorder was substituted for the kymograph and a pendulum with electric contacts was used, in place of the Jaquet. It also registered in one-second periods. In reading the records the right and wrong reactions were first grouped into five-second periods for analysis. As it is impractical to give the actual records, two specimens for eight minutes' work are given in Chart I. These were selected from the work of subject *G* who showed a greater variation in his reactions than any other subject.

Four items are of importance in every record: the averages for the right and the wrong reactions and the variation from these averages. Chart I. illustrates these features. The average of right reactions per five-second period on the eighth day is 5.5 as compared with 3.7 for the fourth, while the average for wrong reactions has risen on the fourth day to .85 from that of .33 for the eighth day. Moreover the variations from the averages has noticeably increased in the records of the fourth day.

The relation of right reactions to wrong is very important. If a subject gives no wrong reactions and his right reactions are comparatively few it is impossible to tell whether he is reacting as quickly as he can or not, for it is obvious that his limit may be above that he is achieving. On the other hand

CHART I. Discrimination reactions, 5-sec. periods.

8th Day
Horizontal Direction = Successive Reactions.

4th Day
Vertical Direction = Number of Reactions.

a large number of false reactions show that the subject is taking chances on his reactions.

Indeed the record for *G* on the eighth day of his experimental work is one of the best he made. The average number of reactions per five-second period is very high for him and the variations from the average are not very large. His wrong reactions are few. But in his record on the fourth day his averages tell another story. He is not as quick, nor as accurate, and there are several five-second periods in which he made no reactions at all.

This last feature is so exceptional that some explanation is necessary. It seems incredible that a subject should remain before a signal for ten seconds, on two occasions in a ten-minute sitting, and register nothing but a few false reactions. Ordinarily such a record would be thrown out as indicating a lack of coöperation on the part of the subject. In the present instance the subject was thoroughly interested in the experiments, volunteered willingly to react, made many sacrifices to be present regularly and gave every assurance of endeavoring to coöperate to the best of his ability. As a matter of fact he was not aware that he made such extensive lapses as his records show, though they appeared quite frequently. In an experiment on continuous simple reactions he showed the same tendencies. His strong tendency is to attend for several minutes and then relax his effort. (In writing an article he would habitually lay down his pen and look out of the window for a few moments and then go back to his task.) According to his introspection these lapses in his record are due to his attention wandering from the task. Sometimes the apparatus would start a train of ideas, which diverted him; occasionally worry about his private affairs intruded and for several seconds his attention be away from the colored light and upon some irrelevant matter. So often did this occur when he was obviously trying to correct the fault and to give a good record that it can only be attributed to a habit of relaxing his effort, or a native trait in the fluctuation of his attention.

Other subjects gave some indication of this trait, but none to the same degree. How typical this is cannot be stated from a small group of eleven subjects. Nevertheless, when it is recalled that eight of these subjects were highly educated men of more than average intelligence it would seem at least probable that such lapses are even more typical of men of less mental discipline. Several workers in applied psychology find this trait a possible explanation for lapses in attention among chauffeurs and engineers which result in disasters otherwise inexplicable.

Chart I. shows the data when the right and wrong reactions are grouped in the five-second periods. Such a grouping is evidently too restricted; for the scattering of the data obliterates any large trends which might be evident. It was therefore found advisable to group the right and wrong reactions in thirty-second periods. Such a grouping smoothes out the evident eccentricities of the data and permits an analysis of trends. The material in the tables is given on the basis of the thirty-second period.

A survey of the tables shows that there is frequently an increase in the number of right reactions when there is an increase in the wrong. The subject is trying to obtain his greatest speed possible with accuracy and frequently outspeeds his accuracy.

Naturally some relation exists between these phenomena. It is the relation which constantly presents itself in experiments involving speed and accuracy. If there were a high correlation between the right and wrong reactions it would be possible to use the 'line of regression' as a means for determining what gain in speed is obtained by the increase in inaccuracy; but there is no such guide; this is because, in addition to the rise in the number of right reactions due to the subjects becoming less accurate, there is also a rise due to a learning curve. It is necessary, however, to compare the performances of the subjects from time to time and, in order to do this, there must be some one expression which represents the number of correct reactions.

To penalize the record of right reactions by subtracting one or two for each wrong reaction would give a score in which both the speed and the accuracy of the reactions are represented, but it would be purely arbitrary. In view of the fact that the wrong reactions are the result of too great haste; or lack of attention, so that the process of recognizing the signal and selecting the correct response is disturbed, and the question of which key is pressed is a matter of chance; we may assume that the increase of the wrong reactions is an indication that the subject is relying less upon careful discrimination and choice of reactions and more upon chance. Chance, then, becomes the prominent factor in the situation. Now, if the subject reacted as quickly as his fingers could move and quite regardless of the colors of the stimuli he would hit the right key once in four trials. In pure chance reacting he would be right with 25 per cent. of the reactions and wrong with 75 per cent. If 75 per cent. of the reactions due to chance are wrong we may assume that in each thirty-second period the wrong reactions constitute 75 per cent. of the chance reactions and that the remaining 25 per cent. of the chance reactions are scattered among the right reactions. If these chance reactions are subtracted from the right reactions we have a figure which adequately represents the subject's reactions not due to chance. Such a figure is serviceable in making a comparison of the different subjects at different times and for plotting curves upon a chart. It is readily obtained by subtracting one third of the number of wrong reactions from the total number of right reactions. By this formula the records of the tables have been entered upon the charts.

Table I. gives the average number of reactions both right and wrong for a series of 10-minute experiments with subjects *A*, *B* and *C*. The recording did not begin until each subject had become thoroughly oriented and when it appeared that the learning curve had levelled out. This, however, was hardly the case with any of the subjects. Some trace of a learning curve, or rather an 'adaptation' curve, appears in

TABLE I

AVERAGE RIGHT AND WRONG REACTIONS IN HALF-MINUTE PERIODS

Experiment	Subjects											
	A				*B*				*C*			
	Right	M. V.	Wr'g	M. V.	Right	M. V.	Wr'g	M. V.	Right	M. V.	Wr'g	M. V.
1.......	29.1	1.6	0.3	0.5	29.6	1.8	0.8	0.7	38.8	1.7	6.4	2.1
2.......	27.2	2.7	1.3	1.1	30.0	1.7	1.3	1.0	35.2	1.8	6.7	2.8
3.......	32.8	1.8	1.1	1.1	28.2	2.1	1.3	0.8	37.4	1.8	7.5	2.5
4.......	28.8	1.7	1.4	1.1	31.6	2.0	1.9	1.6	37.1	1.7	10.1	1.7
5.......	31.4	1.8	0.6	0.7	30.1	2.7	1.7	1.0	37.9	1.4	8.3	2.0
6.......	26.8	2.6	4.3	4.3	30.3	2.6	2.1	1.0	36.2	1.9	6.7	1.9
7.......	30.1	1.9	1.3	0.9	31.6	2.4	2.5	0.9	37.7	2.5	7.0	2.4
8.......	26.7	5.6	0.4	0.6	37.7	1.9	3.1	1.3	41.6	2.0	5.6	2.6
9.......	28.0	2.0	1.5	1.2	33.0	2.5	3.6	1.6	37.8	2.0	5.5	2.1
10.......	29.3	3.4	4.6	1.3	...	...	...	...	...	...	...	...
11.......	29.2	2.2	0.3	0.4	...	...	...	...	...	...	...	...
12.......	31.6	2.7	1.0	0.9	...	...	...	...	...	...	...	...
13.......	32.2	2.8	1.9	1.0	...	...	...	...	...	...	...	...

the table. The experiments usually occurred once each week, but in many cases lapses of one or more weeks were inevitable. Despite the irregularity in the averages from day to day and their tendency to rise, it is clear that there are distinct differences in the subjects. *C* is quicker, but less accurate than *A* and *B*. He does not, however, scatter his reactions as widely as they, as is shown by his smaller mean variations. *B* gradually increased his errors in his effort to better his records.

These records for continuous discrimination reactions show that the average time of reaction for each subject is longer than the usual discrimination time. Thus, the averages ran as follows: $A = .93$ sec. (m.v., .08), $B = .90$ sec. (m.v., .05) and $C = .77$ (m.v., .04). The reason for this lengthened time is obvious. The usual method of obtaining discrimination reactions permits the subject a period of preparation for each reaction before the stimulus is presented, often a signal of warning is given. This results in a preparation which brings the stimulus at the moment of maximum alertness. Moreover, the subject can rest between reactions. The discrimination time may not be taken until the subject has formed habits of reacting in the same way to the same stimulus, which, of course, is obviated in the present method. For these reasons the time of reaction for all of the subjects is not comparable to that obtained in the ordinary discrimination work.

TABLE II

Average Number of Reactions per Half Minute for Subject *D* (1)					Average Number of Reactions per Half Minute for Subject *D* (2)				
	Right	M.V.	Wrong	M.V.		Right	M.V.	Wrong	M.V.
1st Day					*1st Day*				
7.30 A.M.	33.8	3.0	4.7	1.6	8.30 A.M.	27.4	3.4	3.3	1.2
12.30 P.M.	28.6	1.6	5.8	1.6	12.30 P.M.	26.4	2.2	3.9	1.7
6.00 P.M.	27.4	1.9	4.6	2.0	———	—	—	—	—
10.15 P.M.	30.6	2.2	6.1	1.8	10.30 P.M.	30.1	1.6	3.5	1.2
2d Day					*2d Day*				
7.30 A.M.	33.4	2.2	3.4	1.0	8.30 A.M.	29.9	2.0	2.8	1.0
12.30 P.M.	31.7	3.0	5.0	1.7	12.30 P.M.	25.7	2.9	3.2	1.6
6.00 P.M.	30.7	2.6	3.9	1.5	1.30 P.M.	23.2	2.8	4.1	2.6
10.15 P.M.	34.0	1.3	3.6	1.1	10.30 P.M.	—	—	—	—
3d Day					*3d Day*				
7.30 A.M.	33.8	2.1	3.2	1.2	8.30 A.M.	32.0	1.0	2.5	1.7
12.30 P.M.	28.6	2.2	5.0	1.4	1.00 P.M.	27.5	2.4	3.5	1.2
6.00 P.M.	32.6	2.6	4.2	1.7	6.00 P.M.	28.0	2.5	1.5	1.0
10.15 P.M.	27.3	1.7	5.2	1.5	10.30 P.M.	28.3	2.2	3.1	1.2
4th Day					*4th Day*				
7.30 A.M.	32.9	2.7	2.2	0.9	8.30 A.M.	29.6	3.3	2.1	0.9
12.30 P.M.	30.0	2.0	3.8	1.2	1.00 P.M.	30.0	1.6	3.0	0.6
6.00 P.M.	31.3	2.0	3.7	1.8	6.00 P.M.	26.3	2.2	4.0	0.3
10.15 P.M.	32.7	2.4	3.4	0.9	10.30 P.M.	25.0	2.3	6.8	1.6
5th Day					*5th Day*				
7.30 A.M.	35.9	2.5	2.8	0.9	8.30 A.M.	26.6	3.0	4.0	1.2
12.30 P.M.	32.5	2.3	3.9	1.2	1.00 P.M.	27.7	2.7	4.3	1.5
6.00 P.M.	31.5	3.0	4.8	1.5	6.00 P.M.	32.0	1.2	3.5	1.4
10.15 P.M.	32.5	2.0	4.3	1.3	10.30 P.M.	27.8	2.3	4.0	1.3
6th Day					*6th Day*				
7.30 A.M.	36.6	3.1	4.2	1.3	8.30 A.M.	32.5	1.8	2.7	1.6
12.30 P.M.	34.8	2.7	4.8	1.3	12.30 P.M.	31.1	2.4	3.5	1.6
6.00 P.M.	32.4	3.6	5.8	2.0	1.30 P.M.	31.5	3.0	3.9	1.0
10.15 P.M.	35.3	3.1	4.9	1.5	10.30 P.M.	31.1	2.1	4.0	1.4

Table II. gives the records for subject *D* in the first and second years of the experiment. During the first year's work he carried a heavy schedule in teaching and laboratory work. The second year was less laborious and the experiments were made at approximately 8:30 A.M. instead of 7:30 A.M. In both years the second experiment of the day occurred about 12:30 P.M. and the third about 6 P.M., though variations of a half hour from the prescribed time were frequent. The last experiment, at 10 P.M., was made after the close of two hours in a seminar.

The averages for the number of right reactions per thirty-second periods are lower in the second year than in the first until the sixth day. Possibly the explanation lies in the fact that during the second year the subject was constantly worried by a case of illness in his family and did not give the same undivided attention to his work.

The clearest feature of the first year's reactions is the diurnal curve, which appears when the averages for the four experiments each day are made, after deducting a per cent. for the wrong reactions according to the formula.

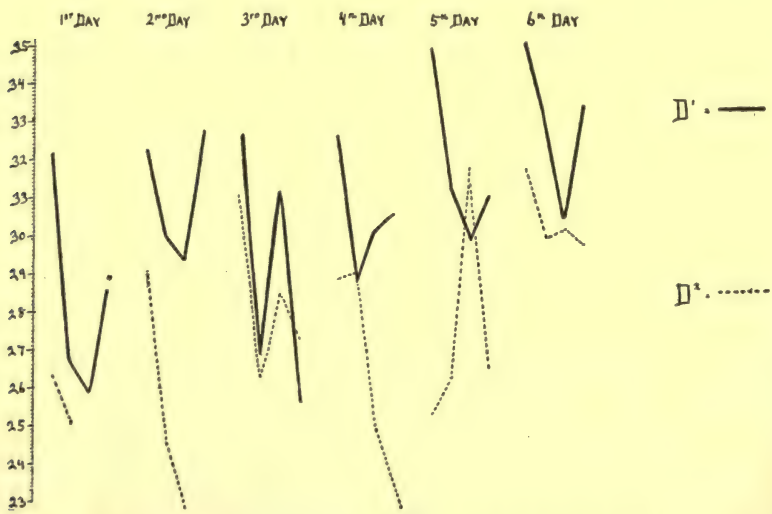


CHART II. Reactions per 30-sec. periods, taken 8 A.M., 12:30 P.M., 6 P.M., 10:30 P.M.

Chart II. shows the diurnal curve for *D*'s first year's work in the solid lines. It will be seen that in every instance, during this year's work, the early morning results were the best with one exception, and in every intance but one the results at night were superior to those in the late afternoon. These curves were so suggestive of a diurnal variation that the experiment was carried on the second year with a view to determining whether the curves were fortuitous or not. The dotted line in the chart shows the averages for the second year. The records for the first and second day are not complete, the third experiment for the first day and the fourth for the last are missing. These dotted curves show no evidence of a diurnal variation. They agree with the findings of the first year in that four out of the six days show better averages in the early morning work than during the day, and this would seem to be the only safe inference from the two sets of curves.

As a further test of diurnal possibilities three graduate students *E*, *F* and *G*, went through the experiments in the morning, midday, afternoon and night. Their results appear in Table III. Several experiments are lacking, to complete the four for each day, with subjects *E* and *F*. Nevertheless it is very evident that there is no semblance of a diurnal curve among these subjects. There is however a great difference in the efficiency shown on many days when we use the formula mentioned for comparisons. Thus, there is a drop for *E* on the third day between morning and midday of 15 per cent., of which the subject was not aware. His introspection is to the effect that his 'attention wavered' more in the morning, and that the work at midday was better. His improvement in efficiency of reaction from late afternoon to night is 27 per cent. The next day his first three records are slightly better in right reactions than any preceding day, while the night record shows a drop of about 11 per cent.; and, again, the introspection contradicts the records, for the subject thought he did his best work at night.

The greatest change subject *F* shows in any one day's records occurs in the third day when his record shows a gain of about 19 per cent. in efficiency. His introspection shows he was aware of this improvement. He was overworked, tired and sleepy all day, he felt somewhat better in the morning, when he was doing the poorer work.

Subject *G* was convinced that his best work, that which required sustained attention, was always done late at night.

Table III

Average Number of Reactions per Half Minute for Subject *E*					Average Number of Reactions per Half Minute for Subject *F*					Average Number of Reactions per Half Minute for Subject *G*				
	Right	M.V.	Wrong	M.V.		Right	M.V.	Wrong	M.V.		Right	M.V.	Wrong	M.V.
1st Day					*1st Day*					*1st Day*				
9.00 A.M....	33.2	1.5	3.2	2.1	9.00 A.M....	26.4	2.1	1.5	1.3	10.00 A.M....	25.1	3.1	1.8	1.0
12.30 P.M....	36.0	1.7	3.3	1.0	1.00 P.M.....	25.3	2.0	1.1	1.9	1.00 P.M.....	27.0	2.0	1.2	0.5
6.00 P.M....	34.0	3.4	3.4	1.2	6.00 P.M.....	26.5	2.8	2.6	1.1	6.00 P.M.....	26.0	2.0	1.3	0.9
9.30 P.M....	33.2	3.4	1.5	1.3	10.00 P.M.....	26.8	2.7	1.6	1.3	10.30 P.M.....	26.2	2.3	1.3	0.7
2d Day					*2d Day*					*2d Day*				
9.00 A.M....	36.6	2.3	4.4	1.5	9.00 A.M.....	30.2	2.2	0.5	0.6	10.00 A.M.....	27.2	4.0	0.8	0.3
12.30 P.M....	39.5	1.7	3.2	1.5	1.00 P.M.....	31.0	2.0	3.3	2.0	1.00 P.M.....	21.0	4.4	3.5	2.1
6.00 P.M....	34.2	3.3	4.5	1.6	6.00 P.M.....	32.4	2.4	2.3	1.2	6.00 P.M.....	20.8	5.1	3.7	2.0
9.30 P.M....	35.2	2.0	4.9	1.5	10.00 P.M.....	31.4	1.6	0.1	0.3	10.50 P.M.....	26.6	3.0	0.2	0.3
3d Day					*3d Day*					*3d Day*				
9.00 A.M....	36.2	1.6	4.2	1.6	9.00 A.M.....	28.2	2.1	5.2	2.9	10.00 A.M.....	26.9	3.4	0.5	0.5
12.30 P.M....	31.6	3.1	6.2	1.2	1.00 P.M.....	28.9	3.2	7.4	2.4	1.00 P.M.....	24.4	2.9	0.2	0.4
4.00 P.M....	32.5	3.2	6.6	2.1	6.00 P.M.....	29.7	2.9	3.9	1.3	6.00 P.M.....	24.7	3.6	0.7	0.6
9.00 P.M....	41.4	1.4	0.6	0.4	10.00 P.M.....	36.7	2.0	5.4	1.4	10.30 P.M.....	24.1	2.0	2.3	1.4
4th Day					*4th Day*					*4th Day*				
9.00 A.M....	43.0	2.6	4.1	1.6	9.00 A.M.....	33.8	2.7	7.1	1.9	10.00 A.M.....	26.9	3.6	3.2	1.7
12.30 P.M....	42.0	1.9	3.7	1.0	1.00 P.M.....	32.2	2.4	6.5	1.5	1.00 P.M.....	23.0	4.4	6.1	2.4
5.30 P.M....	42.6	2.0	5.9	1.9	6.00 P.M.....	29.7	3.0	4.0	2.0	6.00 P.M.....	23.1	3.4	7.0	2.4
10.00 P.M....	39.0	2.4	8.7	2.0	10.00 P.M.....	...	...	...	...	10.30 P.M.....	20.6	3.2	5.8	2.2

TABLE III *Continued*

Average Number of Reactions per Half Minute for Subject *E*				
	Right	M.V.	Wrong	M.V.
5th Day				
9.00 A.M....	48.9	3.0	10.0	5.5
....	...	...	...	...
6.00 P.M....	45.9	2.5	9.9	3.1
10.00 P.M....	46.6	2.3	9.7	1.7
6th Day				
9.00 A.M....	41.8	2.0	12.8	1.8
1.00 P.M....	42.6	2.7	12.2	3.0
6.00 P.M....	43.7	2.3	10.3	1.9
9.00 P.M....	43.4	2.1	9.4	2.6

Average Number of Reactions per Half Minute for Subject *F*				
	Right	M.V.	Wrong	M.V.
5th Day				
9.00 A.M.....	36.9	1.0	6.8	1.5
1.00 P.M.....	35.5	1.6	5.8	2.2
6.00 P.M.....	34.8	3.1	6.8	1.5
10.00 P.M.....	34.8	1.5	5.3	1.6

Average Number of Reactions per Half Minute for Subject *G*				
	Right	M.V.	Wrong	M.V.
5th Day				
10.00 A.M.....	20.8	5.4	4.7	1.4
1.00 P.M.....	24.9	2.7	2.9	1.3
6.00 P.M.....	28.0	3.5	2.0	0.4
10.30 P.M.....	28.0	5.5	2.1	1.1
6th Day				
10.00 A.M.....	24.0	5.0	1.7	1.2
1.00 P.M.....	28.7	3.0	5.0	1.2
6.00 P.M.....	23.1	4.3	4.9	1.8
10.30 P.M.....	25.3	4.2	3.9	2.0
7th Day				
10.00 A.M.....	21.1	3.3	2.2	1.7
1.00 P.M.....	26.7	3.2	2.5	1.7
6.00 P.M.....	30.2	3.1	3.1	1.3
10.30 P.M.....	28.1	4.6	1.9	1.6
8th Day				
10.00 A.M.....	29.1	4.3	4.3	1.3
1.00 P.M.....	30.0	2.4	2.9	2.0
6.00 P.M.....	33.3	2.7	2.5	2.3
10.30 P.M.....	30.7	4.1	5.9	2.1

Habits of years' standing confirmed him in this conviction. His records for attention in this type of work, however, do not show any superiority of the night reactions over those taken in the morning. Often *G* did better work when he felt fatigued, for then his mind seemed less alert and fewer distracting ideas occurred to him.

In general the introspection of these subjects was unable to detect variation in efficiency of 10 per cent., occasionally more, if by 'efficiency' is meant success in performing the task of reacting quickly and accurately. This suggests the inadvisability of relying on one's own judgment as to the character of work done under similar conditions, such as speed and accuracy in taking and sending telegraph messages, in doing linotype work, sorting type, court stenography and such tasks which are performed under pressure for quick and correct results.

Though it is evident from the four daily records of *E*, *F* and *G* that there is no diurnal curve, the results of *H* show the very clear effect of a day's work in teaching upon efficiency in discrimination reactions. See Table IV. In the seven

TABLE IV

AVERAGE NUMBER OF REACTIONS PER HALF MINUTE FOR SUBJECT *H*

	Right	M.V.	Wrong	M.V.
1st Day				
8.30 A.M.	38.0	1.4	2.5	0.7
5.00 P.M.	35.6	1.5	3.0	1.3
2d Day				
8.30 A.M.	38.0	1.7	4.3	1.6
5.00 P.M.	37.7	1.7	3.3	1.0
3d Day				
8.30 A.M.	42.3	1.6	3.0	1.6
5.00 P.M.	40.2	1.6	5.5	1.3
4th Day				
8.30 A.M.	42.1	1.6	5.1	1.5
5.00 P.M.	41.1	2.3	5.6	1.2
5th Day				
8.30 A.M.	44.1	2.4	10.0	2.9
5.00 P.M.	41.3	2.4	10.2	2.4
6th Day				
8.30 A.M.	46.2	1.5	8.7	1.8
5.00 P.M.	41.3	2.4	11.1	1.8
7th Day				
8.30 A.M.	46.2	1.9	11.4	1.6
5.00 P.M.	47.4	2.0	12.1	2.8

days' experiments there was a drop in the number of right reactions each afternoon in six instances and an increase in wrong reaction in six instances. Obviously the school work, with subject *H*, was sufficiently fatiguing to affect his ability to concentrate upon his reactions as well in the afternoon as in the morning. His introspection, however, did not discover this in all cases. On the fifth and sixth days he was aware of the fact that he was not making as good records in the afternoon as in the morning, but he felt that he was more accurate in his reactions on the afternoon of the fourth day than in the morning. A comparison of the day's work (Chart

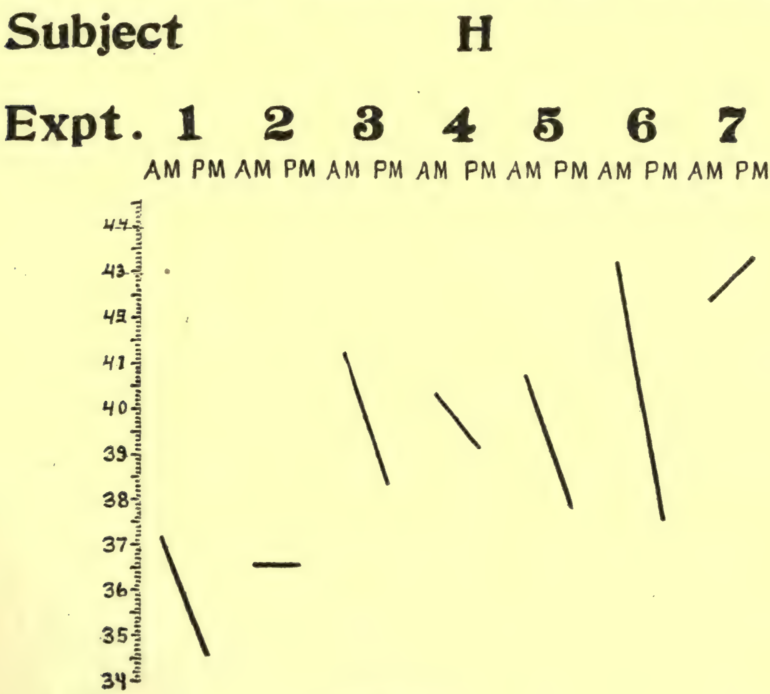


CHART III

III) will show the differences in the efficiency in the morning and afternoon work when the deduction for chance reactions is made according to the formula previously given. These differences are the more impressive as there is a learning curve apparent in the series. Nevertheless, this learning factor did not make the afternoon records superior to those of the morning, a condition which appears in the results of subjects *I*, *J* and *K*. It should be stated that *G*'s school work involved seven periods of teaching, and also a period of correcting papers.

It was not practicable to run the subjects *I*, *J* and *K* through a long series of tests. The records were taken after each of the subjects had become familiar with the apparatus and the requirements of the test, and had made sufficient number of reactions to indicate the large rise in the learning curve had been passed. Table V. shows that there was still a learning factor in their results.

Table V

Average Number of Reactions per Half Minute for Subjects *I*, *J*, *K*

Subject I

	Right	M.V.	Wrong	M.V.
1st Day				
1 P.M.	24.9	2.6	5.5	2.2
5 P.M.	25.2	3.4	4.2	1.5
2d Day				
1 P.M.	31.2	2.7	4.8	1.4
5 P.M.	26.9	3.2	9.7	3.1

Subject J

	Right	M.V.	Wrong	M.V.
1st Day				
1 P.M.	27.2	1.8	0.8	0.6
5 P.M.	29.7	3.0	1.3	1.0
2d Day				
1 P.M.	29.2	2.9	2.3	1.1
5 P.M.	31.7	2.3	1.7	0.7
3d Day				
1 P.M.	27.4	3.5	2.6	0.9
5 P.M.	26.3	4.3	2.8	1.5

Subject K

	Right	M.V.	Wrong	M.V.
1st Day				
1 P.M.	34.1	2.0	4.3	1.8
5 P.M.	36.8	2.0	3.2	1.7
2d Day				
1 P.M.	36.5	2.6	4.0	1.8
5 P.M.	35.9	2.0	3.9	1.8
3d Day				
1 P.M.	38.2	1.3	5.0	2.2
5 P.M.	36.9	2.4	6.5	2.1

The tests were made to determine whether an hour and a half of close application to memory work would affect the attention for the type of task represented by the tests. The memory work involved a rigorous concentration of the attention, which was sustained for the entire period of the experiment. Immediately after this the subjects took the discrimination reaction tests. A comparison of the work done at 1 P.M. and at 5 P.M. on the last day of the testing shows the effect of the memory work upon the reaction tests. The first day with subject *I*, and the first two days with subjects *J* and *K* involved no memory work, and in every case but one there is an improvement in the late afternoon work over that done earlier in the afternoon. After this the subjects took the discrimination reaction tests. Subject *I* shows some improvement in her reactions both for time and accuracy, in the records for her tests at five o'clock, as compared with those for one o'clock, on the first day. The second day her records for the one o'clock work were very much superior to those of the preceding tests, but after the hour and half memory work her quickness of reaction had dropped and her errors had risen conspicuously.

Subject *J* showed a marked improvement in his late afternoon work on both of the days preceding the memory work tests, but after the memory work on his third day he is not so quick and accurate as before; indeed, he gives his poorest record then.

Subject *K* was a quicker reactor than *I* or *J*. His records on the first day show an obvious improvement in the late afternoon work, but the records for the second day do not show this, though the record for the late afternoon work is but slightly below that of the earlier work. However, on the third day his record taken after the memory work shows a clear deterioration in both speed and accuracy of reactions.

It would appear from the results of *H*, *I*, *J* and *K*, that any work which calls for close application during a portion of the day will affect the ability to concentrate the attention, and the results of *E*, *F* and *G* indicate that this effect is more conspicuous than a diurnal variation. *D*'s results corroborate such an assumption.

Conclusions

While no broad generalizations can be drawn from the results of eleven subjects, it would appear that under the conditions of the experiment the following conclusions may be drawn:

1. Continuous discrimination reactions may be used as a means for detecting variations in attention.

2. The time obtained for such continuous discrimination reactions averages longer than that obtained when an appreciable time interval intervenes between a reaction and a succeeding stimulus.

3. Introspection gives a very inadequate description of variations in efficiency in this type of work.

4. Such diurnal variations as appear are due to effects from specific tasks.

5. Conspicuous differences in speed and accuracy appear even in a small and highly selected group of subjects.

EFFECTS OF LOSS OF SLEEP (I.)

BY EDWARD S. ROBINSON AND SAMUEL O. HERRMANN

The University of Chicago

Introduction

The experiments reported in this paper and in the one to follow it [1] were conducted with the purpose of studying the effects of loss of sleep. We were interested in the specific nature of these effects as well as in their general amount and direction.

The qualitative effects of insomnia are apparently quite definite. Roemer[2] and Aschaffenburg[3] have both reported that exhaustion brought on by loss of sleep characteristically increases the proportion of sound and other lower grade associations in the free association test without materially increasing association time. Patrick and Gilbert [4] have described visual hallucinations and other incidents of dissociation and disorganization. Visual hallucinations were also noted by Smith,[5] and Aschaffenburg[6] observed occasional hallucinations of contact during one of his experiments. Hypersensitivity and emotional instability were noted by Smith.

From a quantitative standpoint the outstanding paper is the one already cited, by Smith. She is the first investigator of this subject who, so far as we are aware, has obtained any-

[1] The second study was carried on by E. S. R. and F. Richardson-Robinson.

[2] Roemer, E., 'Ueber einige Bezielungen zwischen Schlaf und geistigen Taetigkeiten,' Dritter internationaler Kongress für Psychol., 1896, 353–355 (a brief summary).

[3] Aschaffenburg, G., 'Experimentelle Studien über Associationen,' *Psychol. Arbeiten*, 1895, I., 209–299, and 1897, II., 1–83.

[4] Patrick, G. T. W., and Gilbert, J. A., 'On the Effects of Loss of Sleep,' *Psychol. Rev.*, 1896, 3, 469–483.

[5] Smith, M.—A 'Contribution to the Study of Fatigue,' *Brit. J. of Psychol.*, 1916, 8, 327–350.

[6] Aschaffenburg, G., 'Ueber die psychischen Erscheinungen der Erschoepfüng,' *Arch. für Psychiatrie*, 1893, XXV., 594–597.

thing like adequate records from tests preceding and following the period of insomnia. The availability of her report renders it unnecessary to describe her work in any detail. Suffice it to say that in two tests in particular (the McDougall dotting machine and an associated words test) she found that her performance improved or remained normal for a day or so immediately subsequent to a period of insomnia. Following this, there set in a period of inefficiency from which she recovered gradually through some days. An interesting result appears in the fact that during her periods of low efficiency following loss of sleep, her efficiency could be brought back to normal rather abruptly by the interpolation of another period of insomnia. It appears that a mixture of physical weariness and mental exhaltation were present during the early period following loss of sleep when efficiency was relatively high. During the period of lower efficiency, feeling tone was practically normal.

Although Smith's results are consistent throughout her study, it does not seem to us that they can be accepted without further verification. In the first place, she employed but one subject, and, in the second place, her findings are quite unlike those of other investigators. It should be said that she herself admits her experiments might well be carried out in more extensive fashion.

The quantitative results of Aschaffenburg, Roemer, and Patrick and Gilbert show that as a result of periods of enforced wakefulness some functions show a marked deterioration while others show little or no modification. The absence of data from adequate tests before and after the loss of sleep, however, make difficult an estimate of the importance of most of those effects. In the case of Aschaffenburg's and Patrick and Gilbert's experiments a certain complication is present due to the fact that tests were applied continuously or at comparatively short intervals throughout the period of wakefulness. A certain amount of variation in these cases can probably be laid to the specific work of taking the tests rather than to the loss of sleep. But in general it can be said that the results of these investigations show a loss in efficiency as a result of deprivation of sleep rather than an initial gain followed by a loss.

There is another point of some importance upon which the data of Patrick and Gilbert and those of Smith are not in agreement. The former experimenters found from a single testing, conducted following a 90- (approximately 90-) hour insomnia and the first subsequent night of sleep, that that night of sleep, which was only from 16 per cent. to 35 per cent. longer than an ordinary one, completely restored the efficiency of the subjects employed. They present an explanatory fact in the form of the extreme soundness of the sleep of one of their subjects just after his prolonged wakefulness. While Smith noticed a prompt recovery from *subjective* consequences of loss of sleep, her *objective* results show that the process of recovery was a very gradual one extending over a number of days. Of course, it is possible that further tests upon Patrick and Gilbert's subjects would have shown the apparent restoration only temporary, and that, following it, there was a period of low efficiency lasting for a considerable length of time. There is no tangible evidence, however, for any such assumption.

Because of the disagreements which we have pointed out, it seemed to us desirable to conduct further experiments with the hope of discovering whether the stimulating effects of insomnia reported by Smith are at all general and of uncovering or verifying other consequences of loss of sleep.

Subjects

On account of the arduous nature of our first experiment, only three subjects were employed. K. R. J. is an undergraduate student, male, 20 years of age. While not overweight, he is large and athletic. At the time of the experiment he seemed in excellent health. W. H. is a graduate student, male, 22 years of age. At the time of the experiment, he was in only fair health. S. H. is a graduate student, male, 22 years of age. His health was very good.

Method

In our first experiment we studied the effects of going without sleep from the ordinary rising time on one day until the ordinary retiring time the second night following, a period of from 60 to 65 hours.

The following tests were employed:

I. *Hand Dynamometer* (*Smedley*).—Three readings were taken for each hand at every test period. The subject chose an optimal setting for the instrument at the beginning of the experiment and this was kept constant throughout the work.

II. *Tapping.*—Up to the seventh or eighth day of the experiment, a lateral tapping instrument devised by one of the writers was employed. Certain unforeseen difficulties arose at that time, and, since there was not time in which to correct them, a shift was made to the ordinary tapping board and stylus. An electric clock counter was used as a recorder. The subject tapped for a period of three minutes with each hand, a two- or three-minute rest being given between the tapping with the right hand and that with the left. The usual precautions were taken as to the position of elbow, forearm, and wrist, and grip on the stylus.

III. *Aiming.*—The materials and apparatus recommended by Whipple[1] were employed. The metronome was set at 69 beats per minute. The subject went through the series of targets forward and then backward and forward and backward with each hand. That is, there were 40 thrusts for each hand.

IV. *Reading Letters.*—This test was adapted from one of Patrick and Gilbert's ('naming of letters'). A single prose passage of 1,000 letters (12-point type) was employed. The material was arranged in a column approximately 3.5 inches in width. The subject began reading the letters aloud from right to left in the lowest line and continued up the column—reading through each line from right to left. The experimenter kept track of the accuracy of the performance by means of a copy of the text, the letters of which were enlarged.

[1] Whipple, G. M., 'Manual of Mental and Physical Tests, Simpler Processes,' pp. 147–151. See the same work for a detailed description of the dynamometer and tapping tests.

V. *Mental Multiplication.*—Twelve two-place by two-place multiplications were performed at each sitting. The problems were chance arrangements of the digits from 3 to 9 inclusive. The problems were read to the subject by the experimenter. No graphic movements were permitted during the performance. Each incorrect digit in the results was counted an error.

The subjects were put through the tests once each day beginning from 11 to 26 days before the period of insomnia and continuing until the fourth or fifth day following the insomnia. The records for each day are given in Figs. I. to X., and the exact number of repetitions for each test can readily be observed in these figures. The testing was done at the same time of day for each subject. W. H. was tested between 6 and 7 P.M., S. H. between 7 and 8 P.M., and K. R. J. between 8 and 9 P.M. Because of increasing speed of performance in certain of the tests, the test periods dwindled from one hour or longer in the early stages of the experiment to from 30 to 45 minutes at the end. The testing during each period was practically continuous, and the tests were always given in the same order.[1]

Where both speed and accuracy were involved (reading letters and mental multiplication), the subjects were instructed to work as rapidly as possible without sacrificing accuracy. During the period of fore-exercise, the subjects had access to their previous records and strove constantly to better them.

So far as we can ascertain, the subjects were free from prejudice in regard to the outcome of the experiment.

The period of insomnia, as we have said, lasted from the usual rising time on one day until the usual retiring time on the second day following. That is, each subject went without sleep on two consecutive nights. During the daytime portion of the period, the subjects engaged in their ordinary activities. At night they did whatever would make their sleeplessness

[1] This order was dynamometer, reading letters, aiming, tapping, mental multiplication.

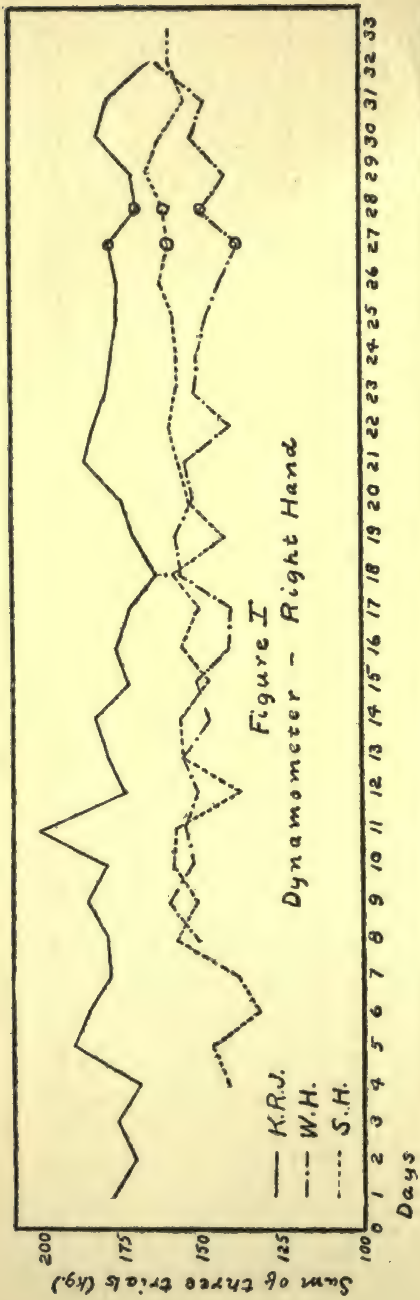


Figure I
Dynamometer - Right Hand
K.R.J.
W.H.
S.H.
Days
Sum of three trials (kg.)

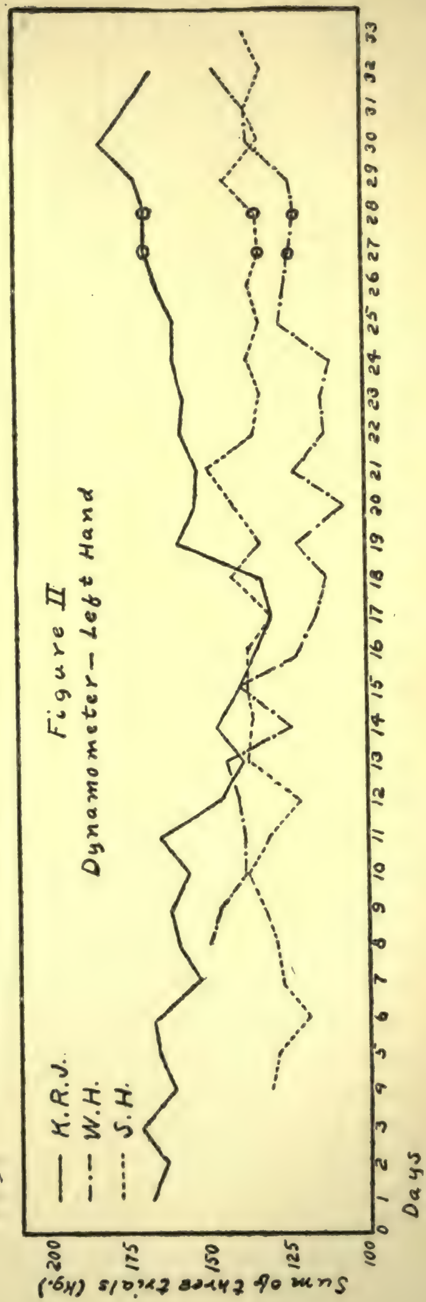


Figure II
Dynamometer — Left Hand
K.R.J.
W.H.
S.H.
Days
Sum of three trials (kg.)

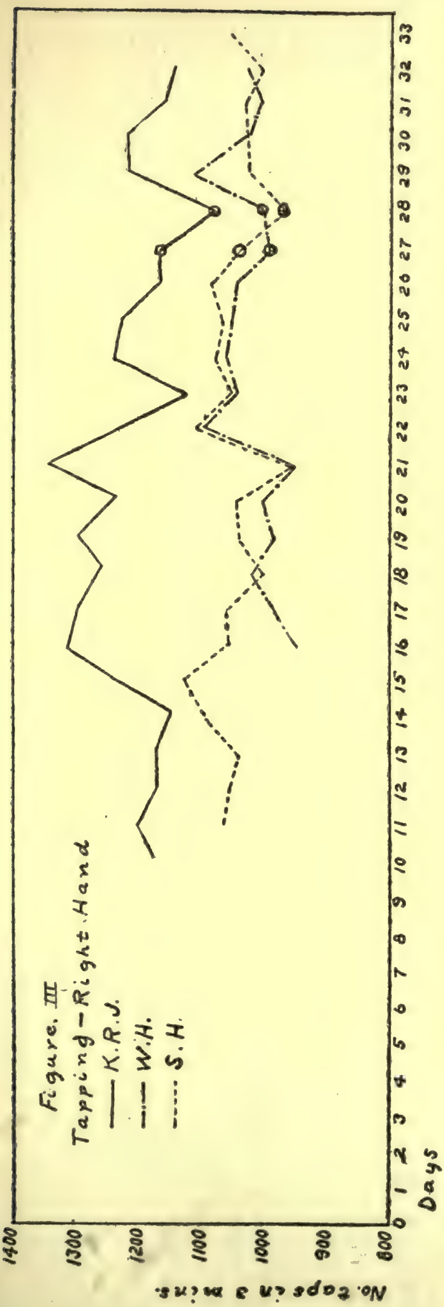


Figure. III
Tapping-Right Hand
K.R.J.
W.H.
S.H.
Days
No. taps in 3 mins.
1400
1300
1200
1100
1000
900
800
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33

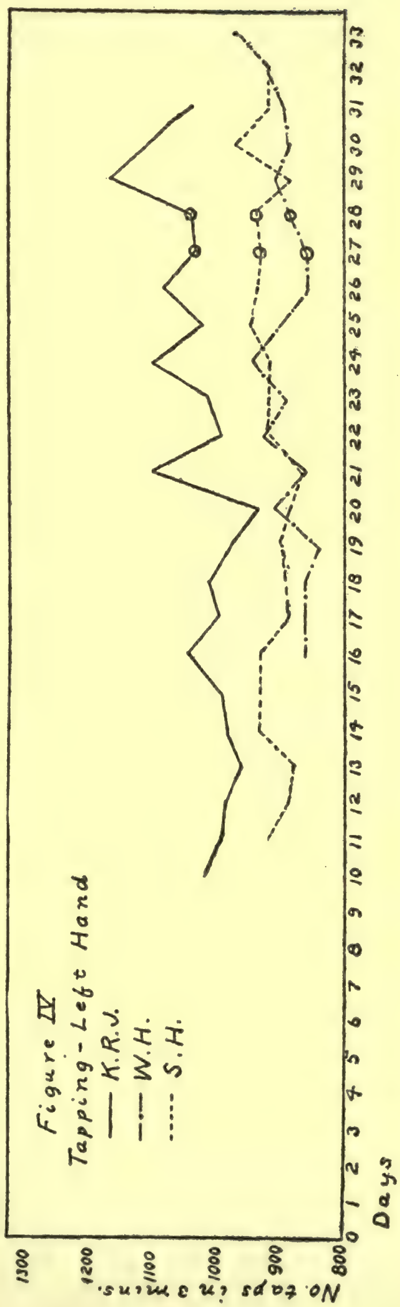


Figure IV
Tapping-Left Hand
K.R.J.
W.H.
S.H.
Days
No. taps in 3 mins.
1300
1200
1100
1000
900
800
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33

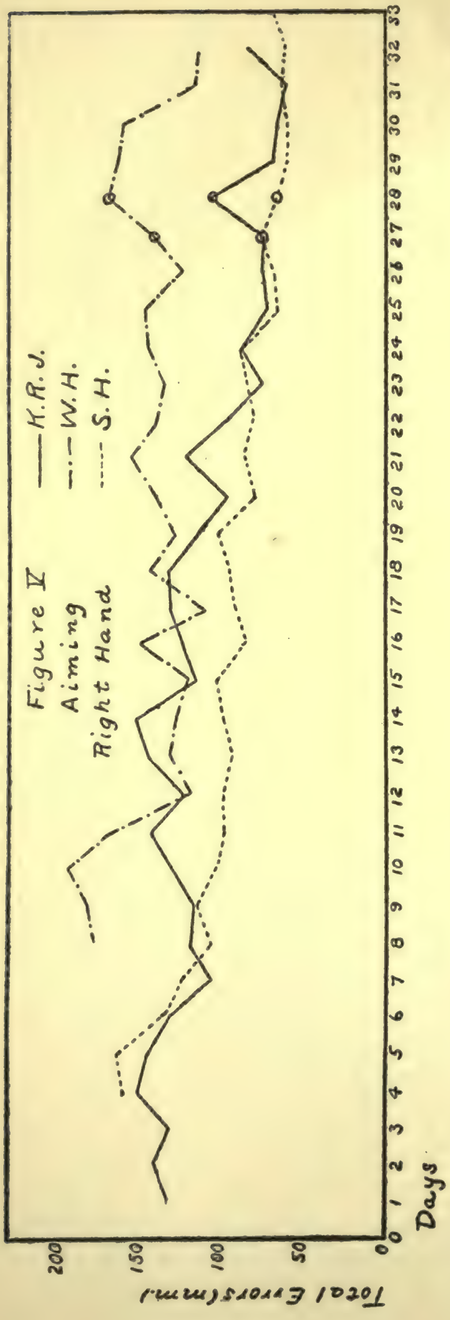


Figure V
Aiming
Right Hand
K.R.J.
W.H.
S.H.
Days
Total Errors (mm.)

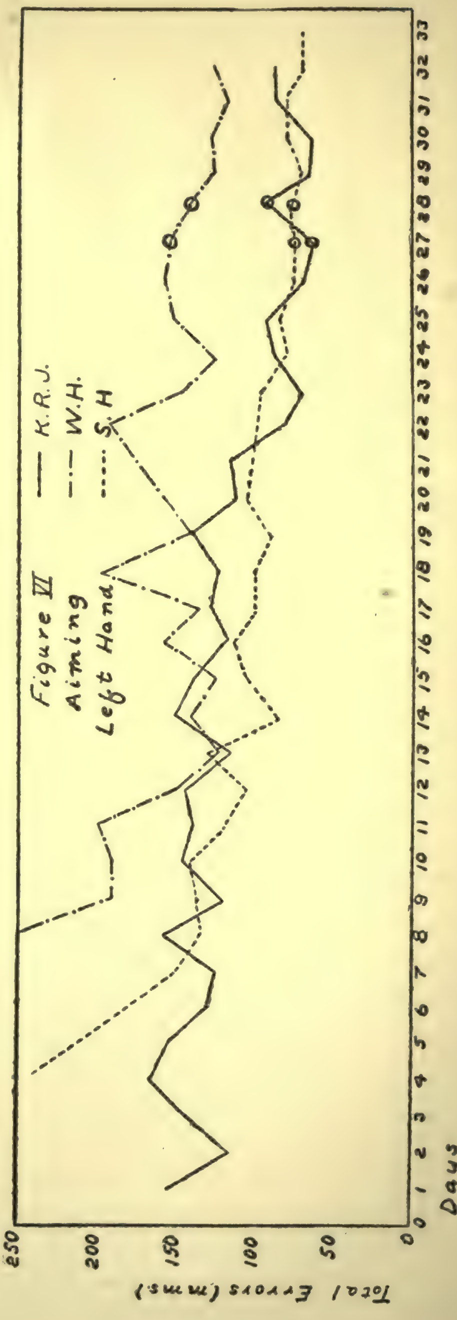


Figure VI
Aiming
Left Hand
K.R.J.
W.H.
S.H
Days
Total Errors (mms.)

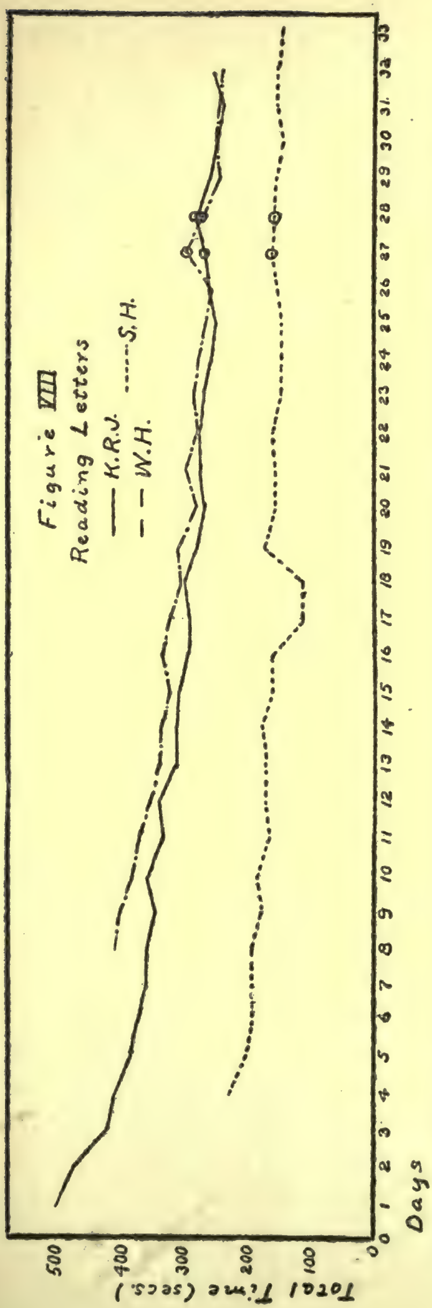


Figure VIII
Reading Letters
— K.R.J.
- - W.H.
------ S.H.
Total Time (secs.)
500 400 300 200 100 0
Days
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33

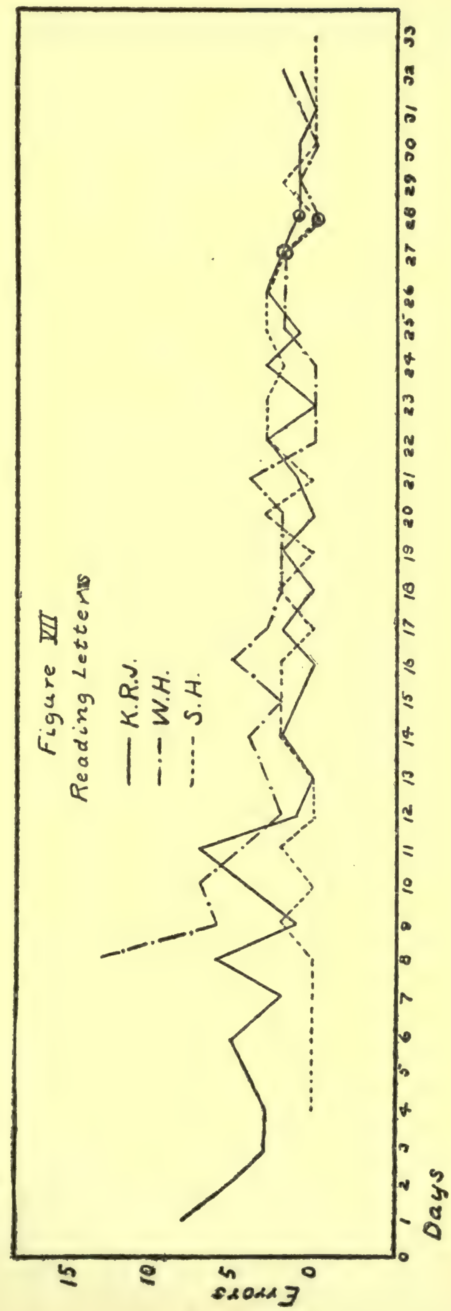


Figure VII
Reading Letters
— K.R.J.
-.- W.H.
---- S.H.
Errors
15 10 5 0
Days
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33

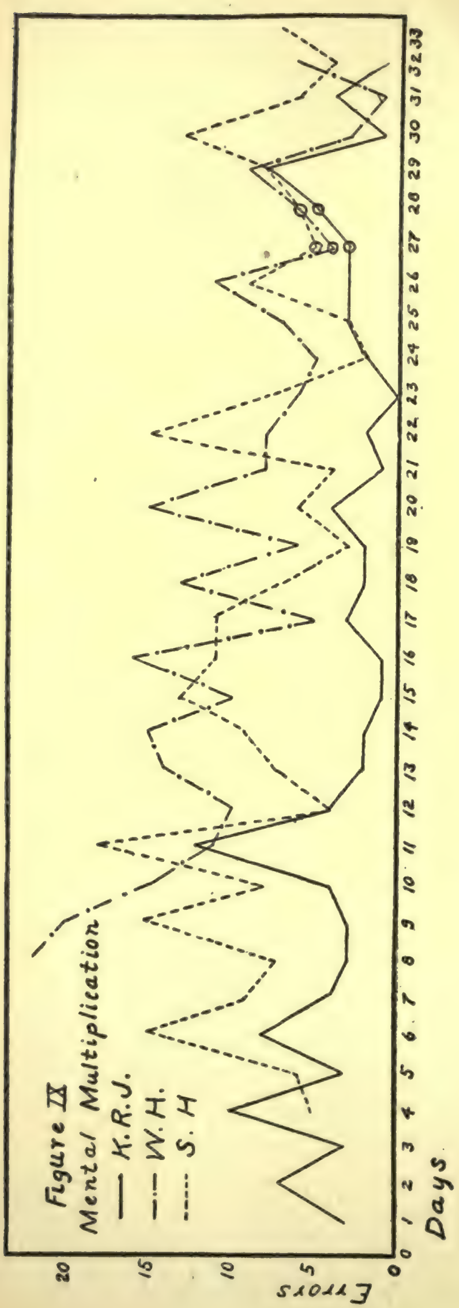


Figure IX
Mental Multiplication
— K.R.J.
—·— W.H.
----- S. H
Days
Errors
0
5
10
15
20
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33

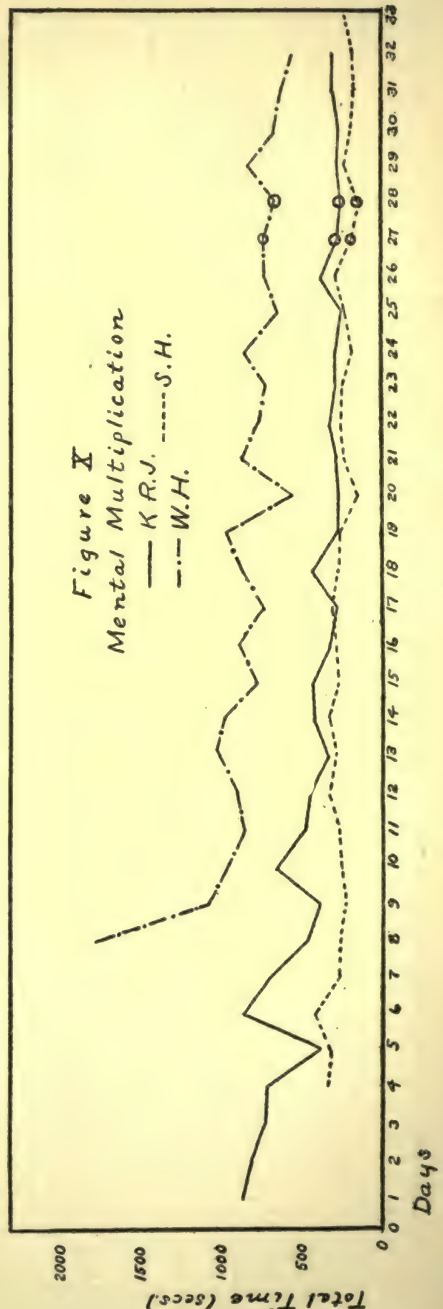


Figure X
Mental Multiplication
— K.R.J. ----- S.H.
—·— W.H.
Days
Total Time (secs.)
0
500
1000
1500
2000
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33

most assured. This varied from reading and studying to walking about, attending the theater, and in one case, taking a short railroad journey.

Results

Figures I. to X. give the plotted records of the three subjects throughout the experiment. The points in the curves marked by circles represent performances during the insomnia. That is, records for day 27 are records obtained on the day following one sleepless night, and records for day 28 are those obtained on the day following the two consecutive sleepless nights. In the interests of more ready comparisons, the curves were placed on the graphs in such a way that the insomnia records of the different subjects coincide chronologically, although as a matter of fact they did not quite do so in the actual experiment.

The effects of the period of insomnia upon performance in these tests can be judged by inspecting the irregularities apparent in the curves of performance before, during, and after the loss of sleep. The irregularities in these curves before day 27, we may be sure, are no more due to unusual amounts of sleep than to undetermined variations in a normally uniform mode of life. Variations as great or consistent as these, then, may be caused by any one of the minor fluctuations of everyday life, and, unless a newly introduced condition, such as a period of insomnia, be accompanied by greater or more consistent variations, there are no grounds for estimating its effects.

The dynamometer curves show no effects from the insomnia. The right-hand curves of K. R. J. and S. H. in the tapping test show a possible loss due to the insomnia, but the evidence furnished by the right-hand curve of W. H. and by all three left-hand curves is negative. Perhaps the accuracy of K. R. J. and W. H. in the aiming test was affected deleteriously so far as their right hands were concerned, but the other curves for this test show no such effect. The results for reading letters are not very decisive. The fact that all three subjects improved slightly in accuracy during the insomnia is not emphasized enough by our data to be taken seriously. All three of the error curves for mental multiplication rise during the insomnia. The time curves for the same test are apparently unaffected.

The main conclusion to be drawn from these curves is that the results of the tests were not affected by the insomnia in any marked or consistent manner. In so far as the curves show any general effect, they show a deleterious one. In only one (Fig. VII.) of ten sets of curves is there any evidence for increased efficiency during the insomnia. Our test results, then, would support the conclusions of Aschaffenburg, Roemer, and Patrick and Gilbert rather than those of Smith.

The fact remains, however, that most of our results are practically negative. This might be explained as due to the actual absence of any marked effects from the 60- to 65-hour insomnia, or to the fact that, feeling certain symptoms of lowered capacity, the subjects expended more effort on the tests during the period of sleeplessness. If this latter explanation is applicable to the almost negative character of our results, it might also be applied to the peculiar results of Smith. And it does not seem so highly incredible that her increased efficiency immediately following a loss of sleep may have been due to over-compensating through added effort for her lowered capacity. She claims that in at least one of her tests (the McDougall dotting machine) the possibility of variations due to changing amounts of effort was taken care of by the nature of the test. But the mere fact that the speed of performance is kept constant or that the test is exciting is, in our estimation, no guarantee that the subject will always exert a constant, in this case a maximum, amount of effort. Another factor which may have masked an actual loss of capacity in Smith and in our subjects is the increased interest in the general experimental situation which one might expect just at the time of actually going without sleep.

The qualitative effects of the insomnia upon our subjects were well marked. During the day following the first sleepless night, K. R. J. felt slightly nervous but otherwise normal. Following the second sleepless night he reported a 'buzzing in the head,' he felt sleepy, and found it almost impossible during the morning to concentrate on a book that he tried to read. W. H. reported headache and burning eyes. On the evening following the first sleepless night, he said that he felt 'almost dead.' On the day following the second sleepless night, he was very nervous and jumpy. When spoken to, he answered rapidly and with absurd emphasis. S. H., from his reports, seems to have felt the insomnia least, but even with him there was a dazed feeling, very like mild intoxication, present after the first night without sleep and growing more pronounced throughout the sleepless period.

From these observations, it is evident that the loss of sleep did affect all of these subjects, and that at least one of them, W. H., was rather severely affected. It being admitted, then, that the loss of sleep had an effect upon our three subjects, the lack of more positive evidence in the test scores may mean that the tests happened to tap capacities comparatively uninfluenced by the insomnia, or that the changes in these capacities were compensated for by some such factor as extra effort. Because of the general nature of the qualitative effects noted by us, and also by other investigators, the second hypothesis seems the more reasonable. We know that many acts of muscular strength can be performed with what is apparently the same efficiency when we are tired as when we are fresh. But acts performed when we are tired usually involve new or additional muscles. This change in the muscular make-up of an act has as antecedents those 'sensations' which we think of as our tiredness. This is probably closely analogous to what happened to our subjects. As they felt the effects of insomnia increase, they may have exerted more effort on the different tests, which really amounts to saying that they began to alter their muscular or ideational means of producing constant or nearly constant objective results.

In some form or other the above theory has been offered in explanation of other failures to obtain markedly positive results with mental tests under conditions where important changes in capacity must surely have taken place.[1] In the case of our experiment, the nervousness, the headache, the dazed condition, and the very noticeable disturbance of speech establish beyond a doubt the deleterious effects of a prolonged insomnia of which the test scores tell nothing. It is significant too, that such a function as mental multiplication is apparently as unmodified as the considerably simpler ones. If our explanation is the correct one, it points out an important limitation in the use of the usual tests for the detection of even relatively great disturbances. A desire to state the amount of such disturbances in purely objective language can be met only by test procedures of some refinement. Yet, even were such procedures readily available for application to a wide variety of functions, it is questionable whether the effects of loss of sleep, for instance, could be more vividly described than in terms of the visual hallucinations, the increased irritability, the emotional instability, the headaches, the nervousness, the dazed condition, and the disturbance of speech which this and other investigations have revealed. Suppose that we had been able to measure the degree of effort expended on mental multiplication on different days of the experiment, and suppose that we had found that efficiency during and after the insomnia was maintained only with x amount of additional effort. Those other results, obtained from what would generally be termed introspective and incidental observation, would still in our judgment be as important as any of our findings.

Our subjects reported that after a single night of sleep (average length, 8.5 hours) all detectable subjective consequences of the insomnia had disappeared. This agrees with Patrick and Gilbert, and with Smith. Our own objective results, in so far as they show any effects of loss of sleep, indicate that those effects did not persist for long after the insomnia. Here we are more in agreement with Patrick and Gilbert than with Smith.

[1] See Morgan's discussion and demonstration of the need for something other than the conventional measures of efficiency in terms of production. *Arch. of Psychol.*, 1916, No. 35.

THE EFFECT OF CHANGED DATA UPON REASONING[1]

BY EDWARD L. THORNDIKE

Institute of Educational Research, Teachers College, Columbia University

The older psychology, perpetuated in current educational doctrines and practices, regarded reasoning as a force largely independent of associative habits, which worked back to correct or oppose them. Our present psychology finds that the mind is ruled by habit throughout, the correction or opposition being of certain more simple, thoughtless and coarse habits, by others which are more elaborate, selective, and abstract. It defines reasoning as the organization and coöperation of habits rather than as a special activity above their level; and expects to find 'reasoning' and habit or association working together in almost every act of thought.

One interesting and rather important consequence of this view is the theorem that "Any disturbance whatsoever in the concrete particulars reasoned about will interfere somewhat with the reasoning, making it less correct or slower or both." It is the purpose of this article to give illustrations of and evidence for this theorem drawn from a simple experiment which can be easily given to any class, and is perhaps deserving of inclusion in a list of group experiments in psychology.

Consider the two sets of tasks in algebra, printed below. Each pair of tasks demands the application of the same principle, but the concrete situation in the one case is that with which our ordinary associative habits have been made, whereas in the other case the concrete particulars are somewhat altered. The alterations vary from slight ones, such as using p instead of x, or $b_1 + b_2$ instead of $x + y$, to a

[1] The experiments reported here were part of an investigation made possible by a grant from the Commonwealth Fund.

change from a very customary statement of a set of relations to a very rare statement, as in No. 9.

Ha.

Habitual

1. What is the square of $x + y$?
2. What is the square of a^2x^3?
3. Simplify $$4ac + \left(\frac{b^2}{c^2} \times \frac{c}{d^2} \times \frac{d^3}{b}\right).$$
4. What are the factors of $x^2 - y^2$?
5. Multiply x^a by x^b.
6. Simplify $ac - [a(b + c)]$.
7. Solve $$\begin{aligned} x + y + z &= 15, \\ 2x + y + 3z &= 22, \\ x + 2y + z &= 25. \end{aligned}$$
8. $e^2 + ef = \dfrac{g}{x}$. What does x equal?
9. There are two numbers. The first number plus 3 times the second number equals 7. The first number plus 5 times the second number equals 11. What are the numbers?

Ch.

First Changed Form

1. What is the square of $b_1 + b_2$?
2. What is the square of $r_1^3 r_{11}^2$?
3. Simplify $$5a^2b + (b^2 \div c^2)(c^2 \div d^2)(d^3 \div b).$$
4. What are the factors of $1/x^2 - 1/y^2$?
5. Multiply 4^a by 4^b.
6. Simplify $p_1p_3 - [p_1(p_2 - p_3)]$.
7. Solve $$\begin{aligned} c_1 + c_2 + c_3 &= 15, \\ 2c_1 + c_2 - 3c_3 &= 22, \\ c_1 + 2c_2 + c_3 &= 25. \end{aligned}$$
8. $e^2 + ef = \dfrac{g}{p}$. What does p equal?
9. $y = ax + b$. When $x = 3$, $y = 7$. When $x = 5$, $y = 11$. What does a equal? What does b equal?

The probability of error and delay in the necessary reasoning is clearly increased when the task is of the habitual sort (Ha.) rather than the changed sort (Ch.).

The subjects which I used were ninety-seven graduate students, divided at random into three groups of 34, 32 and

31. Group *A* did tasks 1 and 7 of Ha., tasks 4, 5, 6, 8 and 9 of Ch. and two other tasks as follows: "What is the square of k^2p^3?" and "Simplify $3ab + [b^2cd^3/bc^2d^2]$." Group *B* did tasks 3, 6 and 9 of Ha., tasks 1 and 2 of Ch. and four other tasks as follows: "What are the factors of $x - y$?" "Multiply x^4 by x^3." "Solve $x + y = 13$, $y + z = 12$, $x + z = 5$." "$e^2 + ex = g/a$. What does a equal?" Group *C* did tasks 2, 4, 5 and 8 of Ha., tasks 3 and 7 of Ch. and three other tasks as follows: "Multiply $(b_1 + b_2)$ by $(b_2 - b_1)$." "Simplify $mn - [2n(m - p)]$." "The cost of a season ticket is b dollars plus k times the cost of a month ticket. When the cost of a month ticket is $6, the cost of a season ticket is $15. When the cost of a month ticket is $15, the cost of a season ticket is $33. What is the value of b? What is the value of k?"[1]

The relative abilities of the groups may be estimated by comparing the scores in tasks estimated to be of equal difficulty, such as:

$(x + y)^2 =$? and Solve $e^2 + ef = g/p$ for p, for *A*,

$x^4 \times x^3 =$? and Solve $e^2 + ex = g/a$ for a, for *B*,

and Factor $(x^2 - y^2)$ and Solve $e^2 + ef = g/x$ for x, for *C*. Using these combinations, the percent of right answers was 71 for Group *A*, 84 for Group *B*, and 63 for Group *C*. There is thus an approximate equalization of ability of subjects between the *Habitual* and the *Changed* series.[2]

All three groups worked by the same time-schedule, which was: "0, Begin; 1 min. 30 sec., even if you have not finished 1, begin on 2; 2.30, even if you have not finished 2, begin on 3; 4.30, even if you have not finished 3, begin on 4;" and so on with 6.30 for proceeding to 5; 8.30 for 6; 9.30 for 7; 12.30 for 8; 13.30 for 9; 16.30 to stop. The subject could thus use time saved on one task for another, absolute uni-

[1] The fractions in the above and throughout were all presented in the usual 'horizontal' form, e.g. $\frac{g}{a}$.

[2] The *B* group is probably really somewhat more superior than these figures show, and the differences shown in Table I. below are somewhat greater than would be the case with perfect equalization. However, there can be no doubt that a substantial difference would remain.

formity being had only in respect to the 16 min. 30 sec. spent on the entire set of nine tasks.

The percents wrong or incomplete for the nine tasks where customary associations were favored and for the nine where some change was made were, in order:

TABLE I

	Customary	Changed	Nature of Change
1........	6	28	$(x + y)$ to $(b_1 + b_2)$
2........	34	47	a^2x^3 to $r_1^3r_{11}^2$
3........	37½	64½	$-$ to $\div$ form
4........	22	41	$x^2 - y^2$ to $\frac{1}{x^2} - \frac{1}{y^2}$
5........	55	70½	x^2 to 4^2
6........	25	53	ac, etc., to p_1p_3, etc.
7........	62	61	xyz to $c_1c_2c_3$
8........	52	53	x to p
9........	16	70	form of problem

In all but two cases (7 and 8) there is a substantial interference with thought by the change. In No. 9 the amount of novelty introduced is much greater than in the other cases, and is perhaps more truly called a change in the principles and operations used than a change in the concrete particulars reasoned about. In No. 6 the use of a minus sign in place of the plus may have added a little difficulty over and above the change from literal to subscript distinctions.

We may now examine certain facts from the accessory tasks which were used partly to conceal the special point of the experiment from the participants and partly to secure additional data.

"Multiply $(b_1 + b_2)$ by $(b_2 - b_1)$" is not strictly comparable with "What is the square of $x + y$?" but the rise in errors from 6 percent for the latter (Group *A*) to 61 percent for the former (Group *C*) would probably have been much less if "Multiply $(x + y)$ by $(y - x)$" had been used.

The change from "What is the square of a^2x^3?" (Group *C*) to "What is the square of k^2p^3?" (Group *A*) raises the errors and omissions from 34 percent to 47 percent.

Simplify $3ab + [b^2cd^3/bc^2d^2]$ produces more errors than the same task when arranged in $+ (\times \times)$ or in $+ (\div \div)$ form (71 percent in Group *A*). This may seem to be in contradiction to our general theorem, but an inspection of the errors makes it probable that the parentheses were a protection against two habits, one of canceling upper and lower numbers seen in proximity, the other of clearing of fractions any expression that contains one. The former leads to canceling the b of $3ab$ with the b of bc^2d^2, which is done by one of the thirty-four subjects. The latter adds an operation in which errors are made by four of the subjects, and makes the needed cancellations more difficult. It also encourages the error of dropping out the common denominator, which is done by four more, and permits the worker to feel that he has done his duty by the task, though he has not cancelled. Finally, there are further erroneous manipulations of the denominator (as in $(3ab + b^2cd^3)/bc^2d^2$) from which the parentheses and signs of multiplication and division were perhaps a protection in the other forms of the task.

The question "What are the factors of $x - y$?" evokes responses which show that when a novel situation is met, one's associative habits do not retire while reason attacks. On the contrary, the forces of habit are nowhere more evident than in the treatment of novel situations. Which habits will act depends on which elements or features of the situation are given weight and upon the amount of weight given to each. With our subjects, the features that this is the difference not of two particular numbers but of any number from any number and that 'any number' is as truly the square of '$\sqrt{\text{any number}}$' as '(any number)^2' is of 'any number,' are usually neglected or underweighted, there being only four answers of

$$(\sqrt{x} + \sqrt{y})(\sqrt{x} - \sqrt{y}) \text{ or } (x^{1/2} + y^{1/2})(x^{1/2} - y^{1/2})$$

The underweighting of these features permits the habits of not factoring prime numbers in arithmetic, and of not factoring expressions like $x - y$ in algebra, to act acceptably, answers of $1(x - y)$ being the commonest given (11 out of 32).

"Multiply x^4 by x^3" is, consistently with our general theorem, easier than x^a by x^b.

On the whole, it seems certain that even such slight changes as from the customary a, b, x and y, to k and p or to p_1, p_2 and p_3 or to p_1, p_{11} and p_{111}, impede thought, and that the general theorem does hold that "Any disturbance whatsoever in the concrete particulars reasoned about will interfere with reasoning."

THE EFFECT ON FOVEAL VISION OF BRIGHT (AND DARK) SURROUNDINGS. V

BY ELLIOT Q. ADAMS AND PERCY W. COBB,

Laboratory of Pure Science, Nela Research Laboratories, Nela Park, Cleveland, Ohio

A study of the variation of the difference-threshold with independent changes in the brightness of the test field and of the surroundings has been reported in an earlier paper.[1] In that paper there were deduced, on the basis of a physico-chemical explanation of contrast, certain theoretical considerations which proved to be in only limited agreement with the results of observation. It has since been found possible with a few simple, and qualitatively[2] plausible, assumptions as to the nature of the nerve processes of the retina, to derive a theoretical expression for the variation of the difference-threshold which is in good agreement with the experimental results.

Assumptions

The assumptions made are:

1. Over the range of ordinary photometric work, with complete adaptation to a field of constant and uniform brightness, the fractional difference-threshold reaches a value which is independent of the brightness, and not greater than that for any non-uniform field.

2. With a non-uniform field, the state of adaptation at any point is that which would have resulted from a uniform field of brightness equal to a weighted[3] quadratic mean of the brightness of the different parts of the field.

[1] P. W. Cobb, J. Exp. Psych., *1* (6), 540–566 (Dec., 1916).

[2] Exact accuracy of the postulates is not claimed. The accuracy possible in determinations of threshold does not appear to warrant refinement of the assumptions made.

[3] 'Weighted' to take account of the relative areas of the different parts of the field, and their visual importance with reference to the test field. The *arithmetic* mean was tried but the formula so derived did not agree so well with the experimental results.

3. The response of a fiber of the optic nerve consists of a series of impulses[1, 2] whose frequency increases, at constant adaptation, with the brightness at the corresponding point of the field.

4. The interval between impulses consists of two parts, a constant 'absolute refractory period,' and a 'relative refractory period' which varies inversely with the brightness at the corresponding point of the field, *i.e.*, after the end of the absolute refractory period following one impulse the sensitiveness (the reciprocal of the least effective brightness) increases linearly with time until its product with brightness reaches a certain critical value, whereupon another impulse follows.

5. Equal differences in frequency of response are equally perceptible.

Constant Adaptation

For simplicity consider first the variation of the fractional difference-threshold for a 'test field' as a function of the variable brightness of the test field. This condition will obtain *approximately* when a test field of the smallest practicable size is viewed against constant surroundings.

The interval between impulses will be

$$t = t_\infty + \frac{kt_\infty}{B}, \tag{1}$$

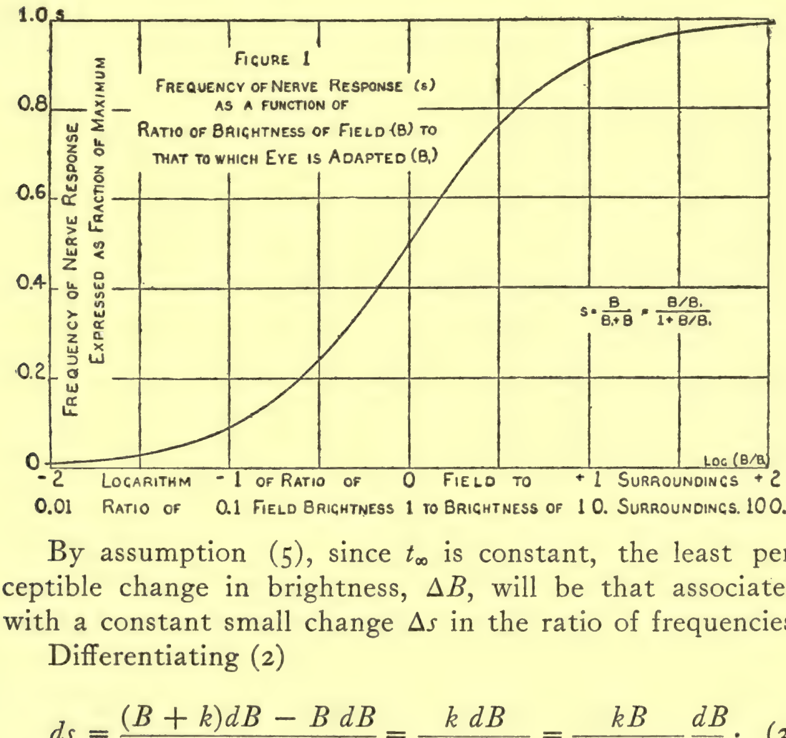
where t is the interval; t_∞ the absolute refractory period, assumed constant; B the test-field brightness and k a constant such that kt_∞/B is the relative refractory period. It can be deduced from postulates 1 and 5 that k will vary proportionally with the adaptation brightness, B_1, over the range within which complete adaptation is possible; and it will be shown below that $k = B_1$.

[1] Keith Lucas, 'The Conduction of the Nervous Impulse,' London (1917), p. 9.
[2] L. T. Troland, *J. Opt. Soc. Am.*, *4*, 160 (1920).

The *frequency* of the impulses will be $1/t$, and the ratio of the frequency to the maximum possible, $1/t_\infty$, will be

$$s = \frac{t_\infty}{t} = \frac{1/t}{1/t_\infty} = \frac{1}{1 + (k/B)} = \frac{B}{B + k}. \tag{2}$$[6]

This ratio of frequencies is shown, in Fig. 1, plotted against log (B/k), the logarithm of the brightness of the test field (in terms of $k = B_1$).



By assumption (5), since t_∞ is constant, the least perceptible change in brightness, ΔB, will be that associated with a constant small change Δs in the ratio of frequencies.

Differentiating (2)

$$ds = \frac{(B + k)dB - B\,dB}{(B + k)^2} = \frac{k\,dB}{(B + k)^2} = \frac{kB}{(B + k)^2}\frac{dB}{B}. \tag{3}$$

Regarding Δs and ΔB as infinitesimal changes, the least perceptible fractional change in B, *i.e.*, the fractional difference threshold, is related to Δs as follows:

$$\frac{dB}{B} = \frac{(B + k)^2}{kB}ds = \left(\frac{B}{k} + 2 + \frac{k}{B}\right)ds. \tag{4}$$

Differentiating the coefficient of ds in (4) and equating to zero to determine the value of B corresponding to minimum threshold,

$$\frac{d}{dB}\frac{(dB)/B}{ds} = \frac{d}{dB}\left(\frac{B}{k} + 2 + \frac{k}{B}\right) = \frac{1}{k} - \frac{k}{B^2} = 0. \qquad (5)$$

Whence $k = B$, but since by assumption (1) the fractional difference threshold is a minimum when the field is uniform, that is when $B = B_1$,

$$k = B_1. \qquad (6)$$

Substituting in (4) and writing ΔB and Δs for dB and ds:

$$\frac{\Delta B}{B} = \left(\frac{B}{B_1} + 2 + \frac{B_1}{B}\right)\Delta s. \qquad (7)$$

The coefficient of Δs in (7) is shown in Fig. 2, plotted against log B/B_1, the logarithm of the ratio of the brightness of the (infinitesimal) test field to that of the surroundings.

Effect of Test Field on Adaptation.

If account be taken of the effect of the brightness of the test field on the state of adaptation, equation (7) will still hold, provided that by B_1 is meant not the brightness of the surroundings B', but the quadratic mean brightness,

$$B_1 = \sqrt{\alpha B^2 + (1 - \alpha)B'^2}, \qquad (8)$$

[1] It is of interest to point out that equation (9) of the earlier paper (1),

$$L = \frac{WB}{S - W}$$

becomes, on solving for W/S

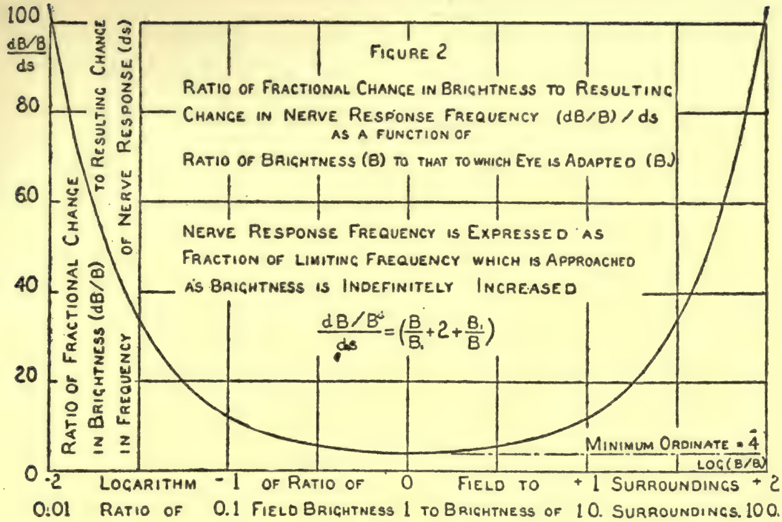
$$\frac{W}{S} = \frac{L}{L + B}$$

or in the notation of this paper, if W/S be identified with s:

$$s = \frac{B}{B + k},$$

which is identical in form with equation (2) above.

It follows that the assumptions of the former paper would, with the addition of assumption (1) of the present article, give the same mathematical results as those derived above.



where α is a coefficient, to be determined empirically, expressing the relative effectiveness of test field and surroundings on adaptation.

Comparison with Experimental Results.

For comparison with the equations (7) and (8), Table I. gives the observed fractional threshold,[1] as a function of the

Table I

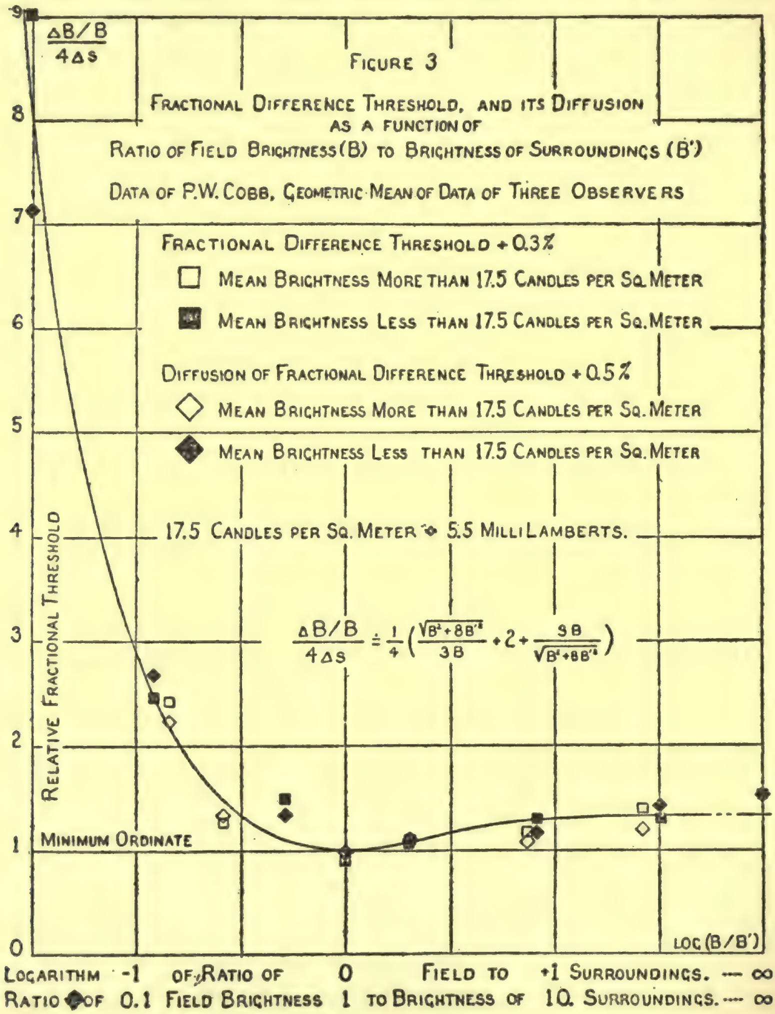
Difference Threshold and Diffusion, Measured as Fractional Changes in One Half of a Test Field of Angular Dimensions 2.0° by 2.6°

	Brightness		Fractional Difference			Brightness		Fractional Difference	
	Test Field	Surroundings	Threshold	Diffusion		Test Field	Surroundings	Threshold	Diffusion
a	436	17.3	0.42%	0.60%					
b	128	17.3	0.35	0.54	*u*	17.5	121	0.73%	1.12%
c	35.3	17.3	0.32	0.54	*v*	17.6	46.0	0.38	0.67
o	17.6	17.3	0.27	0.49	(*o*	17.6	17.3	0.27	0.49)
d	8.92	17.3	0.45	0.67	*w*	17.6	8.64	0.33	0.55
e	2.11	17.3	0.74	1.34	*x*	17.6	2.15	0.39	0.61
f	0.540	17.3	2.71	3.57	*y*	17.6	0.547	0.39	0.71
g	0.0	17.3	∞	∞	*z*	17.6	0.0	0.46	0.76

Brightness is expressed in candles per square meter.

[1] Geometric mean of values for three observers. Quoted from Table V., p. 547 article cited in Footnote 1.

brightness of the test field, and of that of the surroundings. These data are shown by points in Fig. 3, while the line represents the coefficient of Δs in equation (7), when the



coefficient α, of equation (8) is given the value $\alpha = 1/9$. To reduce the three functions to a common basis, namely a minimum ordinate of unity, the values of the difference-threshold have been divided by 0.3 per cent., those of the diffusion by 0.5 per cent., and the coefficient of Δs in (7), by 4. The thresholds are indicated by squares the diffusion values by 'diamonds', Black squares indicate results obtained when the mean brightness was less than 17.5 candles per square meter, white squares when the mean brightness was greater. It will be seen by comparing the number of light and dark squares above and below the curve that increase in the general level of brightness causes a slight but apparently definite reduction in the thresholds. This difference represents the degree in which the assumptions (in particular assumption (1) in the opinion of the writers) fail to take account of the facts.

Fig. 1

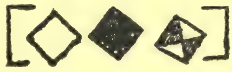


Fig. 2

With a larger test field the loss in accuracy with dark surroundings would be less pronounced.

Conclusion

Based on the assumption that visual impressions are transmitted along each fiber of the optic nerve by a series of impulses whose effect depends only on their frequency (the All-or-None hypothesis of Keith Lucas[4, 5]) quantitative expressions for the effect of light and dark surroundings on vision have been derived, which for foveal vision, with empirically selected values of the coefficient for the relative effect of different parts of the field, accord well with observations of difference threshold, and of its 'diffusion.'

Further experimental work will be needed to establish the way in which the coefficient before mentioned is related to the size and location of the test field, the extent to which the assumptions on which the formulas are based are in accord with fact, and the range of conditions over which they continue to hold.

Nela Research Laboratories,
Nela Park, Cleveland, Ohio,
March, 1922

A NEW LABORATORY AND CLINIC PERIMETER

BY C. E. FERREE AND G. RAND,

Bryn Mawr College

This apparatus was devised in response to a request from a committee appointed by the American Ophthalmological Society to work out a better standardization of the illumination of perimeters and test charts. The request was for a feasible means of illuminating the perimeter arm with light of a good intensity and quality, so that every point on the arm in any meridian in which it may be placed shall receive equal intensities of light. Intensity and quality of illumination, however, are only two of the factors which influence the results of the perimetric determination. In devising the instrument described in this paper it has been our purpose to provide a control also of other factors which are of importance to the work of the office and clinic. The perimeter devised and to be described in this paper has been adopted by the Committee as the standard instrument for office and clinic work.

The variable factors which influence the apparent limits of color sensitivity are, so far as we have been able to discover, the wave-length and purity of the stimulus, the intensity of the stimulus and the visual angle, length of exposure of the eye, the method of exposure (moving or stationary stimulus), accuracy and steadiness of fixation, the intensity of the general illumination of the retina and its state of adaptation, breadth of pupil, and the brightness of the preëxposure and of the background or surrounding field. The most important of these from the standpoint of the office or clinic are perhaps the intensity of the stimulus, the brightness of the preëxposure and the surrounding field, the intensity of the general illumination, and the accuracy and steadiness of fixation.

Perhaps errors in refraction should be included in the above list of factors. They differ from those factors, however, in that they are a source only, or mainly at least, of differences in breadth of field between observers, not for the same observer at different times. They are not therefore a serious source of trouble in the use of perimetry to check up the advance or recession of a given pathological condition, but they are confusing in diagnosis. We have no certain means of telling, for example, how much of the results in any given case of high myopia is due to the refractive condition and how much if any is pathological. To add to our knowledge on this point we are now conducting an investigation to determine the effect of various amounts and kinds of refractive errors on the results of field taking. We can not help but feel, as is stated and discussed in the paper, that a provision should be made in our practice of field taking to include one set of results under the correcting glasses, even though it means either working under conditions which give a narrow field or being content with results which do not include the wider portions of the temporal field.

1. *Intensity of Stimulus.*—By a sufficiently wide variation of this factor alone, the zones of color sensitivity may be made to have almost any breadth within the limits of the field of vision, to differ radically in shape, and even to change or reverse their order of ranking as to breadth. When pigment surfaces are used as stimuli the illumination of the perimeter arm determines the intensity of the stimulus light. Two methods are proposed for securing an even illumination of the stimulus at every point on the perimeter arm and of reproducing this illumination from time to time.

Method 1.—When the source of light is inlaid in the surface of the arm or its continuation, the illumination on this surface will be equal for approximately 180° on either side of the source. The value of this illumination at every point will be equal to the normal flux of light from the luminous surface divided by four times the square of the radius of curvature of the perimeter arm, or four times the square of the distance of the eye from the perimeter arm. A perimeter embodying this principle of illumination is being constructed in the following way. A lamp house is fastened on the continuation of a 90° arm in such a position that an opening in its surface facing the observer lies in the continuation of the surface of the perimeter arm. This opening is filled in with diffusing glass bent to take the curvature of the arm and shaded in such a way as to shield the eye of the physician and the observer without changing or interfering with the distribution of light to the perimeter arm. The lamp house rotates with the arm and thus illuminates it uniformly at every point in all meridians.

The principle by which an even illumination of the perimeter arm is secured by this method may be demonstrated as follows. Let S, Fig. 1, be the source of light inlaid in the surface of the arm; P any point in the perimeter arm that is to be illuminated; X the distance from S to P; y the radius of curvature of the arm or the distance of the eye from the arm; a the angle of emission of the light from the source S; and θ the angle of incidence of this light at the point P. Then the intensity at P will be inversely as the square of the distance of P from S or inversely as the square of X, and directly as the cosine of the angle of emission a, $\frac{1}{2}x/y$, and of the angle of incidence θ, also $\frac{1}{2}x/y$. That is,

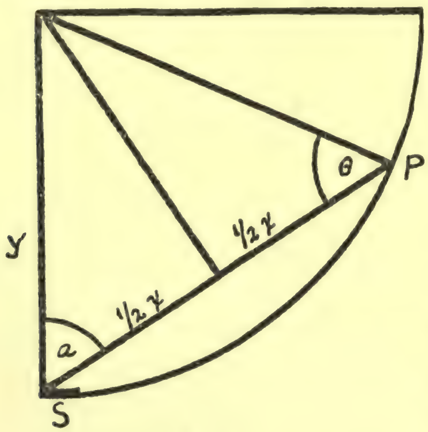


FIG. 1

$$I_1 = I \times \frac{1}{x^2} \times \frac{\frac{1}{2}x}{y} \times \frac{\frac{1}{2}x}{y} = I\,\frac{\frac{1}{4}x^2}{x^2y^2} = \frac{I}{4y^2}\,;$$

in which I_1 is the intensity of light at P and I the intensity at S. From this equation is derived the law of illumination of the arm,—the intensity of light at any point on the arm is equal to the normal flux of light from the source divided by four times the square of the radius of curvature of the arm.

The method has the following objections. (1) The difficulties in construction are not easy to overcome. (2) Evenness of illumination requires an approximately perfectly diffusing glass. This glass is difficult to obtain and prepare and its transmission is apt to be low. Moreover the light incident on the perimeter arm should approximate daylight in composition. The selective absorption required to correct the light from the lamp to this composition further reduces the intensity enormously. This double loss, first by absorption and second by the somewhat wasteful type of distribution employed, renders it difficult to get an adequate intensity of illumination of the perimeter arm.

Method 2.—When the source of light lies in the perpendicular to the plane of the perimeter arm at its center of curvature, it will be equidistant from every point on the arm; also the angles of emission and incidence of the beam of light will be equal for every point on the arm. A perimeter has been constructed embodying this principle of illumination. This perimeter is shown in Figs. 2–5. Two arcs of the same radius of curvature were constructed at right angles to each other; one a 180° arc, the perimeter arm; the other a 90° arc, the lamp arm, at the end of which is placed the source of light. In order that the source of light shall sustain a fixed relation to the perimeter arm for all positions of that arm, the two arms are fastened together at the center of rotation. About the source is a housing which was designed in such a way as to shield the eye of the patient and the physician without interfering with the distribution of light to the perimeter arm. This housing is made of black japanned iron and is painted a mat black on the inside in order that as nearly as possible all of the light which passes to the perimeter arm shall radiate directly from the lamp filament. Its dimensions are $4\frac{1}{4}$ x $4\frac{1}{4}$ x 5 in. A rectangular aperture $2\frac{1}{8}$ in. wide was cut out of the side of the housing facing the perimeter arm, at the bottom and for 3 in. back on the two adjacent sides. The relation of the lamp to the aperture of this housing is such that the light radiates freely without shadows from the filament to every point on the perimeter arm. In order that the lamp may be removed when desired, the bottom of the housing is hinged at the back and is held in place by a latch on either

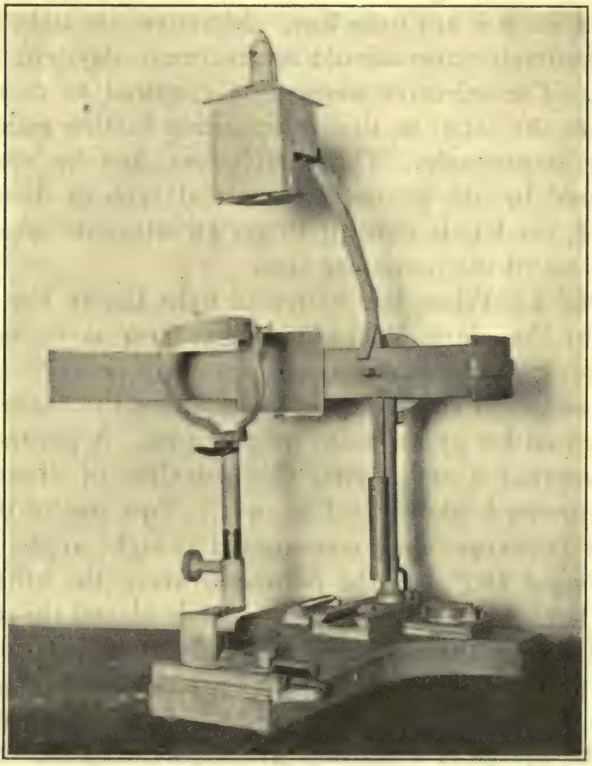


FIG. 2

side. To prevent overheating the housing is well ventilated by especially designed light-tight ventilators, four in the sloping roof of the lamp house and four on each of the sides at the bottom.

Provisions are made in the construction of the lamp house for filtering the light to daylight quality. This is accomplished in two ways. (*a*) A well-seasoned 75-watt type C_2 (blue bulb) Mazda lamp operated by ammeter and rheostat control is used as the source of light. This is the ordinary blue bulb commercial lamp and gives only approximate daylight quality. The surface of the bulb is acid-etched to diffuse the light. The intensity of light incident on the perimeter arm when this lamp is used as source of light is 17 foot-candles. (*b*) A type C (clear bulb) Mazda lamp of

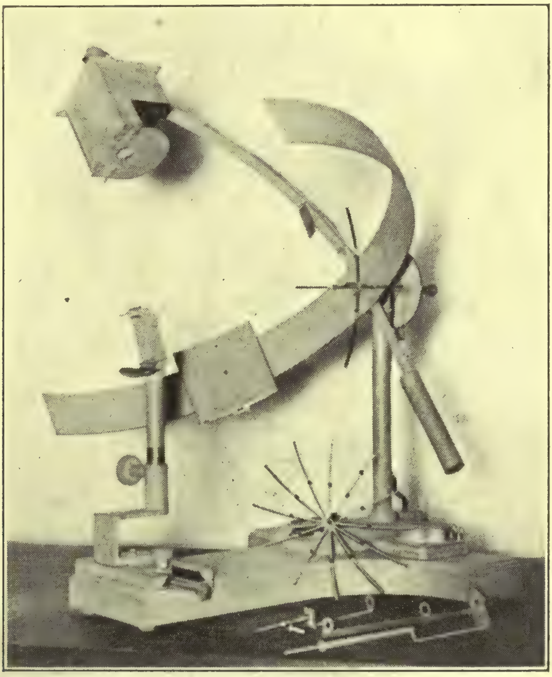


Fig. 3

75, 100 or 150 watts depending upon the intensity of light desired, is used as source. The light from the lamp is filtered to daylight quality by means of a double-etched collar or cylinder of carefully prepared daylight glass $2\frac{1}{2}$ in. broad and 4 in. in diameter, which can be inserted into the bottom of the lamp house so as to fill completely the three-sided aperture. When in position all of the light emitted from this lamp house must pass through the filter. The collar is securely held in position by three pins so placed in the bottom of the lamp house as to form the apices of an equilateral triangle of the appropriate dimensions. This bottom is hinged so that it can be dropped down for the insertion of the collar and is provided with latches so that it can be held securely in position when raised. The light of a type C Mazda lamp

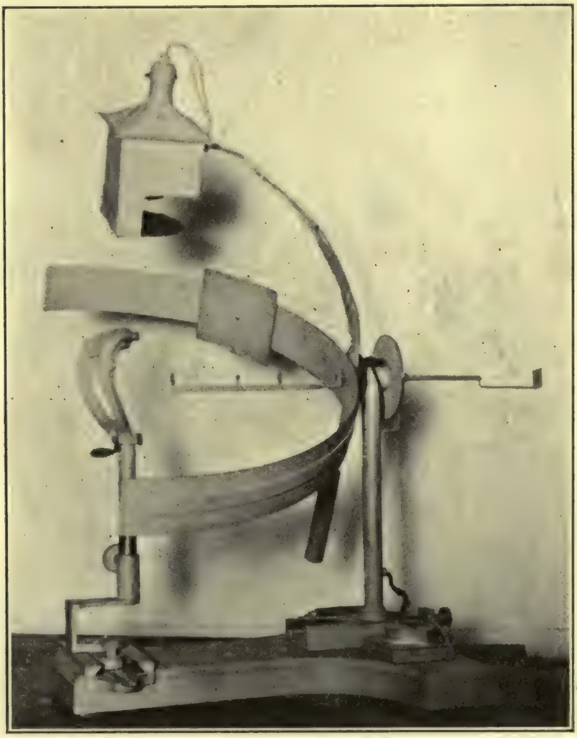


FIG. 4

filtered by this glass is a very close spectro-photometric approximation to north skylight. The coefficient of transmission of the glass is 15 per cent. The light incident through the filter on the perimeter arm from a 100-watt type C Mazda lamp has a value of 2.6 foot-candles,—an intensity very suitable for the taking of a low illumination field the importance of which in diagnosis will be discussed later. Higher intensity fields may be obtained by using higher wattage lamps.

This perimeter is not difficult to construct or to operate. It provides for a uniform illumination of the perimeter arm in all meridians with light of a good intensity and quality; and with it a precision of control is possible which is com-

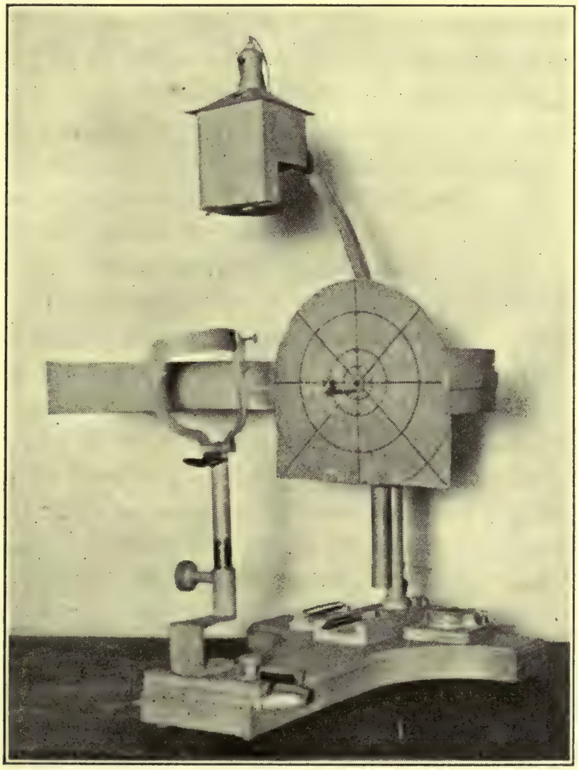


Fig. 5

parable with the work of the physical laboratory. Of the two instruments we have devised, it is without doubt much the more feasible and it is also very probably the more correct in actual practice. Both instruments are correct in theory.

2. *The Brightness of the Preëxposure and the Surrounding Field.*—The brightness of the surface to which the eye is

preëxposed may change the apparent limits in certain meridians as much as 17–20°. A preëxposure lighter than the color gives a dark, and one darker than the color a light after-image. These after-images change profoundly the saturation of color sensation, also its hue. A background or surrounding field lighter or darker than the color produces a similar effect on the limits, but not so great. In this case the disturbing achromatic effect is due to physiological induction or contrast. The variable effect of brightness of preëxposure and surrounding field can be eliminated only by making both a gray of the same brightness as the stimulus color. Here again a precise control of the intensity of the illumination for all points on the perimeter arm becomes important. That is, the shade of gray which is needed to match the color in brightness changes with change of illumination; therefore, the selection of a gray which will match the color in brightness for all points of work presupposes constancy and uniformity of illumination. A further advantage is gained by making the background of the same brightness as the color. That is, when color and background are of the same brightness the stimulus disappears completely when the limit of sensitivity to that color is reached, instead of turning into a gray concerning the colorlessness of which the patient is apt to be in doubt. This gives the effect of the disappearance type of photometer and like it adds greatly to the ease and certainty of making the judgment.

The control of brightness of preëxposure and surrounding field is provided for in the following way. To the stimulus carriage is attached a light aluminum holder, No. 19 B. & S. gauge, grooved to hold a card 5 x 6 in. These cards are covered on one side respectively by grays of the brightness of the four colors, red, yellow, green and blue, of the Hering standard series of pigment papers, as seen in the peripheral retina. At the center of each of these cards is pasted a disc of the appropriate color subtending a visual angle of 1°. To provide for the control of the preëxposure for the stationary method of giving the stimulation, cards identical with the background cards are provided, covered also on one side with a gray of the brightness of the color. The stimulation by this method is given as follows. The stimulus is placed at the point to be tested and covered with the preëxposure card. The observer is told to take his fixation. At a given signal the stimulus is uncovered for one second and re-covered. In case the moving stimulus method is used, the surrounding field serves as the preëxposure.

The perimeter arm and body are painted a gray of a shade which is approximately mid-gray to the blue and yellow, the darkest and lightest of the stimuli employed. In our own laboratory the perimeter is used on a table painted with the same gray and stands before a gray screen. These latter precautions, however, are not necessary.

When the instrument is supplied to the profession, provision will be made that the stimulus and preëxposure cards, a seasoned lamp, and all other perishable parts can be furnished in the quantities desired.

3. *The Accuracy and Steadiness of Fixation.*—All are familiar with the disturbing effect of inaccuracy and unsteadiness of fixation. If correct and reproducible results are to be obtained, the eye must be accurately placed at the center of the sphere in the surface of which lies the perimeter arm, and the line of sight must not shift from the fixation point while the color observation is being made. As an aid to the correct placement of the eye and a check on its steadiness of fixation, two devices have been provided. (*a*) A small circular mirror is used as a fixation object in which the observer sees the image of his own eye (Fig. 2). When the eye is correctly placed with the line of sight normal to the surface of the mirror at its central point, the fact is indicated to the observer by the position of the image of his pupil and iris as seen in the mirror. Not only is this simple device of service in determining the correct position of the eye, but it aids the observer greatly in holding a steady fixation.

We scarcely need to point out that a fixation object does not afford an accurate control of fixation. Exact checking methods show that the patient is not always fixating the object when he thinks he is. The only guide to monocular fixation is clearness of seeing and this is a criterion that presents considerable latitude. However with a mirror the patient has an objective check on the position of his eye.

One of the objections to the use of a mirror as a means of controlling fixation is the liability of glare from its surface probably due to a combined specular and diffuse reflection, rendering in proportion as it occurs the clear seeing of the image of the eye by the patient unnecessarily difficult. This objection has been obviated in the instrument described by cutting off the direct radiations from the lamp from the mirror by a narrow shield which can be turned back out of the way when not in use. With the shield in position the eye receives from the mirror only the light which is first reflected from the eye to the mirror and then back to the eye. This device adds greatly to the ease and clearness with which the patient sees his imaged eye. The elimination of the troublesome glare on the surface of the mirror is rendered particularly simple and easy in case of this instrument because of the plan of illumination employed, *i.e.*, the light all comes from a fixed source above and directly in front of the mirror. If the perimeter were illuminated from a window, for example, and its position with reference to the window changed as is needed in order to give as nearly as possible equal illumination of the stimulus in its various positions, the elimination of glare from the surface of the mirror would not be so simple a problem.

Another theoretical objection to the mirror is that the clear seeing of the iris in the mirror requires the eye to focus for twice its distance from the perimeter arm. When this distance is 33 cm., as in the present case, the use of the mirror as a fixation device throws the refracting system of the eye 1.50 diopters out of focus for the colored stimulus on the perimeter arm. Just how important this is when the problem is the mapping of the limits of color sensitivity and not a determination of acuity is difficult to say *a priori*. The poor imaging for the peripheral field is well known to all who have made determinations of acuity at points far removed from the center of the field. The additional blurring due to an error of focusing of the magnitude described is probably of negligible consequence at points having an excentricity of 30–70 degrees, particularly in determinations of color sensitivity. In any event we have not been able to detect appreciable differences in the limits of sensitivity at 17 foot-candles of illumination with the mirror as fixation device and with the other device provided, in case of which the eye is correctly focused for clear seeing at 33 cm.

(*b*) The second device for the control of fixation is similar in principle to a peep-sight and may be called a parallax or peep-sight device (Fig. 5). A small disc or button (8 mm. in diameter) at the center of rotation of the perimeter arm is viewed through a circular opening (7 mm. in diameter) in a thin round metal disc 9 cm. nearer to the eye. The plane of the disc and the viewing opening are both normal to the line of sight when the eye has its correct position and fixation. When the eye has this position and fixation the relation of size of disc and opening is such that the disc is seen not quite to fill the opening. The disc is painted black, also the edge of the opening, thus when the eye has the proper position and fixation, the edge of the opening is seen concentric to the disc with a narrow ring of the gray of the perimeter arm between. The control afforded by this device is very sensitive. A very slight deviation of the position or fixation of the eye results in the complete or partial extinction of this ring at a point in the direction of the deviation. Like the mirror the device is mounted on a short pin or plug which can be inserted in the fixed tubular axle about which the perimeter arm rotates.

However, while the image in the mirror or the correct sighting of the fixation disc through its opening will indicate to the observer when the line of sight is normal to the surface of the fixation object at its central point, there are two important features in the correct adjustment of the eye over which these devices exercise no control: (*a*) the distance of the eye from the mirror; and (*b*) the agreement of the meridians of the field of vision as read on the perimeter with the meridians of the retina. In order that it may be known when the eye is at the correct distance from the perimeter arm, a light measuring rod 33 cm. in length is provided, to one end of which is fastened at right angles a small metal disc. In making the adjustment for distance one end of the rod is placed against the mirror at its center and the distance of the perimeter from the observer's eye is changed by means of the coarse screw adjustment to be described later, until the closed lid is just in contact with the metal disc.

Perhaps the simplest device for ensuring a constancy of relation of the meridians of the retina to the meridians of the field of vision as laid off by the perimeter arm, in other words for guarding against a slight tilting of the head to one side or the other, is a mouth bit. We have designed (Fig. 3) a very small and unobjectionable mouth bit of light wood to be changed for each observer, so shaped that it can not be bitten too far forward or back, and thus the distance of the eye from the mirror be changed, or too far to one side or the other. There seems, however, to be an insuperable prejudice against the use of a mouth bit by both the physician and the patient. We have designed, therefore, (Figs. 2, 4 and 5) a head rest which follows approximately the outlines of the forehead, side of the head and face, furnished with a suitably cupped chin rest, the height of which is adjustable. To provide for individual differences in shape and breadth of forehead, an adjustable forehead piece or band of thin spring steel extending well around to the side of the head is screwed at its central point to the forehead piece of the head rest. This forehead band is adjusted to fit foreheads of different shape and breadth by means of a set screw on either side near the two ends of the band. When the chin rest is adjusted to its proper height and the forehead band is made to fit the forehead, the patient's head is held comfortably in position and sufficiently rigid, it is believed, to satisfy the needs of office and clinic work. At least the probability of tilting the head to one side or the other, thus causing a disagreement of the meridians of the field of vision as indicated by the perimeter readings with the meridians of the retina is very greatly lessened, if not entirely obviated. This it will be remembered, is the especial purpose of the device, the other features of the control being taken care of in other ways.

In order quickly and conveniently to locate the patient's eye at the center of the perimeter system, three adjustments are provided: a rack and pinion to raise and lower the head, a second rack and pinion to shift the head to right or left, and a coarse screw adjustment to change the distance of the perimeter arm from the eye. In the process of getting the eye in position, the patient bites the mouth bit or adjusts the head in the head rest, brings the eye to the level of the fixation device (mirror or peepsight) with the first rack and pinion, and centers its image in the mirror with the second rack and pinion. The physician gets the correct distance of the perimeter arm from the patient's eye by means of the screw adjustment and the measuring rod already referred to. Once these adjustments are made for an eye, they need not be made again during the process of taking the fields for that eye; that is, the biting of the mouth bit or the return of the head to the head rest guarantees that the eye always returns to the same position for which the original adjustments were made.

A very great practical need in a clinic perimeter is a method of controlling fixation for patients who have a central scotoma or pathological blind area. With the eye properly adjusted for taking the fields these patients are not able to see a central fixation object. A device has been constructed (Fig. 3) for controlling the fixation of such patients in the following way. The perimeter arm is made to rotate about a hollow fixed axle. Into this fixed axle telescopes the stem at the end of which is the mirror used for patients with normal or approximately normal central vision, also a hollow stem carrying the device used to control the fixation of patients with a central deficiency. This latter device consists of four light arms or rods at right angles to each other curved to follow the arc of the perimeter arm, and of sufficient length to provide for all probable breadths of scotoma. Each of these arms carries a small stimulus the distance of which from the center of the field is adjustable. In adjusting the patient's eye by this device the physician looks through a small telescope contained in the hollow tubular axle and lines up the pupil of the patient's eye with the cross hair in the field of the telescope. When the patient's eye is observed to be in position, two or all four of the stimuli as may be desired are adjusted so that they can just be seen by the patient at the edges of the scotoma. These stimuli serve as the control of the patient's fixation, his instructions being so to direct the eye that all are visible.

This fixation device intended to control the fixation during the taking of the fields can be made serviceable for mapping the scotoma itself by adding 12 or more graduated arms equally spaced, making 16 in all, provided with stimuli similar to those already described. Then when the fixation is obtained by the adjustment of the four stimuli designed for that purpose, the further mapping of the scotoma is accomplished by moving the remaining 12 until they are on the edges of the scotoma. Or, if desired, the four stimuli designed for the control of the fixation may not be used for that purpose at all. The physician may watch the patient's eye through the telescope directing the fixation by means of the cross hairs, while all of the stimuli are moved into position on the edges of the blind area. This objective control of the fixation may also be used both in mapping the scotoma or in taking the fields when a central scotoma is present.

If desired the patient's eye can be rendered more visible by reflecting the light from the lamp directly on to it. This is provided for by placing a small oblong mirror of specular metal on the lamp arm at such a position and angle that the light received from the lamp will be reflected on to the iris. This mirror is hinged to the lamp arm and can be turned back against it when not in use. The hinge is provided with a stop so that when the mirror is turned down it can not go beyond the position required to reflect the light on to the eye. The visibility of the eye in the field of the telescope is considerably increased also by excluding as far as possible scattered light. This was done by putting the objective lens well back in the tube, painting the inside of the tube a mat black and placing on the end of the tube nearer the eye a short tubular hood which shields the opening from the direct rays from the lamp and admits only those reflected from the eye.

Another important need in a clinic perimeter is a method of giving the correct location and fixation to eyes suffering with high myopia. Eyes with myopia ranging from 8–20 diopters would have great difficulty in seeing a fixation object at a distance of 33 cm. Because of the grave pathological changes which take place in the retina and choroid of eyes suffering from high myopia, particularly in the region of the macula and nerve head, it is of great importance to be able to use both the perimeter and the tangent screen to be described later, in the examination of eyes in the more advanced stages of myopia. Three provisions have been made for this. (1) The mirror may be mounted on a rod sufficiently long to permit of its location at any point in the line of sight between the perimeter arm and the eye. This rod may be inserted in the tubular axle on which the perimeter arm rotates. (2) A peep-sight device is provided similar in principle to the one already described and so constructed that it may suffice as a fixation control for values of myopia ranging between 8 and 20 diopters. As shown in Fig. 4 this device has been so contructed that it can be used either for the normal or the myopic eye. That is, it is provided with two sets of fixation targets and viewing openings each at their proper distance from the eye. Both sets are hinged to the narrow steel bar which serves as their support so that either can be turned down out of the way when not in use. The diameter of the target for use with the normal eye is 8 mm.; the diameter of the viewing opening, 7 mm.; and the distance of the opening from the target, 9 cm. The diameter of the target for use with the myopic eye is 8 mm.; its distance from the perimeter arm is 16 cm. The diameter of the viewing opening is 6 mm. and its distance from the target is 9.5 cm. (3) The perimeter arm may be illuminated with two intensities of light—one carrying the fields well towards the periphery of the retina, the other giving limits narrow enough to fall within the corrected field of the glasses which are worn or may be worn by the patient. This feature provides also for the correction of high astigmatisms, the presence of which make field-taking annoying and uncertain as a diagnostic procedure. There are other advantages too of providing for the taking of fields at more than one intensity of illumination. (*a*) Because of the concentric arrangement of the fibers in the nerve trunk and their order of distribution in the retina, it may be of importance as a point of diagnosis to sample the responses of the retina at different degrees of excentricity. And (*b*) the low illumination fields are in general more sensitive to the influence of the pathological factors which cause the fields to have different breadths. This is largely due to the fact that low illumination fields are narrow fields. That is, sensitivity falls off gradually near the center of the retina, therefore smaller changes of sensitivity are required near the center of the retina to expand or contract the field. It is probably also due in part to the change produced in the state of the retina's sensitivity at low illumination.

The control of fixation for the presbyopic eye also presents a problem to the perimetrist. The eye with a high degree of presbyopia would have considerable difficulty in seeing with the necessary clearness a fixation object at a distance of 33 cm. By the use of the mirror as fixation control this distance is extended to 66 cm. The satisfactory use of the mirror requires that the image of the eye be seen fairly clearly, and an eye without power of accommodation even if there is no hyperopia for far seeing is approximately 1.50 diopters out of focus for an object at a distance of 66 cm. When 1.50 diopters out of focus the eye can not see its image in the mirror with a satisfactory degree of clearness. However the mirror can be used with a fair degree of satisfaction for lesser degrees of presbyopia.

We have three proposals to make for the control of fixation for the presbyopic eye. (1) The use of the mirror for the lesser degrees of hyperopia. (2) The use of an illumination sufficiently low to bring the color fields within the field of the correcting glasses. And (3) the use of a peep-sight or parallax fixation device (Figs. 3 and 4)[1] similar in principle to the one already described, with the target at a distance great enough to be seen with the necessary degree of clearness by the eye without power of accommodation. Sixty-six cm. has been chosen for this distance because (*a*) the target can be seen at 66 cm. with sufficient clearness to determine whether or not it is at the center of the viewing opening even though the eye is as much as 1.50 diopters out of focus; and (*b*) a greater distance presents difficulty as a matter of feasibility of construction. The device is provided also with a lens the distance of which from the target can be varied from its focal length to that which is needed to render the target clearly visible with 3 diopters or more of accommodation. With this lens at its focal length from the target the rays of light are rendered parallel and the apparent or optical distance of the target is 6 meters. By suitably changing the distance of the lens from the target the apparent distance of the target can be varied over a wide range. By means of this device therefore the apparent distance of the target can be adapted to patients having all possible degrees of presbyopia. However, as already stated, it is quite possible to secure a good control of fixation without the use of the optical attachment.

The fixation device consists of a tube 17.5 cm. long and 8 mm. in diameter, and a light metal rod fastened at one end of the tube and supporting vertically at the other a small plate of hard sheet aluminum 3 x 3 cm., at the center of which is a disc 4 mm. in diameter. This disc and plate serve as the

[1] As already stated the device for control of fixation in high degrees of myopia is also shown in Fig. 3 in position for use. When using either, the other should of course be removed.

fixation target. To use this device as a control of fixation the telescope is removed from the tubular axle of the perimeter and the tube of the fixation device is inserted in the axle to a depth of 4.5 cm. To control the depth of this insertion and thus the distance of the target from the eye, the tube is provided with a stop or shoulder 4.5 cm. from the end inserted. The tubular axle and its extension thus form the opening through which the fixation target is viewed, and when the eye is in position at the center of the perimeter system the target is seen to be centrally located in this opening. The target itself is painted black and the plate at the center of which it is mounted is painted in the same shade of gray as the remainder of the perimeter. The target receives its illumination from the lamp which illuminates the perimeter arm. An unobstructed passage of light from the lamp to the target is possible because of the distance of the target 33 cm. behind the perimeter arm. The lens which is used to vary the optical distance of the target from the eye is mounted in the end of the main tube, facing the target. This tube telescopes to the desired depth into a shorter tube at the end of which on a short supporting rod is the target.

The steadiness of fixation is greatly influenced by the method of giving the stimulation. One of the serious objections to a moving stimulus is the difficulty of holding a steady fixation while the object to be observed is moving. The alternative procedure is the use of a stationary stimulus. That is, the stimulus is placed at the desired point on the perimeter arm and covered with the preëxposure card. The observer takes his fixation and at a given signal the stimulus is exposed and re-covered. By this method of giving the stimulation more time is consumed but a much greater precision of result is possible. A compromise procedure is recommended. That is, the approximate location of the limit is determined with the moving stimulus and the exact location with the stationary stimulus. By this compromise but very little more time is required and there is no sacrifice of precision.

In order to provide for the mapping of the normal blind spot and for the quick detection and mapping of central and paracentral scotomata, it has been deemed advisable to add to the perimeter a tangent screen subtending a visual angle of 60 or more degrees. Provision is made so that this screen can be quickly and conveniently attached to the stimulus carriage and moved into position. The stimulus carriage and the tangent screen have at their exact center a circular opening equal in size to the cross section of the tubular aperture about which the perimeter arm rotates. Thus when the tangent screen is in position, *i.e.*, with its central point in the axis of rotation of the perimeter arm, the tubular opening in the perimeter is continued through to the front surface of the tangent screen. This both provides for the exact adjustment of the screen and permits of the convenient use with it of all of the fixation controls which we have described. Cards of white or black as may be desired with the fields laid off on a tangent scale are provided for the mapping of the area deficient in the light sense, and of grays of the brightness of the colors for mapping the color deficiencies. This is shown in Fig. 5. It is quite possible that the stimulus preëxposure cards used for the field-taking and the larger backgrounds or cards for the tangent screen could be made of thin hard sheet aluminum coated with a flat enamel paint of white, black or grays of the brightness of the colors. If this were done they could be kept clean by washing and the small paper disc of pigment color used as the stimulus for the field taking is all that would have to be renewed as the result of long periods of use. These paints can be bought in a flat black and a flat white. By using the two in the desired proportions the shades of gray desired may be obtained.

In our own work we have found it convenient to use the large screen just described for a quick survey of the field for scotomata and a smaller screen similar to the one used to carry the colored stimulus in field-taking for the actual detailed mapping of the scotomata and the normal blind spot. This screen was made considerably larger than the screens which serve as backgrounds for the colored stimulus in field-taking in order that it may serve for the mapping of large scotomata and pathologically enlarged blind spots. It is mounted in a carriage of its own and is shaped to take the curvature of the perimeter arm. This screen is intended not as a substitute for the larger screen but as a supplement in cases in which such a supplement is found to be convenient and desirable. This device has the following advantages over the large central tangent screen for the actual mapping of the blind areas. (1) It can be moved to any part of the field from the center out to 90° in any meridian and its center located at the center of the area to be mapped. When the screen is properly centered the mapping can be done as it is on any tangent screen. (2) Upon each screen is drawn 16 meridians, radially from the center of the screen. These meridians are finely graduated so that the limits of sensitivity in any meridian can be read off for the permanent record and transferred to properly planned maps at the convenience of the experimenter. (3) Blind areas are most easily and precisely mapped when the stimulus is made to follow lines radiating from the center of the area. Unless there are such guiding lines it is difficult to pass from within out or from without in consistently when determining the limits of the blind area or when checking up the location of a limit by a second or third determination. On a large fixed screen these lines would have to be drawn specially for each scotoma. (4) All of the evidence points towards the importance of mapping the blind areas with colored stimuli, particularly the Mariotte spot. It is highly important that the determinations for the different colors be made on backgrounds of the same brightness as the colors, as will be shown by blind-spot studies to be published later. It is much more feasible to arrange for this in case of the smaller movable fields than in case of a large fixed field. When the background for the smaller field is laid off in graduated radial lines it can serve for the mapping of many blind spots and scotomata before it need be replaced. In case of a large fixed screen, this superior adaptability and long service would not be possible. (5) With the movable smaller field it is more nearly possible to map all blind areas under the conditions of equal illumination which obtain in the perimetry of the color fields than it is in the use of the large fixed screen. In the latter case the illumination at which the mapping is done in one part of the field may be quite appreciably different from that at which it is done in another part of the field. The control of illumination, while not so important for mapping the blind areas to the light sense stimuli, is very important in the mapping of the blind areas to color.

With the controls provided in the perimeter recommended, a careful worker can without difficulty reproduce the limits of sensitivity within 1 or 2 degrees.

SOLVING INTERCORRELATIONS BY POLAR COÖRDINATES

BY HERBERT A. TOOPS

Institute of Educational Research, Teachers College

In another place[1] the author has shown how intercorrelations of a number of variables may be efficiently solved on the adding machine by means of transmutation of gross scores into class numbers with the aid of standard grouping tables. The method of solving intercorrelations here presented is efficient in point of speed, although not quite so efficient, and in addition possesses the advantage of yielding a scattered diagram of the data. With reasonable care, computational errors are unlikely to occur. This advantage is secured by the numerous checks obtainable in the plotting method [2] used by the author.

In the first-mentioned method of solving intercorrelations, the transmuted class scores of the Y-variable are borne into the midst of the data sheet by means of carbon copy strips in order that the XY products may be readily computed mentally. This function is performed in the new polar method by a 6-foot radius bar, H, (Fig. 1) which bears the Y-marginal classification successively into the bodies of a number of polar coördinate correlation plots arranged radially about a common center, C. Each of the plots is a small arc-sector of a circle 12 feet in outside diameter. The arrangement of the charts is shown schematically in Fig. 1. This method employs the transmuted class scores derived by the use of grouping tables.[1] The equation employed is:

$$r = \frac{[\Sigma X^2 + \Sigma Y^2 - \Sigma(X - Y)^2]\frac{N}{2} - \Sigma X \cdot \Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\,\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}\,, \qquad (1)$$

[1] Toops, H. A., 'Computing Intercorrelations of Tests on the Adding Machine.' To appear shortly.

[2] Toops, H. A., 'Eliminating the Pitfalls in Solving Correlation; A Printed Correlation Form,' Jour. of Exp. Psych., Vol. 4, No. 6, 1921, pp. 434–447.

in which X and Y are steps of the polar correlation plot, or are transmuted class scores derived from the grouping tables. These transmuted class scores are the classes under which the gross measures would be filed if there were not over eighteen classes covering the total range of variation of the variable.

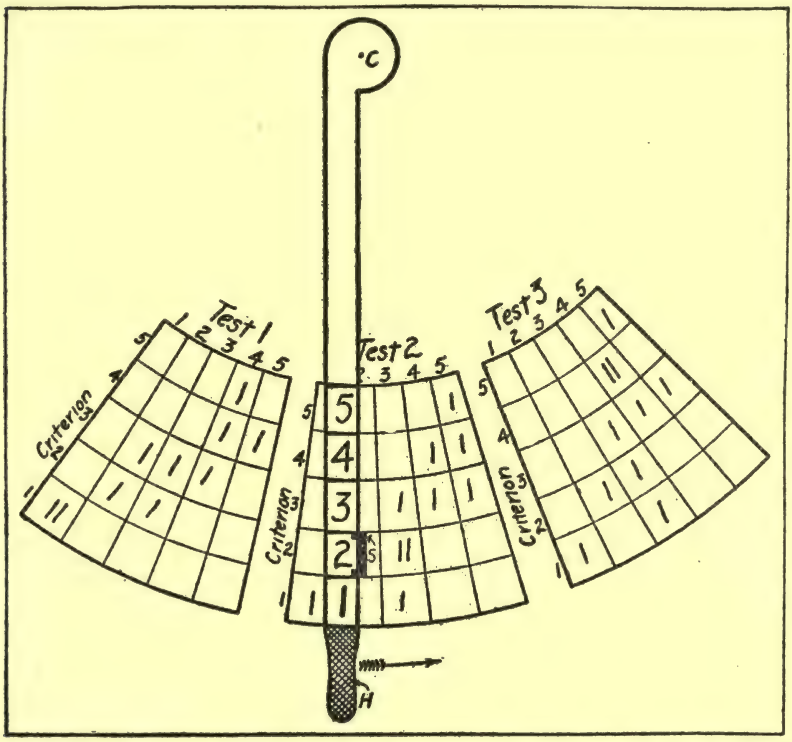


FIG. 1. Schematic diagram of the polar correlation plotting method. This shows how eight to ten polar correlation sheets may be thumbtacked in a semi-circle about a common center, C, and all be plotted simultaneously. The slider, S, is set at the score common to all charts (criterion score 2, above); thereupon the given subject's scores in the various tests are plotted opposite the slider in the proper columns of each sheet in turn.

It will be seen that (Fig. 2) with a twelve-foot diameter of circle, the small portion of arc required for a scatter diagram of eighteen classes is such as to make the horizontal arcs almost straight lines while the radial lines are almost parallel. The coördinates are thus practically rectangular coördinates.

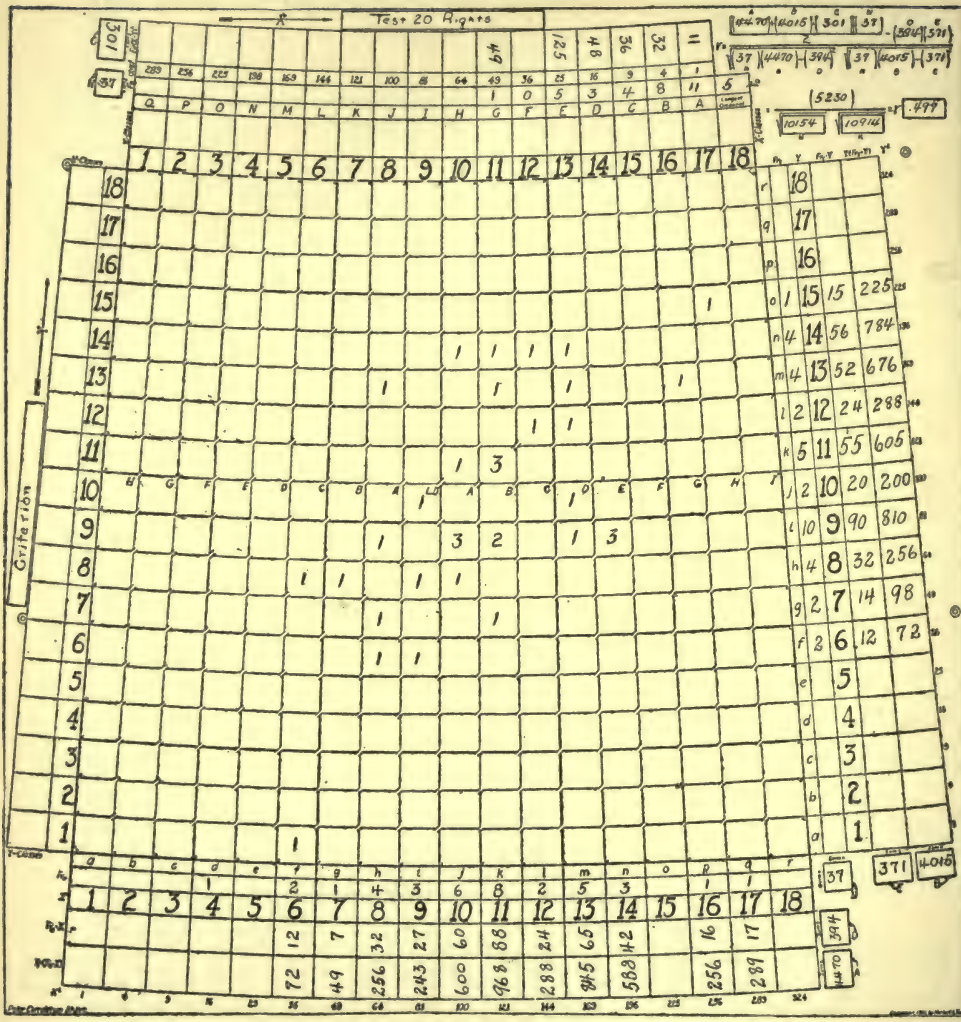


FIG. 2. The computation of r by the polar coördinate sheet method.

In use, a number of the correlation plots are thumbtacked to the top of a soft pine table arranged so that all the radial lines have a common center.[1] The slider, *S*, on the radial bar is set to indicate the *Y*-class common to the n' correlations being plotted, of Person 1 in the variable, Criterion, say, with which the intercorrelations are at the time being computed. It is evident then that all the check marks for Person 1 on the various tests (plots) will fall on the same horizontal circle indicated by the slider. Experimenter *A* calls the transmuted classes of three tests, Test 1, Test 2, Test 3, to Experimenter *B*, who swings the radius bar to the right to the proper *X*-class successively of the three tests, recording in each a tally mark opposite the slider; *A* then reads the transmuted classes of the next three tests of the same person, and so on until *B* has reached the end of the swing of the bar, which is at the last test of the series of intercorrelations. Then the same procedure is gone through with for Person 2, Person 3, etc., in turn until all *N* cases have thus been plotted. The long bar, if made of thin material, will be found to respond very readily to rapid changes of position. It may readily be lifted entirely from the chart if not pivoted too tightly at the center.

The plotting compartments are sufficiently large, approximately 7/16 inches square at the center of the chart, to prevent plotting errors if reasonable care is taken. The printed classes at both the top and bottom sides of the chart, as well as the heavy vertical lines after every third compartment make errors in the vertical direction easy to avoid if *B* calls back his entries to *A* as rapidly as he writes down the tally marks; errors will never be made in the horizontal direction if the slider is set correctly each time. If errors are made in any plotting, these will be readily detected by comparing the obtained marginal frequency in each case with the standard marginal frequency as soon as a standard (correct) marginal frequency is obtained. (It is well worth the extra time to obtain a correct list of all marginal frequencies by tabulation at the outset.)

[1] The author is revising the chart so as to be used with a revolving table, the bar, *H*, being stationary and charts moving while the plotter remains seated at one point opposite *H*.

The A-diagonal frequencies for this chart are found by adding all frequencies in *both* A-diagonal rows, the eye being guided in this process by the narrow lanes[1] leading from one diagonal compartment to another; the B-diagonal frequencies by adding all frequencies in *both* B-diagonal rows, and so on.

The computation of r follows the procedure outlined in the diagram below (Fig. 3).

Figure 3 is a systematic job analysis solution of the equation,

$$r = \frac{[A + B - C]\,\overline{\frac{N}{2}} - D \cdot E}{\sqrt{NA - D^2}\,\sqrt{NB - E^2}}, \tag{2}$$

arranged so that all similar operations are done at one time. Only the intercorrelations of four variables are shown.

The procedures involved in Fig. 3 are as follows:

1. Enter the ΣX^2-values by variable numbers at the top and bottom of the chart in the designated rows and repeat as ΣY^2 in the left-hand column, by variable numbers, and check. These are systematically added, the ordinate plus abscissa of a given compartment (x, y) in which the solution of r_{xy} is being sought, thus going to make up for the given (x, y) large compartment the $(\Sigma X^2 + \Sigma Y^2)$-values $= A + B$, or the first entry of each large compartment. These sums check for symmetry of rows and columns about the diagonals.

2. Enter the C-values, $\Sigma(X - Y)^2$, as the second entry of each large compartment throughout the table independently from the original data *twice*—once in the upper diagonal half of the table and once in the lower—and then check the entries for symmetry of rows and columns.

3. Subtract C from $A + B$ in all compartments. Check for symmetry of rows and columns about the diagonals to prove the correctness of this operation.

4. Multiply $(A + B - C)$, already obtained in Procedure 3, by $N/2$ if N is an even number; if N is odd, multiply first by N and then enter half of the product (always an integral quantity) as the fourth entry of each large compartment.

[1] The writer is indebted to Dr. A. S. Otis for the suggestion of this helpful device.

Var. No.			I	1	2	3	
Var. No.	$\Sigma X = D$		371	394	61	405	D_y^2
	$\Sigma X^2 = A$		4015	4470	127	4607	
I	371 ΣY $= E$	4015 ΣY^2 $= B$	1.0000	8485 301 8184 151404 146174 5230 .497	4142 2940 1202 22237 22631 − 394 − .121	8622 324 8298 153513 150255 3258 .389	10914
1	394	4470	8485 301 8184 151404 146174 5230 .497	1.0000	4597 3315 1282 23717 24034 − 317 − .101	9077 405 8672 160432 159570 862 .107	10154
2	61	127	4142 2940 1202 22237 22631 − 394 − .121	4597 3315 1282 23717 24034 − 317 − .101	1.0000	4734 3402 1332 24642 24705 − 63 − .025	978
3	405	4607	8622 324 8298 153513 150255 3258 .389	9077 405 8672 160432 159570 862 .107	4734 3402 1332 24642 24705 − 63 − .025	1.0000	6434
ΣX^2 $N \cdot \Sigma X^2$ $(\Sigma X)^2$ $N \cdot \Sigma X^2 - (\Sigma X)^2$ $= D_x^2$			4015 148555 137641 10914	4470 165390 155236 10154	127 4699 3721 978	4607 170459 164025 6434	

The first entry of a large compartment is $\Sigma X^2 + \Sigma Y^2 = A + B$
" second " " " " " " $\Sigma(X - Y)^2 = C$
" third " " " " " " $A + B - C$
" fourth " " " " " " $N/2(A + B - C)$
" fifth " " " " " " $\Sigma X \cdot \Sigma Y = D \cdot E$
" sixth " " " " " " simplified numerator of r

$$= \frac{N}{2}(A + B - C) - D \cdot E$$

" seventh " " " " " " r (solved by slide rule)

$$= \frac{\text{Numerator}_{xy}}{\sqrt{D_x^2}\,\sqrt{D_y^2}}$$

FIG. 3. The systematic solution of equation (1) in solving intercorrelations. $N = 37$ persons.

(Note: $A + B - C$ *must always be* an *even* number and $(A + B - C)(N/2)$ an integral number; if this is not the case, C is undoubtedly in error and should be corrected from the original plots. It is assumed that ΣX^2 has been recomputed several times so that there can be no doubt as to the correctness of these values.) These remainders check for symmetry.

5. Enter (ΣX)-values $(= \Sigma Y)$ by variable numbers at left and top of the chart in the proper compartments of the column and row, and check carefully. Multiply these in systematic order, *i.e.*, by rows across the page, *ordinate times abscissa* throughout, to get the $\Sigma X \cdot \Sigma Y$ products for a given compartment, the fifth entry of a large compartment. These products check for symmetry.

6. Subtract the products of Procedure 5 systematically by rows across the page from the result of Procedure 4. These remainders are the simplified numerators of the correlation coefficients. Check for symmetry of rows and columns as a check upon the accuracy of all the preceding work; note carefully whether algebraic signs are correct.

7. Compute the simplified squares of the denominators in systematic fashion at the foot of the table, obtaining $N\Sigma X^2 - (\Sigma X)^2 = D_x^2 = D_y^2$. Enter these values again as D_y^2, in the extreme right-hand column by variable numbers, and check carefully.

8. Solve for r systematically on the slide rule, using the formula

$$r = \frac{\text{Numerator}_{xy}}{\sqrt{D_x^2}\ \sqrt{D_y^2}}, \tag{3}$$

in which D_x^2 and D_y^2 are the simplified denominators respectively of the D_y^2 row and D_x^2 column (result of Procedure 7), denominator coördinates of the xy-compartment, the r of which is being computed. This is to be done at one continuous operation on the slide rule. Using a calculating machine, it is an advantage to compute $1/\sqrt{NA - D^2}$, whereupon the solution for r becomes a multiplication problem, the numerators of each row being first multiplied by the denominator-reciprocal of the row, and then these products by the denominator reciprocal of the column systematically. Using a calculating machine, some of the above procedures may be dropped out by performing two or more of them at one time. Two people and a calculating machine can solve the equations of several hundred correlation coefficients in a few hours when making use of the above job analysis procedure. The *r*'s finally obtained must check for symmetry of rows and columns in sign, decimal point, and absolute magnitude. To secure maximum speed without confusion, the operations should be done in exactly the order outlined above.

NOTE ON ELECTRIC COUNTERS[1]

BY W. R. MILES

Nutrition Laboratory of the Carnegie Institution of Washington, Boston, Mass.

The automatic counter is a useful accessory in combination with various laboratory apparatus. It is frequently desirable to operate these counters electrically. Several pieces of apparatus recently seen by the writer show that each designer has solved for himself this rather troublesome problem. Probably in many such cases a large saving in time and expense could be made by using the telephone 'p. b. x. message registers,' commercial electric counters which seem to have escaped the notice of most laboratory workers.[2]

The 'message registers' are very well made and are compact. (See Fig. 1.) The counter and actuating electro-

FIG. 1.

magnet are rigidly mounted together in the same frame. The device is arranged for operation in the horizontal position. The pawl connected with the armature and the locking pawl are both kept in position by gravity, but spring tension may easily be made to do this if it is desired to operate the counter in a vertical position. The entire mechanism is rugged and very accessible for adjustment or alteration. The outside dimensions are about 4 cm. wide by 3 cm. high and 15 cm. long. Total weight is less than 0.5 kg. The digits are 5 mm. high, black on white-metal and very legible. There are four counting wheels. These cannot be set back to zero.

The armature is of the hinge type and acts against a spring.

[1] This note has been prepared at the request of scientific colleagues.

[2] Manufactured by Western Electric Co., Inc., Boston, Mass. Prices range from about 5 to 10 dollars, according to production of the different types.

The maximum movement of its free edge is less than 5 mm. The manufacturers state "that these registers may be operated 120 times per minute and the contact must be in a closed position for a quarter of a second and on the open position one quarter of a second." This is a very conservative statement. Using the registers at rated current we have found them to record upwards of 300 contacts in a minute. If the current is increased contacts as short as .04 to .05 sec. can be made to give positive operation. The spring requires nearly .07 sec. to return the armature. Therefore it is possible with this commercial apparatus to register more than 500 contacts per minute for short periods of operation. It is quite impossible to turn up more than one count with one movement of the armature.

There are a variety of message registers from which one may select. Those most likely to be useful are as follows:

Type	Winding	Rated Ohms	Operating	Production
5-B.........	Single	0.27	1.1 amps.	Special
5-D.........	Single	34.0	.117 amp.	Special
5-H.........	Single	0.27	1.4 amps.	Standard
5-L.........	Inner Outer	37.5 463.	25.5 volts	Standard
5-M.........	Single	280.0	.036 amp.	Obsolete
5-R.........	Inner Outer	120.0 600.0	.165 amp.	Obsolete
5-S.........	Single	5.0	.33 amp.	Standard

Those marked 'obsolete' or 'special' can usually be obtained as there are telephone systems which require them. Most of these mentioned types have a pair of contacts which is closed when the armature is fully moved and a count actually registered. These contacts are serviceable to operate a signal circuit or to provide an auxiliary graphic record which can be inspected for 'chatter,' or failure to operate, if the event counted is quite regular as revolutions of a wheel, number of steps or respirations. We have found this electric counter fast enough for most purposes and a reliable recorder.[1]

[1] See Benedict, Miles, Roth, and Smith, 'Human vitality, etc.,' Carnegie Inst. Washington, Pub. 280, 1919, p. 127.

79

Journal of

Experimental Psychology

VOL. V., NO. 2 APRIL, 1922

AN INDIVIDUAL CURVE OF LEARNING: A STUDY IN TYPEWRITING

BY BLANCHE M. TOWNE

Interest in the curve of learning seems to be increasing from year to year, with the result that an even larger number of studies are being published.

In 1897 Bryan and Harter (1) made their much-discussed studies on the 'Physiology and Psychology of the Telegraphic Language' and followed it two years later with their theory concerning the acquisition of a hierarchy of habits. In 1903 Swift (9) made an extensive study into the acquisition of a physical habit, a mental habit, and a complex of the two. These and other men working along similar lines found that the typical curve of learning was hyperbolic in form, increasing rapidly at first and then flattening out as practice was continued. In 1919 Chapman (7) published the results of an investigation of twenty students who were learning typewriting. He found that there was no curve to which individuals approximated, even though they reached the same degree of skill after the same number of practices; but that for many there was a rapid initial rise. It will be necessary to have the data from several hundred cases before we can really determine the characteristics of a learning curve. In order to secure from such data results, comparable in every detail, one must make the practice material, the practice conditions, etc., similar for all.

It was a desire to know into what sort of a curve my own learning process would resolve itself and how I should proceed that led me to keep the data from the first one hundred practices on a Corona typewriter, and later to continue the practice for seventy-six days longer, making a second series of records. I had had no previous experience. Instructions were procured from a teacher in typewriting and the method outlined in 'Rational Typewriting' by Cutler and Sorelle was followed in detail. All but the last or twenty-fifth lesson, were completed in the first hundred days of practice. The lessons were on graded material. After six months and a half had elapsed, during which time I had used the typewriter but very little, I again began practicing and continued for seventy-six days. Seventy days were spent on the first hundred and eighteen pages of 'Essays in Exposition' by Kurts, Cory, Blanchard, and MacMinn, dealing with material of approximately the same difficulty. There were extracts from three essays by John Stewart Mill, three by John Henry Newman, and a part of one by Matthew Arnold. As these were manifestly more difficult than the material which I had ceased to use at the end of the first hundred days, for purposes of comparison, the last six days were spent on material of the same degree of difficulty as that of the hundredth lesson. Forty-five minutes, between seven and eight o'clock in the morning, were devoted to the work. The original plan was to make the practice days consecutive. This was made impossible by a four-day vacation, three long breaks occasioned by necessary absence from the city, and four Sundays, during the first hundred days of practice; and one break of two weeks and one of three during the last seventy-six days of practice. No attempt was made to work at maximum speed, but rather to proceed with the same amount of effort that would ordinarily be put forth upon any work. Introspections were written down almost daily. No collateral reading was done during the first practices, nor was any scoring done. Practically all the reading was done before the last series, and the results were tabulated from time to time.

The Curve

The correctness of the curve depends upon the uniformity of the practice periods, the time of day when they were taken, the amount of work done during the periods just preceding them, etc. These were all rigidly watched and care was taken to avoid even slight variation.

The first four days were used in practicing the letters in the order in which they appear on the keyboard. The number of letters written was large and the errors few. This work served, therefore, merely to establish a 'set' for typewriting. The real work began on the fifth day, when words were first written. So, for the purpose of our curve, we shall consider the fifth day of our practice as the first day.

The curves described in most publications are based on the same materials, or materials of the same difficulty, written at regular intervals. Our use of material of increasing difficulty in the earlier series, makes it impossible to make any direct comparisons with such data. A study of it may, however, throw some light on our curve.

In the first curve of accomplishment we counted off five correct letters for each error. The space between the letters and the shift between the lines counted as one letter, while any letter, figure, or mark requiring the use of the shift-key received the value of two.

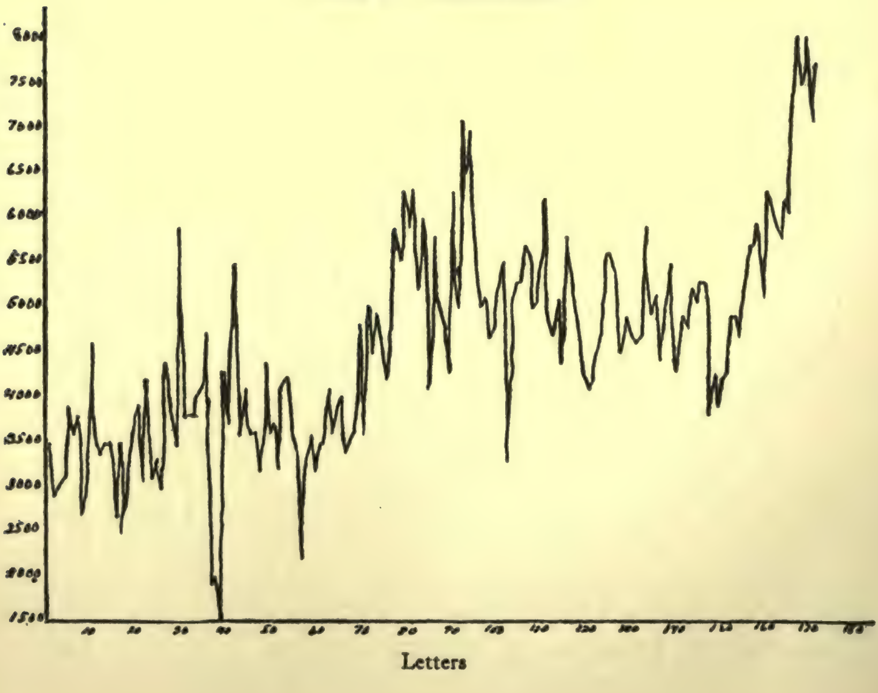
The curve has no characteristic shape. It rises slowly during the first forty-five days, seems to remain at a level for the next twelve days, and then rises rapidly again to the end of the first hundred days of practice, except during eleven days, when the curve seems to remain at the same level, a very high one. At the beginning of the second series of practices, the curve drops to a much lower level and maintains it for about fifty days. Then there is a decided drop for five days, followed by a marked rise when the level of the last plateau in the first series is reached. The last six days of practice reached a level not attained at any other time during the whole practice. The learning was not uniform but proceeded irregularly, attaining a high score only to fall back for a long or short period before attaining it or passing it again. Swift found that the period between high scores was short but later investigations seem to show that it may be long. Typewriting is such a complex process and affected by so many different varying conditions that it seems only correct to expect this. Physical and mental condition, difficulty of material, use of new parts of the machine, speeding up beyond the average, etc., could scarcely interfere and aid in the same degree each succeeding time.

Explanation of the Curve of Learning or Accomplishment

The rise on the fifth day was due to the repeated practice of a review exercise with very few errors. On the next day the drop was due to the new material's causing more errors, which occurred in groups and which left a very confused feeling. The sudden rise on the tenth day was due to the use of old material with few mistakes and to a feeling of exhilaration after having been detained from practice for a few days. A drop on the next day was caused by new material and the writing of many letters with many mistakes. The use of the shift key kept the curve down for the next few practices. Slow, painstaking, and careful work seems to account for the decided rise on the nineteenth, twentieth, and twenty-first days. During the next few practices I found myself confused whenever I tried to watch the copy. Up to that time I had been looking into space or closing my eyes instead of watching the writing. The movement of the paper and the little letters flying up and down annoyed me. Besides, I found myself stealing glances at the keyboard. This irritated me and I felt that I was making no gain in the work and that it had become a drudgery. The very low score made on the twenty-first day may be accounted for, in part at least, by the frequent use of the shift key for numerals. The next day the score was very little better, and the work was very exhausting. The same feeling continued during the nine days that I worked on this lesson. This probably came from the strain on the muscles. A feeling of calmness followed, and I worked slowly and carefully, reaching a higher score than previously except for the one attained on the thirtieth day. Then the beginning of a new lesson brought the score down again. I began to feel some pleasure in the work, and the weather became cooler. The more careful application which resulted soon brought the number of errors down to one for the whole practice. However, the slow work reduced the number of letters and, in consequence, the curve dropped to a new low level. A new lesson requiring new parts of the machine kept it down for two more days. Up to this time there had been no plateaus at all. But when the curve rose again it remained at a level for about twelve days. This period exactly coincides with the time spent on lessons ten and eleven in the typewriting book. These contain the first whole paragraph and the first whole page to be written without repetition. The introspections reveal almost daily how great a difficulty this was. I was also emphasizing speed during this period. There is not one mention of pleasure and once a great dislike of the copy is expressed. This period of stress and strain left a feeling of discouragement, which persisted until the twenty-seventh lesson. Each day I had to drive myself to the practice. Even a perfect copy brought no desire to start another lesson. It was, however, just at this time that the score went up very rapidly. The second plateau, from the seventy-seventh day to the ninety-second, maintained a clear and distinct level, although a higher one than had hitherto been reached. Both decrease in the number of errors and increase in the number of letters written combine to increase this high score. The material, while in reality no easier, seemed so. The exercises were short and of more interesting content. The introspections contain many allusions to growing confidence and pleasurable feelings. For the first time I had actually become happy in the work. During the last few practices of the first series, the curve again started upward, accompanied by pleasurable feelings and self-assurance.

In the second series the drop, which we find is to be expected because of the much more difficult character of the material, both as to spelling, punctuation, and thought content, which, because of a habit of long standing, I could not divorce from the copying. The first fifty days were on the same level. I went to the task each day as one of my regular duties. I felt no false elation at the task nor any great depression that I was not gaining. It all seemed a part of the day's work. Occasionally, I tried to spurt up but the gain would be of very little significance. Then I decided to put the emphasis on accuracy. The result of this was a decrease in the number of errors and also a decrease in the number of letters written, which, of course, caused a fall in the curve. Up to this time the score was being reckoned by dividing the number of letters written by the number of errors. As this seemed too great a price to pay for an error, it was decided to change the plan to the one which we have described above. Then I found my attitude changing from one of a certain watchful attitude toward errors to one which was a little more relaxed and which emphasized speed. This accounts for the following rise. Each day I felt a little freer than on the previous day, and, therefore, could speed up a little more. I, also, feel that this accounts for the higher level which was reached; a level, not very much higher than the preceding plateau but about proportionate to the relaxation from strain

Curve of Accomplishment



and control which was felt before. During the next six days of practice the material used was of the same difficulty as that used at the end of the first series. The result was what might have been expected. The curve rose a little higher than at any other time during the practice, and maintained that height during the period. This shows how little real gain was made during the previous seventy days of work, and indicates my level for the present, at least, in work of that difficulty. The level last reached on the difficult material is the one which shows the level of my capacity for such material.

Lower and Higher Order of Habits

At the outset the lower order habits dominated. Locating the desired key, placing the correct finger upon the key, and then striking it required individual attention. The letters 'a,' 's,' and 'l' soon needed no attention, while 'g,' 'b' and 'y' were still causing trouble on the sixtieth day of practice. 'X' was the one letter which had not become automatic on the last or one hundred seventy-sixth day of practice. 'As,' 'tion,' and 'es' were completely automatic before fully one half of the single letters had become so. By the end of the sixtieth day of practice, short, familiar words such as, 'and,' 'the,' and 'we' caused no trouble at all. The first group of small words were written automatically about the fortieth day. all lower order habits were steadily taking care of themselves in an ever increasing degree. Letters, syllables, words, and groups of words were dropping into the background. In the last few practices of the first series the thought of whole passages would be dominant and the lower order habits seemed not to interfere. These passages became longer during the second series, but strange combinations of letters in foreign or unfamiliar words brought the lower order of habits once more to the front.

Such observations as these completely disprove the theory that the lower order habits are completed before the higher orders are begun, or, indeed, that there is any regular order. Swift (9), Batson (5), and later investigators have reached the same conclusions.

Plateaus

Nature provides resting times in all her plans of life-processes. Therefore, it is not surprising that the learning process shows real resting places or plateaus. These may be interpreted as slumps in enthusiasm. Interest and attention wane at such times, energy seems entirely lacking, and the work is described as very tedious to the learner. A plateau often occurs after excessive effort or when the physical or mental condition is not good.

Bryan and Harter (2) thought that a plateau was a slowing up of the processes in order that certain lower order habits which were approaching their maximum development. could become sufficiently automatic to leave the attention free to attack higher orders of habits. The length of a plateau would thus measure the difficulty of making these lower order habits sufficiently automatic.

Book (3) thought that in learning typewriting there were two levels at which plateaus occurred. The first, so he believed, occurred when the speed was about one hundred ten strokes per minute and the second when it was one hundred forty-five strokes.

Chapman (7) found only a few short plateaus which he considered genuine. In the twenty cases that he studied, these plateaus did not occur at fixed places which were the same for each individual, but varied greatly. Neither Hill, Ryall, and Thorndike (8) nor Bradford (4) found any plateaus. They feel there is no need for these and that it is not necessary for habits to be developed and perfected in such a monotonous way, and that they do not 'suddenly shoot together into new combinations' and thus cause the curve to rise rapidly after a plateau. Peterson feels that the type of a curve drawn by some investigators does away with plateaus which would otherwise show up.

Swift (9) found that a small amount of effort to spurt up is helpful if it is not too severe. Overstrain exhausts the learner and hinders progress by bringing into consciousness those processes which help most when kept in the background.

The three plateaus in the curve which I have drawn from my practices are each quite different in their characteristics. The first one is undoubtedly caused by the increased difficulty in the material. This did cause a big slump in enthusiasm. Interest and attention waned but the continued attempts to speed up do not indicate that energy was diminishing. It did not occur after excessive effort or when the physical or mental condition was not good. On the other hand the condition is reported as being better than it was at the beginning of the work. Therefore, this plateau could scarcely be called a true one in the sense that it is spoken of when it arises during work on material of the same difficulty.

The second is more nearly a true plateau than the first. The material was not markedly more difficult than that which had just preceded it, but came after what seemed to the learner a very strenuous period. Much effort had been used just to keep up the daily practices. The exercises themselves seemed long and tedious. At the beginning of the plateau the reaction to them became quite different. It may be likened to the pleasure of basking in the warm, spring sunshine after the winter's cold. Interest and attention were undoubtedly at a lower level during the eleven practices, for a high degree of energy and alertness is not compatible with the state of mind just described. Habits, too, were probably becoming automatized. In the practices that followed directly after the plateau they seem to have dropped away to such a degree that whole passages could be written with little or no attention directed to the actual running of the typewriter. Then, too, interrelations were fast becoming established. Near the end of the plateau the machine came to seem like a tool with but a single purpose instead of a composite with several ends to be attained before the written sheet could be produced. At no time during the period was I aware that the work was at a standstill. On the contrary, I thought that I was making more progress than at any time previous. Neither was I aware that my work was improving during the sudden rise which just preceded the plateau when it seemed that no progress was being made at all. Many investigators have noted that the judgment of the learner is often very erratic. Contrary to the usual idea, the work at the time of the plateau was not tedious, but very enjoyable. But the variation in the material makes the comparison with other's results rather uncertain as to value. The level of my plateaus are not comparable with Book's findings because of the difference in the way the data was taken. They are more nearly comparable with those of Chapman who holds that individuals have their own individual levels for resting places. Possibly more effort could have been put forth during this second plateau so that the curve would have continued to rise had there been consciousness on the part of the learner that the work was at a standstill. Personally I find no regret that the first prolonged pleasure in the work was not interrupted. I believe that I was almost too self-satisfied to have been spurred on by any interest aroused by an outside force.

The third plateau occurred because I had reached the level of my ability, for work of the grade of difficulty used. Although the work of the last six days of practice on this material showed a somewhat higher level I feel certain that the whole plateau would have assumed this same level had the restraint been thrown off at the beginning of the plateau. Therefore, in the true sense of the word, this was no more a real plateau than was the first one.

Effect of Physical and Mental Condition upon the Curve of Accomplishment

Physical and mental condition apparently played a part in our results. Hot, sultry days seem to cause irritation, make the work unpleasant, and keep the curve down. A cool day usually brought the curve up and gave added enthusiasm for the work, all other things being equal. The physical exhaustion of the muscles kept the score down in several cases, especially using the shift-keys for the first time. The high points of the first sixty practices were accompanied by pleasurable feelings as also was the second plateau which maintained a much higher level than had been reached at any previous time. The pleasure which occurred at this time when the learning was making no apparent progress may not seem so incompatible with the prevalent idea of displeasure during the plateau period when we remember that our plateau did attain such a high level.

Book's statement, that it is intense effort that educates, may be substantiated at least once during our practices. This was during the rise which just precedes the second plateau.

The practices of the second series were very little affected by physical or mental conditions, although the practices were continued during the same hot sultry months as were those of the first series. This was probably because so much of the work had become reduced to habit that other things had little or no effect upon the work. Intense effort did not raise the level of the third plateau, because, in a sense, the learning was complete. I had reached my level.

Periods of Absence

Hill, Ryall, and Thorndike found that the Sunday rest had no influence on the curve. A few learners gained and others lost by it.

The periods of absence or breaks in the regularity of practice occurred as follows:

After the 8th day of practice for 3 days.
After the 26th day of practice for 4 days.
After the 57th day of practice for 1 day.
After the 63d day of practice for 1 day.
After the 71st day of practice for 1 day.
After the 75th day of practice for 1 day.
After the 77th day of practice for 29 days.
After the 89th day of practice for 20 days.
After the 96th day of practice for 6 and ½ months.
After the 136th day of practice for 21 days.
After the 154th day of practice for 11 days.

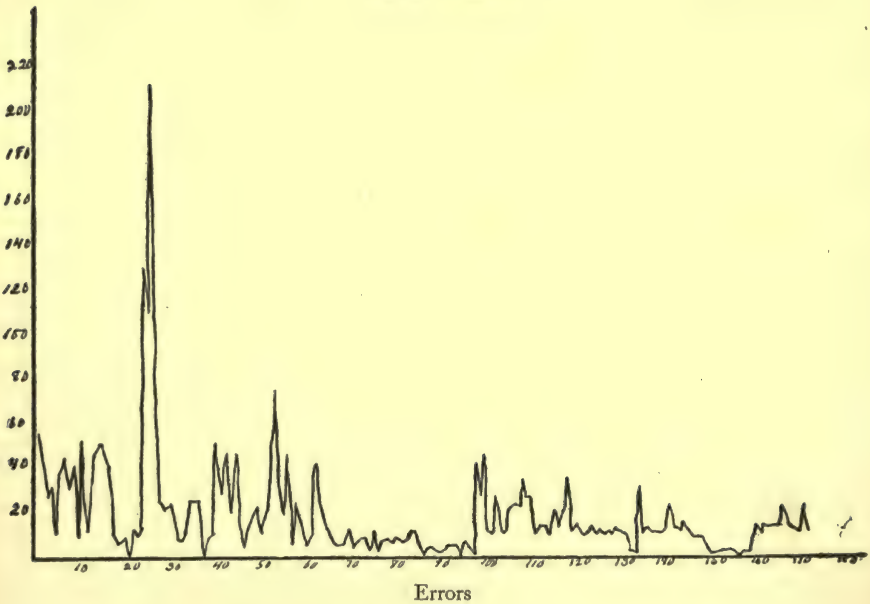
There was no perceptible loss or gain in the accomplishment even after long breaks of six and a half months, when the curve kept the level it assumed at the beginning of the series.

The Curve of Errors

The general shape of an error curve shows a rapid decrease at first. This becomes much more gradual as practice continues. It is not a smooth line, but is full of fluctuations.

Bergström found that there were fewer errors and a smaller number of keys hit, letters said, etc., when the rate was slower; but Thurston found no relation between accuracy and the rate of learning. Speeding up over the average is sure to result in a very large number of errors. One error is

Curve of Error



usually followed by several because the processes thus blocked result in a confusion. It, then, requires a little time to regain poise and composure. In his study on 'Psychomotor Mechanisms of Typewriting' Wells, also, found that there was a conscious or unconscious slowing up of all processes before errors.

My curve, based on material of increasing difficulty, shows only a very gradual decline toward a vanishing point, flattening more perceptibly after the sixtieth practice. The curve, beginning with the second series, where the material was of the same difficulty, remained at the same level until emphasis was put upon accuracy, when it decreased toward a vanishing point. As soon as the emphasis was removed it again sought the level it had maintained through the practices of this series. I, therefore, seem to have a level for errors which is not the minimum possible, but which could be reduced at the cost of speed. The plateaus are practically coincident with the plateaus in the curve of accomplishment. One error was usually followed by several, especially during the first sixty practices, when most of the errors occurred. This was less true in the second series, as the knowledge which I had gained as to the psychical condition at this time usually helped me to stop and get possession of myself before continuing.

In general, the errors were caused by (1) new or unusual material, as groups of words, sentences, poetry, business contracts, etc.; (2) hard material which required the frequent use of shift-keys; (3) the use of new parts of the machine and (4) the speeding up above the general average number of letters per day for the lessons that had just preceded.

The per cent. of error for the first week was 11.69 and for the twentieth 1.35, a difference of ten. Thurston found that his group started with 58 per cent. of error and decreased to 36 per cent. at the end of twenty weeks. This is a difference of twenty-two. Here, as elsewhere, we are comparing two things which are not essentially alike.

Conclusions

1. Habits of all orders make progress simultaneously, some in a less and some in a greater degree.

2. A single, real plateau which was short and genuine occurred in the practice after the seventy-seventh day. The others were not real.

3. Intense effort, consciously or unconsciously applied, helps, all other things being equal.

4. The days of greatest accomplishment were generally accompanied by a pleasant feeling, except after a hard struggle against great odds, when the unpleasant feeling held over for a long time.

5. Breaks in the regularity of the practices seem to cause neither gain or loss in the net result.

6. The error curve decreased gradually for the first sixty practices and then proceeded more slowly toward the vanishing point.

7. There is a level in both the curve of accomplishment and the curve of error which is normal for me.

8. Errors were caused, (*a*) by new material, (*b*) by use of new parts of the machine, (*c*) by extensive use of the shift-keys, and (*d*) by speeding up over the general average.

9. Physical and mental conditions have an effect upon the results produced, the better the condition, the greater the accomplishment.

10. In graded material, work which is much harder than that just preceding, causes a slump in enthusiasm and a dislike for the subject in hand.

Bibliography

1. Bryan and Harter. Studies in the Physiology and Psychology of the Telegraphic Language. *Psych. Rev.*, 1897, p. 27.
2. Bryan and Harter. Studies on the Telegraphic Language, same, 1899, p. 345.
3. Book. Study of Typewriting. University of Montana Studies.
4. Bradford, C. G. An Experiment in Typewriting. *Ped. Sem.*, 1915, p. 445.
5. Batson. Ball Tossing. *Mon. Suppl. Psych. Rev.*, 1916.
6. Bergström. Relation of Interference to the Practice Effect of an Association. *Am. J. Psych.*, Vol. 6, p. 443.
7. Chapman, J. C. The Learning Curve in Typewriting. *Jour. Ap. Psy.*, Vol. 3, p. 252.
8. Hill, Ryall, and Thorndike. Practice in the Case of Typewriting. *Ped. Sem.*, 1913, p. 516.
9. Swift, James E. Acquisition of Skill in Typewriting. *Psy. Bull.*, Vol. 1, p. 295.

EFFECTS OF LOSS OF SLEEP (II.)

BY EDWARD S. ROBINSON AND F. RICHARDSON-ROBINSON

The University of Chicago

Introduction

The effects of loss of sleep may vary considerably from individual to individual. There is a suggestion of this fact in the results of our first experiment.[1] It was determined, therefore, to carry out this second experiment with a larger group of subjects in order to observe the general effect of loss of sleep upon a group and its more specific effects upon individual subjects.

Subjects and Method

The insomnia group consisted of 31 college students. Twenty-five of these, 3 women and 22 men, took all of the tests. The records of the other 6, being incomplete, were disregarded. This group of 25 will be referred to hereafter as the I., or Insomnia, group.

Our control group consisted of 39 college students, 2 women and 37 men. This group will be referred to hereafter as the S., or Sleep, group.

Both groups were given forms 5, 7, and 9 of the alpha intelligence examination in the order named, at the class hours, 10 A.M. or 11 A.M., on the mornings of three successive days. The alpha examination was employed because of its several nearly equal forms. The tests were scored according to the methods described by Yoakum and Yerkes.[2]

The members of the S., or control group lived *as usual* during the experiment. The members of the I. group abstained from sleep during the first day and night of the experiment, and until the completion of the second testing. No attempt was made to control their activities after that. Both groups were asked to state their usual hours for rising and retiring. These estimates afforded an average of 7.7 hours

[1] E. S. R. and Samuel O. Herrmann, Jour. of Exp. Psych., Vol. V, pp. 19–32.

[2] Army Mental Tests, 1920, pp. 53–79.

of sleep for the S. group and 7.4 hours of sleep for the I. group. For the night following the insomnia the S. group reported an average of 7.2 hours of sleep, and the I. group, 9.6 hours.

On the evening of the first day, upon which form 5 had been given to both groups, the members of the I. group gathered in the psychological laboratory. They had been asked to appear at the laboratory at eleven o'clock or soon thereafter. By midnight all the subjects were assembled. They had been given permission to bring work and play materials as they chose. They were also permitted the use of the various rooms, and they had soon chosen quarters for solitary or social pursuits. Sleep only was forbidden and doors were to be left unlocked for continuous inspections guarding against it. There was no evidence of attempts to sleep, and the restless ones who made continuous rounds during the night corroborated the reports that no one slept. At one o'clock a lunch was served. Work, conversation, and games filled the remainder of the night. At six-thirty in the morning, the students 'checked out.' They received instructions to take no sleep before the second testing. While their later reports showed that two students took 'cat-naps' of not to exceed a total of fifteen minutes each in their eight o'clock classes, the others had no sleep. At the class hour form 7 was given both groups. No instructions were given as to sleep during the following 24 hours, and on the third day both groups were given form 9. As we have stated, the I. group slept on the average 9.6 hours or 2.4 hours more than the S. group, between the testing on the second day and that on the third.

Test Records

Table I. gives the average examination scores of both groups for the three days of the experiment.

Table I

Averages (Test Scores)

Group	Day 1		Day 2		Day 3	
	Av.	$P.E._{av.}$	Av.	$P.E._{av.}$	Av.	$P.E._{av.}$
Insomnia	160.20	±3.2	168.16	±3.0	168.48	±3.7
Sleep	161.00	±2.0	170.51	±1.7	171.54	±1.7

The scores of both groups increase from day to day, due, we may be sure, to the effects of practice in taking examinations of this kind.[1] Furthermore, there seems to be no significant difference in the amount of improvement shown by the two groups, even though one of them, the I. group, had no sleep between the testing on day 1, and that on day 2. What difference there is between the two groups favors the S. group, but it does not do so in any convincing fashion.

TABLE II

AVERAGES BY DAYS FOR BOTH GROUPS ON EACH TEST

Groups	Day 1	Day 2	Day 3
Test 1—Oral Directions			
Insomnia	10.36	11.28	11.36
Sleep	10.13	11.46	11.49
Test 2—Arithmetical Problems			
Insomnia	13.16	14.76	15.96
Sleep	12.79	14.62	15.67
Test 3—Practical Judgment			
Insomnia	12.36	13.24	13.72
Sleep	12.36	13.49	14.21
Test 4—Synonym-Antonym			
Insomnia	29.28	28.40	30.04
Sleep	30.41	29.74	30.79
Test 5—Disarranged Sentences			
Insomnia	17.44	20.44	19.04
Sleep	19.08	21.51	19.69
Test 6—Number Series Completion			
Insomnia	12.56	14.72	15.20
Sleep	12.79	13.85	15.36
Test 7—Analogies			
Insomnia	32.00	34.48	33.00
Sleep	29.82	34.95	32.90
Test 8—Information			
Insomnia	33.04	30.84	30.16
Sleep	33.69	30.92	31.44

[1] The amount and nature of these practice effects in the control group are discussed in this journal, Vol. IV, pp. 300–317.

A consideration of the results of the individual tests of the examination (Table II.) also inclines us to judge the differences between the two groups to be insignificant. On the whole, the records of the Sleep group show more improvement from day to day than those of the Insomnia group, but very slightly so. The Sleep group improved more or lost less than the Insomnia group in only 5 out of 8 tests between day 1 and day 2. Between day 2 and day 3, the Sleep group improved more or lost less in only 3 out of 8 tests.

Correlations

In order to ascertain whether the loss of sleep had affected the relative standings, or ranks, of the individuals of the I. group, the correlations were figured between the performances of different days. Table III. shows the values of

$$\rho\left(\rho = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}\right)$$

for both groups.

Table III

Coefficients of Correlations between Ranks on Different Days

	Insomnia		Sleep	
	ρ	P.E.	ρ	P.E.
Days 1 and 2	.86	.04	.75	.05
Days 1 and 3	.88	.03	.81	.04
Days 2 and 3	.91	.02	.84	.03

The coefficients are slightly higher for the I. group, indicating that shifts in ranking from day to day were less in the group subjected to an insomnia. In other words, there is no evidence that some of the individuals in the I. group were deleteriously and some advantageously affected by loss of sleep.

Feelings of Tiredness and of Effort

The students were asked before the examinations on each day to rate their feelings of freshness or tiredness, as follows: very fresh, fresh, as usual, tired, or very tired. Table IV. presents the results for these ratings.

TABLE IV

REPORTS OF FEELINGS OF TIREDNESS, GIVEN IN PERCENTAGES OF STUDENTS RATING EACH TIREDNESS ON EACH DAY

	Day 1		Day 2		Day 3	
	Insomnia	Sleep	Insomnia	Sleep	Insomnia	Sleep
Very tired.........	4	7.9	41.6	2.6	8	13.2
Tired.............	36	26.3	41.6	47.3	60	36.8
As usual...........	52	60.5	17.6	44.7	28	50.0
Fresh.............	8	5.3	0	5.2	4	0
Very fresh.........	0	0	0	0	0	0
Totals............	100	100	100.8	99.8	100	100

It will be seen that on day 1, 40 per cent. of the I. group and roughly 34 per cent. of the S. group were more tired than usual, *i.e.*, tired or very tired. On day 2, 83 per cent. of the I. group and 50 per cent. of the S. group were more tired than usual, and on day 3, 68 per cent. of the I. group and 50 per cent. of the S. group were more tired than usual. In other words, the members of the I. group were but slightly more tired at the beginning of the experiment, felt much more tired after the insomnia on the second day, and on the third day were again near the stage of weariness represented by the normal group. It is apparent from the estimates of the S. group that the students' notions of what it is to feel *fresher* than usual were based upon a rather high standard of *freshness*.

Immediately after the examinations on each day, the subjects were asked to rate the amount of effort felt to have been put forth on the taking of the examination. The scale was as follows: best effort, good effort, ordinary effort, not much effort, and practically no effort. Table V. presents the results of these ratings.

On day 1, before the insomnia, 83 per cent. of the I. group and 90 per cent. of the S. group brought more than an ordinary amount of effort to bear upon the task, while a slightly greater percentage of the I. group than of the S. group put forth their best effort. Fewer of the S. group reported moderate effort, and none of either group on this day reported less than moderate effort. On day 2, 75 per cent. of the I. group reported their best effort as over against 38 per cent. of the S. group. On day 3, approximately 50 per cent. of the I. group and 18 per cent. of the S. group were putting forth their best effort. There is some evidence that the S. group became bored as the experiment progressed. This seems to be indicated by the decreasing percentages giving 'best effort' and the appearance of positive values for 'not much effort.' This boredom was noticeable in the attitude of the class when the forms were produced on the third day.

TABLE V

REPORTS OF FEELINGS OF EFFORT, GIVEN IN PERCENTAGES OF STUDENTS RATING EACH DEGREE OF EFFORT ON EACH DAY

	Day 1		Day 2		Day 3	
	Insomnia	Sleep	Insomnia	Sleep	Insomnia	Sleep
Best effort	41.6	35.8	75.0	38.4	49.9	17.9
Good effort	41.6	53.8	20.8	46.1	29.1	58.9
Ordinary effort	17.6	10.2	4.2	12.3	20.8	17.9
Not much effort	0	0	0	2.6	0	5.1
Practically none	0	0	0	0	0	0
Total	100.8	99.8	100	99.4	99.8	99.8

The instructor asked for volunteers to take a fourth form of the examination on the fourth day, and the fact that twelve members of the I. group and two of the S. group appeared attested to the relative interest of the two groups in the experiment.

The ratings of effort, considered in conjunction with those of tiredness, suggest that loss of sleep causes a general lowering of feeling tone and that this effect is likely to be accompanied by an increase of effort put forth upon a test task. This increase of effort may so compensate for any ultimate loss in capacity that that loss may be relatively unobservable. Of course, it is possible that, although increasing tiredness and increasing effort appear, they are unrelated to the manner or result of test performance. We believe, however, that the theory that they compensate in some degree for each other is more reasonable. Another closely related factor which may have minimized any actual loss in capacity suffered by the Insomnia group was the decreasingly interested condition of the S. group at the times of the second and third tests.

Effects of Loss of Sleep upon Individual Subjects

The subjects were requested to hand in statements of any specific effects of loss of sleep which they noticed. While there were marked individual differences, almost all reported some abnormalities. Nervousness, irritability, dullness, light-headedness, dizziness, and loss of motor control were prominent in the reports. Sleepiness was no more common than some of the other effects. Although unaware of the issues involved, a number of the subjects reported that their symptoms were in abeyance while they were taking the post-insomnia examinations. This, also, may be interpreted as indicative that the members of the I. group took unusual pains with the post-insomnia tests.

Conclusions

The qualitative results of this experiment, like those reported in our first paper, show that loss of sleep is marked by a variety of manifestations of disorganization which differ considerably from individual to individual, but which are prominent enough to be important. This finding accords with the results of previous investigators. Our attempts to measure the magnitude of the effects of loss of sleep for one or two nights by means of standard test procedures brought practically negative results. This does not mean, however, that the functions represented by the tests were unaffected, but merely that the effects were not too great to be compensated for by an extra amount of effort or interest, the existence of which our ratings of effort give good evidence. An interesting problem is furnished by the appearance of the extra effort in the Insomnia group. These subjects felt various symptoms of their condition as is shown by their tiredness ratings and the more specific reports, and these may have stimulated that alteration in their method of work which we loosely describe as work with greater effort. There was every evidence, too, that the Insomnia group, working under unusual conditions, took interest in the experimental situation after the Sleep group, unaffected by loss of sleep, had become relatively bored. Perhaps a combination of these factors was operative.

We believe that our two experiments cast pronounced suspicion upon the procedure of placing too much emphasis in such studies as this upon test scores. Such data are far easier to handle than those obtained by less objective methods, but they are not therefore more significant.

PRACTICE EFFECTS IN INTELLIGENCE TESTS

BY EDWARD L. THORNDIKE,

Institute of Educational Research, Teachers College

Yoakum and Yerkes ('20, p. 50) report an average gain of 8 points for a second over a first trial in Alpha. Dunlap and Snyder ('20, p. 395) report the gains of second over first, and third over second trial with Alpha (forms 6, 8 and 5 respectively), for 44 college students scoring 88 to 195 in the first trial. In a fourth trial (form 9) there was a large loss compared with trial 3, due, so they report, to decline of interest and effort. The median gain from trial 1 to trial 2 was about 15; that from trial 2 to trial 3 was about 12½. Allowing for the differences in difficulty of forms 6, 8 and 5 as per the data of Table 177, p. 660 of Psychological Examining in the United States Army, the median gain from trial 1 to trial 2 would be about 16 and that from trial 2 to trial 3 about 7.

Richardson and Robinson ('21, p. 302 and 305) find (using 39 individuals scoring 102 to 198 in the first trial) a gain of 9½ points for a second over a first trial of Alpha, and about 5½ points for a third over a second trial (1.03 as the actual gain in score plus 4.5 as the estimated greater difficulty of the form used in trial 3).

Combining the last two results we may expect a gain of about 12½ points for a second over a first Alpha, and of about 6 points for a third over a second, in the case of college students. As will be shown, we may expect that 10 minutes of fore-exercise could be used to reduce the former of these gains substantially.

The facts to be reported here are in general like the above. They differ in three respects. The number of cases studied is very large; the influence of inequality of difficulty of forms is usually eliminated by using two forms in changed order. Where allowances are made for differences of the forms in difficulty, the allowances vary only from + 3 to − 3 and have probable errors of about 0.5. The test used was not the Alpha, but a test made up of harder tasks of the same sort plus five others, and given with all directions printed so that the person tested used as much or as little of the test time as he wished in mastering the directions. This test was prepared originally for use by the Examining Boards of the United States Air Service. It is printed in fifteen alternative forms. The scores of college entrants vary around 100 as a central tendency, ranging with few exceptions from 70 to 135.

Consider first the results when two forms are given in immediate succession after 10 minutes of fore-exercise spent in examining another form and doing the first two or three elements of each of the thirteen tests.

Forms *D* and *E* were given in the order *D E* to three groups of college entrants, numbering 312, 134, and 99 respectively. The median $E - D$ was $4\frac{1}{2}$, 4, and $3\frac{1}{2}$, respectively. The same forms were given in the order *E D* to 72 college entrants. The median $D - E$ was 12. The effect of the practice here was thus a median gain of 8.

Forms *I* and *M* were given in the order *I M* to 320 college entrants, 69 graduate students and 24 theological students. The median *M–I* was $3\frac{1}{2}$, − 1, and 2 respectively. As a weighted total result we have $M - I = 3-$. The same forms were given in the order *M I* to 419 college entrants and 409 normal-school students. The median *I–M* was $12\frac{1}{2}$ and $11\frac{1}{2}$ respectively. The effect of the practice was thus 7 +.

Forms *L* and *O* were given in the order *L O* to 273 college entrants. The median *O–L* was 8. When they were given in the order *O L* to 54 high-school seniors the median *L–O* was $8\frac{1}{2}$. The effect of the practice was thus 8 +.

Forms *N* and *F* were given in the order *F N* to 167 college entrants. The median *N–F* was $7\frac{1}{2}$; when the order was *N F* (70 entrants) the median *F–N* was 12. The gain from practice was thus 10 −.

Forms *E* and *O* were given in the order *E O* to 33 first-year engineering students. The median *O–E* was 6. When given in the order *O E* to 20 students, the median *E–O* was 10. The gain from practice was thus 8.

A general weighted average of all these results gives 8 + with high reliability. Other scattered data confirm this estimate of 8 as the median gain from first trial to second trial after ten minutes of fore-exercise.

I have elsewhere shown that the practice effect from a first to a second trial when no fore-exercise is given is $12\frac{1}{4}$. The ten minutes of fore-exercise thus seems a useful proceeding. If, instead of it, a full regular trial of 30 minutes is given, the practice effect from a second to a third trial is $3\frac{1}{2}$.

I have no data for high-school graduates beyond a third trial, but by the kindness of Miss Henrietta Race, I have records of twenty gifted children about eleven years old, and nineteen children taken at random from the same school grade as the gifted children, who took fifteen forms in succession, one each school day. The average results, allowance being made for variations in the difficulty of the forms,[1] are as follows:

Trial	1	2	3	4	5	6	7	8
Gifted	87.5	108	109	109	106	113.5	109.5	110.5
Ordinary	51	76	67.5	79	82.5	80	78.5	80

Trial	9	10	11	12	13	14	15
Gifted	111.5	115	108	112	111	115	111
Ordinary	81	81.5	82	85.5	88	88	86.5

Combining the results for the two groups, and drawing a smooth curve to fit the observations (Fig. 1), it becomes obvious that the gain from practice falls off very rapidly and is, from about the fourth trial on, very small in comparison with differences between individuals. It does not, however, entirely vanish even in the fifteen trials. In so far as the results from so small groups may be trusted, it does almost vanish for the gifted group after about trial 6, as shown in Fig. 2. For them, indeed, all the practice after about the

[1] These allowances were computed from records of high-school graduates. They were for forms *A*, *B*, *C*, etc., in order, + 2, − 3, + 2, + 3, − 1, 0, − 1.5, + 1.5, + 2.5, + 1, 0, − 1.5, − 2, − 2, − 2.

second or third trial means only a very slight gain of about five points.

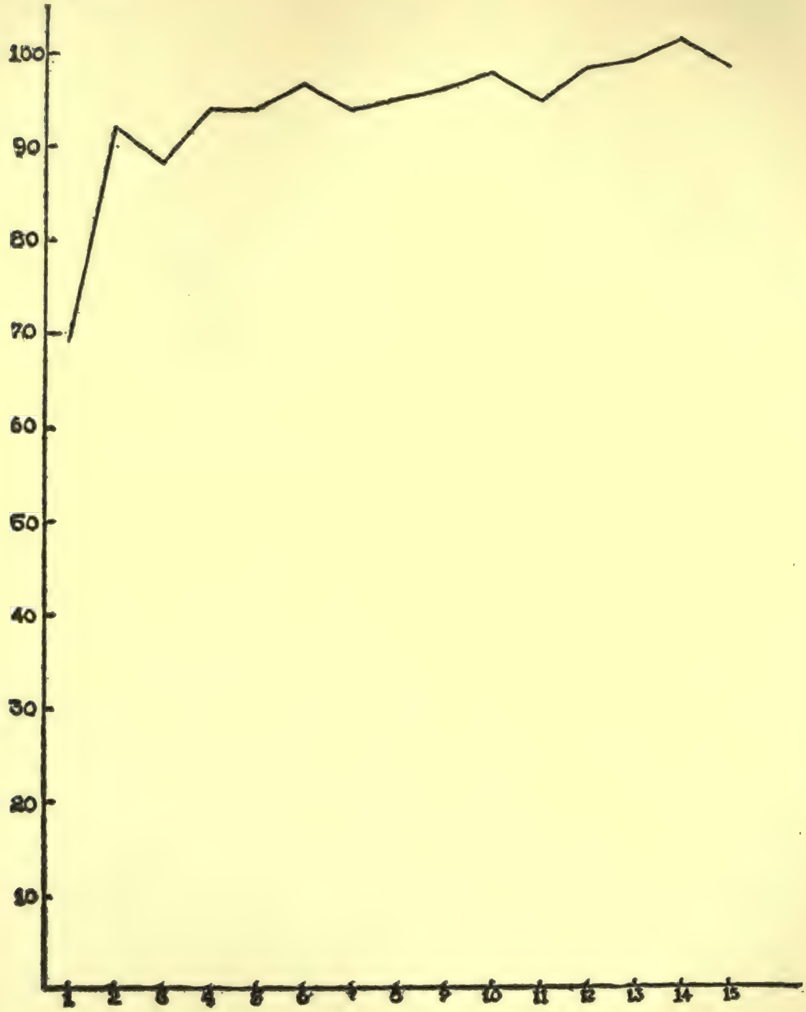


FIG. 1.

We may then hope that in experiments with individuals of the intelligence level of college students in respect to the influence of fatigue, drugs, distraction, and the like, the disturbing effect of practice after the third trial will be very, very slight. We may estimate the probabilities as follows (10 minutes fore-exercise being given):

	Per Cent.
Gain Trial 1 to Trial 2	8
Gain Trial 2 to Trial 3	3
Gain Trial 3 to Trial 4	1 to 1½
Gain Trial 4 to Trial 10 or 15	2

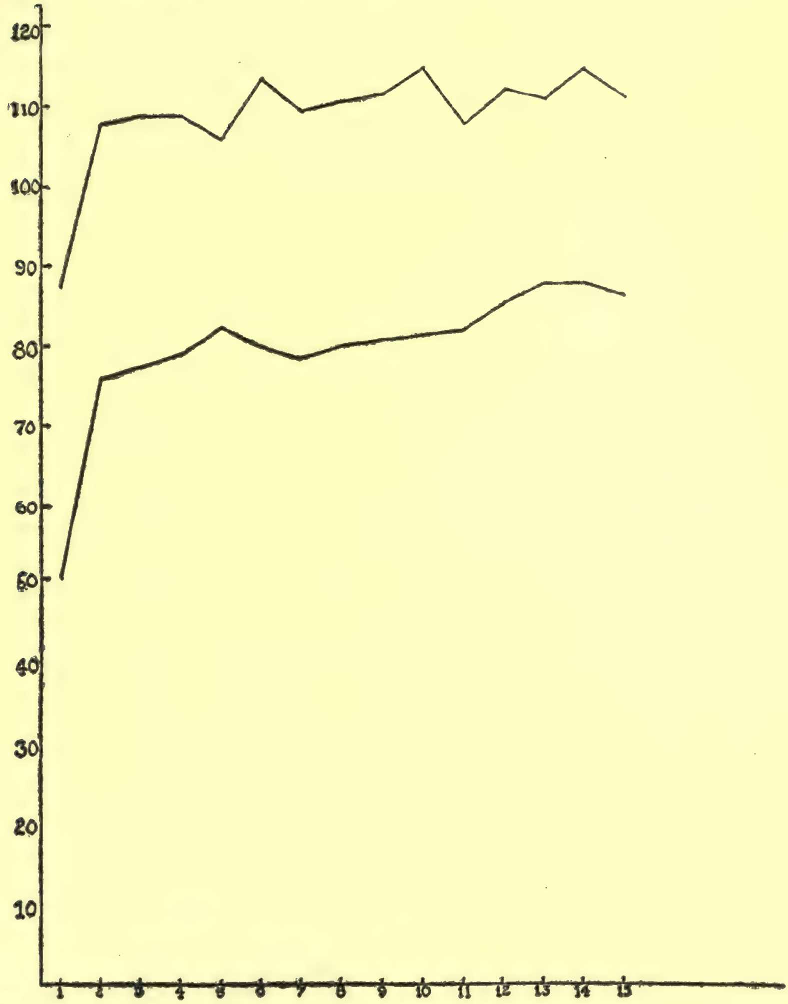


FIG. 2.

The practice effect is approximately the same regardless of the ability of the individual within the limits of 70 to 130. The correlations between intelligence score in the entire examination of 150 minutes and gain from first to second trial of Part I are:

467 high-school graduates, male	−.11
254 high-school graduates, female	−.15
222 normal-school students (mostly female)	−.04

The correlations between intelligence score in Part I and gain from first to second trial of Part I are:

342 high-school graduates, female	−.04
402 high-school graduates, female	+.02

The more intelligent candidates thus made slightly less gross gain in the second trial over the first, probably because they grasp the nature of the tests better from the practice trial and first regular trial. In cases where we wish to infer what a second trial score would be from a first trial score, or to allow for practice in other ways, we may allow + 8 (or, a little more precisely, + 7, + 8, and + 9 respectively) to the brighter quarter, middle half, and duller quarter, of a group of the general level of high-school graduates.

In the preparation of future issues of the Intelligence Examination which is being used as one element in the record of candidates for entrance at Columbia University and elsewhere, we have studied tests of many kinds in respect to their freedom from influence by general practice in taking tests, special practice on the particular sort of test in question, and actual coaching. A report of this work will be made later, but it may be noted now that a test in reading difficult paragraphs and answering difficult questions about them, is certainly one of the very best tests in these respects. I find the practice effect very, very slight with university students, and Gates ('21, p. 381) has found the same with children in grades 4 to 8 of the elementary school. We are preparing a series of such tests to be of known, and approximately equal, difficulty suited to the intelligence level of college students, as the McCall-Thorndike series of ten forms is suited to the level of pupils in grades 3 to 10. These will be especially valuable for short-term experiments with fatigue, drugs and the like where it is not practicable to give an hour or so to fore-exercise. They have, however, two disadvantages in comparison with Army Alpha or the instrument discussed in this article. First, the test has to be rather long to secure sufficient reliability (by sufficient variety in the topics of the paragraphs and otherwise). Second, the scoring requires judgment, and provision must be made to free the results from the personal equations of the scorers. We are now experimenting with a form where the individual *selects* the right answer instead of writing it, which will do away with this second disadvantage. It remains to be seen, however, whether this selective form will show the same high correlations with a criterion of intelligence that the customary form has shown.

References

GATES, A. I., '21. An Experimental and Statistical Study of Reading and Reading Tests (Continued). *Journal of Ed. Psy.*, Vol. 12, pp. 378–391.

DUNLAP, K., AND SNYDER, A., '20. Practice Effects in Intelligence Tests. JOURNAL OF EXP. PSYCHOLOGY, Vol. 3, pp. 396–403.

RICHARDSON, F., AND ROBINSON, E. S., '21. Effects of Practice upon the Scores and Predictive Value of the Alpha Intelligence Examination. JOURNAL OF EXP. PSYCHOLOGY, Vol. 4, pp. 300–317.

YOAKUM, C. S., AND YERKES, R. M., '20. Army Mental Tests.

THE CONDITIONED PUPILLARY REACTION

BY HULSEY CASON

Columbia University

Section 1

Introduction

Much of the work which has been done on the conditioned 'reflex' is open to the criticism that the experiments involved 'voluntary factors' which complicated and in some instances vitiated the conclusions reached. We are not absolutely certain that all of the conditioned reactions which have been secured were 'reflexes' in a strict sense. The rôle which is played by 'conscious' factors, in the opinion of the present writer, looms up as a formidable factor. Much of the work of Pawlow and his associates, Watson's emotion experiment with the baby, and the latter's work on animals is strong argument in favor of the conditioned reflex being a real thing. The writer is of the opinion that the experiment now about to be described furnishes conclusive proof that a reaction may be conditioned without any direct aid from the 'voluntary factors.'

In beginning work on the conditioned reflex, it was our purpose to arrange an experiment which would eliminate these voluntary factors as far as possible. I was somewhat surprised to find that very few reflexes were suited to such an undertaking. Furthermore, it might be observed here that in almost any sort of conditioned reflex experiment, it is desirable (*a*) that the reflex used be one which can be called out repeatedly, (*b*) the reflex should be one which does not change markedly in appearance or fatigue easily with repetition, and (*c*) the reflex response should be adapted to objective registration. Just a little thought on the matter will, I think, convince the reader that there are few reflexes which meet all of these requirements. The pupillary reflex, selected for this particular experiment, has the following very obvious advantages: (*a*) it can be readily evoked by a change in the intensity of light falling on the retina, (*b*) it does not fatigue markedly if proper precautions are taken, (*c*) changes in the size of the pupil can be easily observed with suitable apparatus, which I may say was available for the present experiment, (*d*) the fundamental stimulus (change in light intensity) calling out the pupillary reflex can be easily controlled, and (*e*), what is probably most important of all, this particular reflex cannot be controlled voluntarily in a strict sense. The fact that the pupillary reflex is not subject to voluntary control in a strict sense makes the conditioned pupillary reaction which was secured in this experiment both interesting and suggestive on theoretical grounds.

This experiment was carried out in the Columbia University laboratory, during the year 1921–1922. The writer wishes to make a very grateful acknowledgment to Professors R. S. Woodworth and A. T. Poffenberger, and to Mr. Arthur L. H. Ruben for many valuable suggestions and criticisms.

The 'fundamental' stimulus used was a change in the intensity of light falling on the retina. The 'conditioned' stimulus was generally the sound of an electric bell. The change in light intensity (followed of course by a change in the size of the pupil) and the bell stimulus were repeated together a large number of times. The purpose of the experiment was to see whether the change-in-size-of-the-pupil could not eventually be conditioned to the bell, that is to say, whether giving the bell stimulus alone could not eventually cause a pupillary reaction.

I was fortunate in having the use of an instrument which was constructed especially for measuring the size of the pupil and for controlling the light stimulus. This apparatus was devised by Karl Weiler in Munich, and is referred to as the 'Weiler apparatus.' I have included a brief description of its most essential features in section 3 of this report.

At the beginning of each experiment, I measured the natural effect of the bell stimulus on the size of the pupil.

The bell caused a slight dilatation in the case of every subject tested. Next followed a long period during which I would cause the pupil to make a marked change in size, by altering the light stimulus, and I would give the bell stimulus while this *change* was taking place. At the end of an experiment with a subject, I again measured the size of the pupil, with and without the bell, and found that the 'training' had left its mark. The change in light intensity, immediately followed by the bell stimulus, will hereafter be referred to as 'training stimuli.'

Two different procedures were used in giving these training stimuli. In one procedure, the light was turned *off*, and the bell stimulus was present during a *dilatation* of the pupil, as the training stimuli. It will be seen later that after this training a dilatation of the pupil was conditioned to the bell, *i.e.*, ringing the bell would now cause the pupil to dilate more than it had done at the beginning of the experiment. In the other procedure, the light was turned *on*, and the bell stimulus was present during a *contraction* of the pupil, as the training stimuli. After this training had been continued for some time a contraction of the pupil was definitely conditioned to the sound of the bell, *i.e.*, ringing the bell would now cause the pupil to actually contract. The same light which was used during the training period was always present when measurements of the size of the pupil were being made, both before and after training, and with and without the bell. The first procedure will be referred to as the 'training towards dilatation'; the second as the 'training towards contraction.'

The next section is a brief statement of certain general characteristics of the pupillary reflex which concern us particularly. The rest of the report is in the form of the usual laboratory experiment. Section 3 contains a short description principally of what the apparatus can do. Section 4 contains an outline of the procedure. In section 5 the results of all the experiments with the pupillary reflex are discussed somewhat in detail; and in section 6 the main conclusions which seem justified from the whole experiment are stated.

Section 2

The Pupillary Reflex

Of the two layers of unstriated muscle fibers of the iris, one layer is circularly arranged so that by its contraction it acts as a sphincter and causes the pupil to contract, the other layer is arranged radially so that by its contraction it causes the pupil to dilate. When light is thrown on the eye, there is an average latency in pupillary contraction of 0.2 sec. (19, 164). The pupil then makes a rather quick contraction, gradually slowing up until 0.87 sec. have passed. This is followed by some slight fluctuations; but 1.2 sec. after the original stimulus was received the pupil practically reaches a level (19, 168–169). These are only average figures, even normal people showing marked variations from them.

If a light of a certain intensity is falling on one eye, and a light of the same intensity is thrown on the other eye, both pupils will make a further contraction. Using very bright light, Weiler was able to produce a 'secondary reaction' of about 0.45 mm. (19, 167). Increasing the light for one eye caused an average contraction of 1.3 mm.; while increasing the light for both eyes caused an average contraction of 1.7 mm.

The pupil appears to make very fine reactions to a great many stimuli. The result in man is that the iris is probably always making fine, oscillatory movements (19, 111; 12, 185–248).

Some of the circumstances under which *contraction* of the pupil occurs are as follows:

1. Increase in the intensity of light. In man, the light falling on one eye causes simultaneous contraction of both pupils (19, 169; 11, 267). If a light which has caused a very large contraction of the pupil is continued, the pupil gradually relaxes with the adaptation of the retina to light.

2. Convergence of the visual axes (16, 73–100). A change in convergence, without a change in accommodation, is sufficient to provoke the pupillary reflex; the contraction being due solely to convergence, to the exclusion of the

modifications of accommodation (15; 12, 220–226). The movements of accommodation are simply associated with those of convergence (11, 80–81; 13, 22; 12, 225). These are the only causes of contraction of the pupil which concern us particularly, and it will be seen later that these two factors were controlled.

Some of the circumstances under which *dilatation* of the pupil occurs are the following:

1. Decrease in the intensity of light. Generally speaking, the size of the pupil is determined by the relationship between the adaptation of the eye and the illumination (13, 23). Almost complete adaptation for darkness occurs after 30 min., but if there is any light present at all the pupil contracts again as the eye becomes dark adapted.

2. Divergence of the visual axes.

3. Emotional states. According to Waller (17, 548), dilatation of the pupil may be provoked by painful impressions, suddenly aroused attention, and dyspnea. I conducted a short experiment to determine the effect of different mental states on the size of the pupil, designed to cover most of the conditions met with in this experiment, but I did not arrive at any very definite conclusions. Some careful research is needed to clear up this point.

4. Other sensory or sensorial stimuli (16, 101ff.; 12, 228–241). It seems that various stimuli, muscular, visceral, and sensorial, cause indirectly a dilatation of the pupil.

I shall not attempt to give a complete account of the nerve fibers which go to and from the iris, because this would take us somewhat afield, and is besides not necessary. The reader can find rather thorough descriptions in the works of Magitot (12), Weiler (19), Bach (1), Vidal (16), and Lutz (11). Bach goes more into detail than any of the others.

The size of the pupil, in normal people, is the result of a balance of forces maintained by the sympathetic nerves, which act as dilators of the pupils on the one hand, and the motor oculi which opposes them and causes contraction of the pupils on the other (6). If the sympathetic is stimulated, the pupils may dilate, because the balance is upset.

Conversely, irritation of the third nerve tends to contract the pupils. Thus the sympathetic is mydriatic in function, while the motor oculi is miotic. If the sympathetic fibers are stimulated, and at the same time the third nerve is inhibited, a considerable dilatation of the pupils may be expected. Since stimuli may affect the controlling nerves at various centers in their devious courses, it is to be expected that a change in the size of the pupil may be brought about by a variety of causes (6, 170). According to Magitot (12), two sympathetic centers, the one mesencephalic, the other bulbo-medullary, preside over the movements of the pupil, each through the medium of a peripheral ganglion: the constrictor center by the ganglion ciliare, the dilator centers by the superior cervical ganglion. It is thus a question of a centrifugal chain to two motor neurones.

In most of the experiments to be described, the 'conditioned' stimulus was a sound, generally of an electric bell. We are therefore concerned with the path of the cochlear nerve, which I shall not attempt to describe (9, 216).

Section 3

Apparatus

Through the kindness of Dr. George H. Kirby, Director of the Psychiatric Institute at Ward's Island, New York, I obtained the use of an instrument for measuring the size of the pupil and at the same time controlling the light stimulus. This instrument is known as the 'Weiler apparatus.' Weiler gave a very complete description of this apparatus in 1910 (19). In the same article may also be found an account of a systematic exploration of the pupil, for both normal and pathological people, together with a description of the accessory apparatus which he employed in photographing the movements of the iris and in recording the time of contraction and dilatation. I shall give only a brief description of the essential features of this apparatus (see Fig. 1).

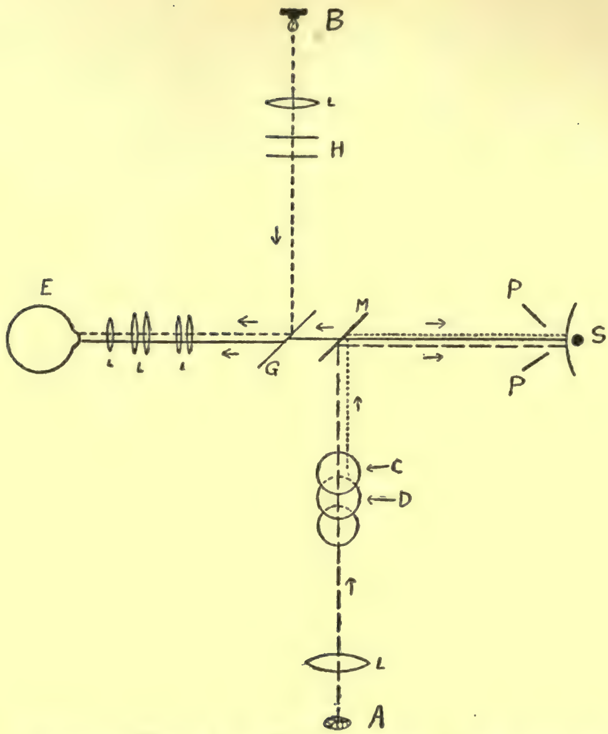


FIG. 1. Schematic Plan of the Weiler Apparatus. Not drawn to scale. *E*, experimenter; *S*, subject; *A*, main source of light; *H*, parallel hairs; *B*, light for parallel hairs; *M*, mirror; *G*, transparent glass; *D*, light filter; *C*, glazed paper which may be moved aside; *L*, lens. Light travels mostly in the direction of the arrows. The mirror (*M*) is to the side of the line of regard passing from *E* to *S*. The parallel hairs may be moved in and out by a thumb-screw. They seem to the experimenter to be superimposed on the subject's eye.

The experimenter is located on the left in the drawing (at *E*); the subject has a position on the right (*S*). Light from the two sources (*A* and *B*) is so controlled that it can travel only in the direction of the four broken lines of the figure. That coming from the bright source (*A*), which is enclosed in a box, is reflected by the mirror (at *M*) to the subject's eye. The luminosity of the light which reaches *S* can be regulated by light filters (at *D*). If we have a light of low intensity falling on *S*, a light of higher intensity may be readily secured by pushing in a spring button of the apparatus, which moves aside a piece of glazed paper (*C*), placed horizontally, and at the same time makes an electrical connection which was generally used in this experiment to start the bell.

The mirror (*M*) can be revolved horizontally, directing the light from *A* through the subject's eyepiece (*PP*), and at the same time enabling the subject to see clearly the translucent paper (*C*), or the glass filter (*D*), as the case may be. The central portions of both *C* and *D* appear to the subject to be of uniform luminosity. The mirror (*M*) is just to the side of the line of regard from the experimenter (*E*) to the subject (*S*); this line passing through a piece of glass at *G*.

Light of low intensity (from *E*) comes down through a tube about 1.75 inches in diameter, passes 'by' the parallel hairs which appear to the experimenter to be placed verticalyl, and is reflected by the glass (*G*) to the observer (*E*), so that the hairs seem to the experimenter to be superimposed on the subject's eye. The intensity of *B*, and hence the luminosity of the outline of *H*, can be regulated by a rheostat, until the brightness of the outline of the vertical hairs is most favorable for seeing both the subject's eye and the hairs simultaneously.

The five lenses near *E* can be moved to the right and left (in the figure) so as to secure the most favorable focus on *H* and *S*. The small light at *B* can be moved up and down to assist the experimenter further in getting a focus on *H*. The vertical hairs (*H*) may be moved toward and away from each other by means of a large thumb-screw until they are each tangent to and on opposite sides of the subject's pupil. The diameter of the pupil is indicated automatically on a circular scale to the tenth of a millimeter. In the diagram, the lower right portion (*A–M–S*) of the whole apparatus is stationary. The left upper section (*E–G–B*) can be readily moved to the right and left, and up and down, until the vertical hairs are brought on the subject's eye. The subject keeps his head fairly steady by using the chin rest and at the same time resting the lower portion of the frontal bone against the projecting eyepiece.

From the above description, one might infer that the light stimulus can be regulated for only one of the subject's eyes at a time. The section *A–M–S*, however, is in duplicate,—one part for each of the subject's eyes. After measuring the size of one pupil, however, it is necessary for the experimenter to shift the part *E–G–B* to the right or left if the subject's other pupil is to be measured. The light stimuli for both eyes can be manipulated simultaneously; or the light for just one eye may be changed. In the case of every subject except *C* and *E* (first experiment only) I used only that part of the apparatus relating to the subject's left eye.

The 'make' and 'break' connections in the apparatus were generally used to start and stop the bell, at the same time regulating the light stimulus. Certain simple modifications in the electrical wiring were necessary, but these need not be described. The experiment was conducted in a dark room. This description is sufficient, I hope, to give an idea of the facilities I had at hand for measuring the size of the pupil.

Section 4

Procedure

The plan of procedure in this experiment required each subject to remain in a dark room continuously for a period of about 3 hours and 20 minutes. My obligations to those who were kind enough to serve as subjects are therefore very great. All of the subjects were graduate students in psychology except Z, who was a college freshman.

The results of 4 subjects, where the training was towards dilatation, are grouped together in the discussion of the results; and the procedure employed for these 4 subjects will now be described. This procedure was as follows:

1. At the 'ready' signal, the subject put himself in position and looked (with his left eye) at the light stimulus. The right eye was in almost total darkness.

2. After an interval of varying length, but always after the pupil had about reached a level of contraction, I turned

the light completely off, and gave the bell stimulus while the pupil was dilating. In this particular procedure, I had made the glazed paper (*C* in Fig. 1) opaque to light.

3. If the subject makes some particular movement each time the conditioning and fundamental stimuli are present, this movement may itself be 'conditioned' to something, and thus serve to complicate the results. Therefore, while the bell was ringing (and while the pupil was dilating), I had the subject remain in the same position and look in the same direction, without winking or otherwise moving the muscles of his face.

4. After a period of about 1 sec. the bell stopped ringing and the subject took a comfortable position to wait for the next 'ready' signal, which came 15 sec. later.

The results of 4 subjects, where the training was towards contraction, are likewise grouped together in the discussion of the results. The procedure used for these 4 subjects is as follows:

1. At the ready signal, the subject put himself in position to receive the light stimulus.

2. The light was turned on, and the bell stimulus was present while the pupil was contracting.

3. While the pupil was contracting the subject kept the same position and continued looking at the light.

4. When the pupil had stopped contracting (generally after about 1 sec.), I turned off the bell and the light together, and the subject rested until the next 'ready' signal was given.

For subjects *Y* and *Z* the light changes were made for both eyes instead of for only one. These two subjects, however, looked with the left eye at the luminous field on that side.

All measurements were taken with the same intensity of light, the same light which I turned on and off during the training series. That is to say, the change in light intensity was from a bright light to total darkness, or vice versa. This bright light is 'the light' which is so often referred to in this report, and it was constant throughout all experiments. It is to be particularly noted that this same light was always *present* when measurements of the size of the pupil were being made. My purpose in having the change in light intensity from the bright light to total darkness, or vice versa, was to cause as large a *change* in the size of the pupil as possible each time the training stimuli were given.

Some preliminary experiments had indicated that to satisfactorily condition the pupillary reflex to the sound of a bell it would be necessary to have a rather long and continuous training series. For all subjects except *F* (in Table VII.) the schedule to which I rigidly adhered was as follows:

1. A period of 20 min. was allowed to elapse in the dark room, to take care of the adaptation factor (19, 154).

2. I took several measurements of the subject's pupil, with and without the bell, to determine the 'natural' effect of the bell on the size of the pupil. An individual measurement was executed about as follows: with the left hand, I moved the observer's telescope and the vertical hairs until the latter were 'over' the pupil. At the same time, with the right hand, I caused the two hairs to move in and out until they were tangent to and on opposite sides of the pupil. This operation generally consumed about 5 sec.[1]

The measurements of the size of the pupil, with the bell and without the bell, were mixed up in an irregular order and followed each other at intervals of about 20 sec. The reader may wonder why certain rests, etc., were of the stated length of time. Many of these were quite arbitrary on my part. My purpose was to cause as many changes in a sub-

[1] My personal error in measuring the size of a pupil was checked up by an objective method. There is no reason to believe that there was a constant error of any kind. Variable errors, however, in measuring the same pupil had a variability (S.D.) of 0.04 mm.; or, as the figures are used in the table, an S.D. of 0.4. This means that variabilities obtained throughout the experiment are given a little *higher* than they should be. If the obtained S.D., for example, of a distribution of measurements of the size of the pupil is 2.5, and if the variability (S.D.) of the measurements of the observer (personal variable error) amounts to 0.4, then the most probable S.D. of the distribution, if there had been no personal error in the observer (or, if compared with the 'true'), would be given by $\sqrt{(2.5)^2 - (0.4)^2}$, or 2.47. I do not make use of this correction in the computations to follow, although there is no reason why it should not be legitimately used.

ject's pupil as possible during a single experiment; but it is very probable that a better distribution of rest periods, etc., could be worked out.

3. A rest of 3 min.

4. The training series. This lasted 2½ hours, and consisted of 400 repetitions of the training stimuli. The schedule of rest periods, etc., which I followed here was as follows:

(*a*) 16 repetitions of the training stimuli, each repetition being followed by a rest of 15 sec.
(*b*) A rest of 1 min.
(*c*) The above, (*a*) and (*b*), repeated 5 times.
(*d*) A rest of 5 min.
(*e*) The whole group, *a*, *b*, *c*, and *d*, was itself repeated 5 times, making altogether a training period of 2½ hours.

During this training series, the bell stimulus was never present except when the pupil was changing markedly in the direction of the training. The reverse is also true: the pupil did not change markedly in the direction of the training except when the bell stimulus was present. That is to say, neither was present without the other. I did not take measurements of the pupil during the training series; the subject's eye being exposed to the bright light only long enough to secure the desired pupillary reaction.

For subjects *A*, *B*, *C*, *D*, *W*, *Y*, and *Z*, I used a small electric bell, driven by 4 dry-cell batteries, as the conditioning stimulus. This bell was screwed down to the table, and made a very aggressive sound. For subject *E* (in Tables III. and IV.), I used the same bell, but driven by one weak battery. For subjects *X* (in Table V.) and *B* (in Table VI.), I used the same bell again, but it was not fastened down to the table. For subject *F* (Table VII.), the conditioning stimulus was a weak but definite electric shock, applied to the right hand. For subject *X* (Table II.), I started off with a very loud electric gong, driven by 4 dry cells; but after 260 repetitions the clapper of this gong broke completely off, and could not be immediately replaced. I finished the experiment with the smaller bell, and the results were of the same general nature as those obtained from other subjects.

5. After the training series I again measured the size of the pupil, with and without the bell stimulus, to determine what the effect of the training had been. During these measurements, the bright light was always present. At this stage, the procedure was necessarily a little irregular. I generally took 2 or 3 measurements, immediately followed by several repetitions of the training stimuli. This had to be done because the conditioned reflex which I secured was rather fragile in constitution and soon passed out. In such a case as this, the total number of times the conditioned reaction is evoked is somewhat arbitrary. If you call it out only a few times the reliability of the result will probably not be very high; while if you call it out a large number of times the reaction disappears. So I tried not to be greedy in the matter, and took only a moderate number of measurements. These are shown in the various tables of results, with which we are next concerned.

Section 5

Results and Discussion

All of the 9 different subjects were graduate students in psychology except Z, who was a naïve subject. Six subjects: *B*, *C* (same as *W*), *E*, *F*, *Y*, and *Z* were men; and three subjects: *A*, *D*, and *X* were women. While there were 9 different subjects, there were 13 different experiments.

The results of the experiments with the first 8 subjects are shown in Tables I. and II. Table I. gives the results of 4 subjects, where the pupil was trained to dilate when the bell stimulus was given. Table II. gives the results of 4 subjects whose pupils were trained to contract when the bell stimulus was given. In every experiment, the bell caused a slight dilatation of the pupil at the beginning of the experiment. This dilatation varies with different individuals, and in the same individual at different times.

Let us consider subject *A*, in Table I., to illustrate the procedure. Before training, I measured this subject's pupil 7 times without the bell (distribution *a*), and 7 times with the bell (distribution *b*). After training, I measured her pupil 11 times without the bell (*c*), and 11 times with the bell (*d*). I thus have four distributions. Consider now the two distributions before training, *a* and *b*. The natural effect of the bell is given by $b - a$, in this case $+ 1.14$. This means that the bell increased the size of this subject's pupil 1.14 *tenths of a millimeter* at the beginning of the experiment. The P.E. of this Diff. is 0.56. I performed a similar operation on the two distributions after training. The bell now increased the size of the pupil 3.09, and the P.E. of this Diff. is 0.47. The effect of the training series on the size of the pupil is given by $(d - c) - (b - a)$, or $+ 1.95$. While the reliability of the results in Table I. are fairly high, they are not conclusive unless they are considered in connection with Table II., where the pupils of 4 subjects were trained to actually contract when the bell stimulus was given.

In Table I. the conclusion seems established that a bell may be made to cause more dilatation after training (towards dilatation) than before. The conclusion that the bell will cause greater dilating effect on the pupil after training than before training is true for the 4 subjects, *A*, *B*, *C*, and *D* with reliabilities of 96 per cent., 86 per cent., 98 per cent., and 99 per cent., respectively. It is to be noted that the *same* positive conclusion was reached in the case of *each* of these 4 subjects with the reliabilities as just indicated. Now the reliability of the conclusion itself is obviously greater than any one of the the 4 individual reliabilities given above. In fact, the chances are 99.9997 out of 100 that the same conclusion itself is correct; and 81.55 out of 100 that the same conclusion would be reached again in *all* 4 cases.

Table II. gives the results of 4 subjects, *W*, *X*, *Y*, and *Z*, whose pupils were trained to contract. Let us consider subject *W* for illustration. Before training, the bell dilated the pupil 3.50. After training the bell *contracted* the pupil 1.73. The total drop is $- 5.23 \pm 1.45$; and the chances are 99 out of 100 that, if the experiment is repeated under the same conditions, the bell will have more effect in the

Table I

Effect of Training Towards Dilatation

Subjects	*A*	*B*	*C*	*D*
Before Training:				
(*a*) Light:				
Meas.	7	9.	10	8
Av.	40.43 ±0.33	41.33 ±0.58	65.30 ±0.53	52.50 ±0.34
S.D.	1.29	258	2.49	1.41
(*b*) Light and bell:				
Meas.	7	9	10	8
Av.	41.57 ±0.45	43.67 ±0.30	69.40 ±0.32	56.25 ±0.55
S.D.	1.76	1.33	1.50	2.33
Bell, or $b - a$	+1.14 ±0.56	+2.34 ±0.71	+4.10 ±0.62	+3.75 ±0.65
Training: Bell with dilatation of the pupil				
After Training:				
(*c*) Light:				
Meas.	11	12	9	12
Av.	38.82 ±0.39	41.67 ±0.33	52.00 ±0.76	54.25 ±0.46
S.D.	1.19	1.74	3.40	2.35
(*d*) Light and bell:				
Meas.	11	12	9	12
Av.	41.91 ±0.27	45.33 ±0.24	59.67 ±0.48	63.08 ±0.38
S.D.	0.81	1.25	2.14	4.37
Bell, or $d - c$	+3.09 ±0.47	+3.66 ±0.41	+7.67 ±0.89	+883 ±0.85
Effect of training, or $(d - c) - (b - a)$	+1.95 ±0.73	+1.32 ±0.82	+3.57 ±1.08	+5.08 ±1.32
Chances out of 100 that $d - c$ will be equal to or greater than $b - a$	96.41	86.12	98.68	99.53

The numbers in the above table represent the horizontal diameter of the subject's pupil, to the tenth of a millimeter. The measurements of the pupil at the beginning of the experiment, with and without the bell, were mixed up together (distributions *a* and *b*). The measurements of distributions *c* and *d* were also mixed up with each other. For each distribution, the number of measures made, the average, the standard deviation, and the probable error of the average, are shown in the table. The effect of the bell on the size of the pupil is represented by $b - a$, before training; and by $d - c$, after training. The probable errors of these differences were also calculated. During the 'training series' the light was turned *off*, and the bell sounded while the pupil was dilating. This was repeated 400 times. Some measure of the *effect of this training* is obtained by subtracting the effect of the bell before training $(b - a)$ *from* the effect of the bell after training $(d - c)$. This gives a difference between two differences, or $(d - c) - (b - a)$. Finally, the reliability of this difference was obtained. For example, if $(d - c) - (b - a) = + 1.95 \pm 0.73$, the chances are 96 out of 100 that, if the experiment is repeated under the same conditions, $d - c$ will be equal to or *greater* than $b - a$. Or, stated in another way, the chances are 96 out of 100 that the bell will dilate the pupil *more* after training than before training, if the experiment is repeated under the same conditions.

Table II.

Effect of Training Towards Contraction

Subjects	*W*	*X*	*Y*	*Z*
Before Training:				
(*a*) Light:				
Meas.	4	4	4	5
Av.	57.50±1.11	41.25±0.69	41.00±0.00	43.40±0.79
S.D.	3.30	2.05	0.00	2.65
(*b*) Light and bell:				
Meas.	4	4	4	5
Av.	61.00±0.63	43.00±0.77	44.00±0.41	47.40±0.65
S.D.	1.87	2.12	1.22	2.15
Bell, or $b - a$	+3.50±1.28	+1.75±1.03	+3.00±1.41	+4.00±1.02
Training: Bell with contraction of the pupil				
After Training:				
(*c*) Light:				
Meas.	7	4	7	7
Av.	51.11±0.48	42.25±1.08	40.00±0.33	38.85±0.75
S.D.	1.88	3.21	1.31	2.95
(*d*) Light and bell:				
Meas.	7	4	7	7
Av.	49.28±0.50	37.25±0.51	39.90±0.25	36.85±0.87
S.D.	1.98	1.52	0.99	3.42
Bell, or $d - c$	−1.73±0.69	−5.00±1.19	−0.10±0.41	−2.00±1.15
Effect of training, or $(d - c) - (b - a)$	−5.23±1.45	−6.75±1.57	−3.10±0.58	−6.00±1.54
Chances out of 100 that $d - c$ will be equal to or less than $b - a$	99.25	99.81	99.98	99.57

In the above table, the statistical handling of the results was the same as in the case of Table I. The training, however, was in the other direction: towards contraction. For each of the 400 training stimuli, a light was turned *on* and a bell sounded while the pupil was contracting. $d - c$ here is negative, showing that after training the bell contracted the pupil in every case.

direction of contraction after training than before. After this training towards contraction, the effect of the bell was to contract the pupils of all 4 subjects. The conclusion is therefore justified that the training had *some* effect; and the individual reliability of this conclusion is over 99 per cent. in each one of the 4 cases. For all practical purposes, there are 100 chances out of 100 that the conclusion itself is correct; and 98 out of 100 that the same conclusion will be reached again in *all* 4 cases. The results of Table II. are fairly conclusive in themselves. However, it might be urged that negative auditory adaptation and fatigue could account for *some* of the decrease in size of the pupil. No exact measure of the importance of these factors can be obtained. This makes it necessary to consider the results of Table II. in connection with those of Table I. In Table I. the bell caused the pupil to dilate *still more* after the training had been given, and the factors of fatigue and adaptation would not be sufficient to account for this effect. The fact that the pupil could be made to actually contract, as shown in Table II., is positive evidence that the emotional factor does *not* account for the results. Judging only from the results of Tables I. and II., there is a possibility that a *dilatation* of the pupil was not conditioned to the bell, since the emotional factor might account for the findings of Table I. This point is cleared up satisfactorily in later experiments, where we shall have more to say about this emotional factor. There is no doubt at all, however, that a *contraction* of the pupil was conditioned to the bell, because adaptation and fatigue cannot account for any contraction below zero when the bell stimulus is given.

The matters of negative auditory adaptation and fatigue had the *same* opportunity of operating when the training was towards dilatation (Table I.) as when the training was towards contraction (Table II.). After training, in the former case, the pupil dilated still more; after training, in the latter case, the pupil actually contracted. These two factors therefore cannot explain the effects produced.

One further possibility might be considered at this time. During the training towards contraction, the subject may have become accustomed to the often-repeated procedure: the bell ringing just after the light was turned *on*. Also, during the training towards dilatation, the subject may have become accustomed to the bell ringing just after the light was turned *off*. Near the end of either experiment, when the subject was being 'tested' for the conditioned reaction (*i.e.*, the bell stimulus present *without* the change of light), there is a vague possibility that there was an increase in emotional tone, due to this somewhat irregular procedure. I mention this possibility not because I think it is important, but simply because someone might suggest it as an explanation for my results. According to the introspective reports of the subjects, taken at the end of the experiment in each case, they were not thinking about the matter to any extent. I am not willing to admit that this 'emotional factor' is sufficient to explain the results of the experiments where the training was towards dilatation, because the results were far too regular to be accounted for in this way. It is entirely out of the question to suppose that this factor could explain *any* of the results where the training was towards contraction, because in these cases (Table II.) the pupil actually contracted at the end of the experiment!

There is rather conclusive evidence from various sources that the size of the pupil cannot be changed voluntarily. I have tested this matter out in a short experiment, and obtained negative results. One occasionally sees a statement in the literature on the pupillary reflex that certain individuals can voluntarily change the size of the pupil. It is only claimed for very rare individuals, and even these claims are frequently discredited by writers and investigators of the subject. Since I succeeded in conditioning the pupillary reflex in every experiment which was tried, 13 in all, the chances are evidently exceedingly remote that *all* of these subjects were controlling their pupils voluntarily. Weiler, who has done perhaps the best work on the pupil with the serial photographic method, found that 'thinking' caused the pupil to dilate a little, whether the thinking be about dilatation or contraction. For further discussion of this subject, the reader is referred to the original article by Weiler (19, especially 181–182).

In a study which has been made of the connections between the cerebrum and the iris, by the electrical stimulation method, no evidence at all was found for cortical centers in the true sense of the term (3, 42). Magitot holds that there are no cortical centers of dilatation, but only nonspecialized points of departure (12, 241).

Judging from the combined results of Tables I. and II., it is *certain* that the pupillary reflex was conditioned. This is the most interesting consideration just now, on theoretical grounds. There is one further remote possibility, which I shall mention. It is possible that some of the subjects were occasionally 'thinking' about the changes which were taking place in their pupils, during the training series. For example, when the light was turned off, and the bell stimulus given during a dilatation of the pupil, subject *A* might have 'thought to herself' each time: "Now my pupil is dilating!" Later, the bell and the thought-of-dilatation might have become associated; so that eventually the bell was sufficient to cause a dilatation, via the subjective route. According to the introspective reports of the subjects, taken at the end of each experiment, they were thinking about the changes which were taking place in their pupils very little, either during the training series or at the end of the experiment. Perhaps the reader will agree that it is difficult to think about your own pupil changing at *any* time. Near the end of all the experiments, measurements of the pupil and the training stimuli were all mixed up together, and, according to the introspective reports of the subjects, they were thinking mainly about being in the right place at the right time. As further evidence against this supposition, it might be mentioned that subject *Z* in Table II. was a perfectly naïve subject. His pupil was definitely trained to *contract* when the bell stimulus was given. If this subjective factor was at all efficacious, which I do not grant, then the change in the size of the pupil was perchance not conditioned directly to the bell, but to a secondary reaction called out by the bell. If such were the case, the level of the conditioned reaction would be changed, but the fact of the conditioned reaction would not be affected.

In his presidential address before The American Psychological Association, in 1915, Watson made the following remarks about the conditioned reflex in general and about the pupillary reflex in particular: "I have dwelt at some length upon this subject for fear some might advance the view that the conditioned reflex is nothing more than the so-called 'voluntary reaction.' . . . The strongest argument against such a point of view is the fact that it apparently can be set up on processes which are presided over by the autonomic system. To test this we have made a series of experiments having for their object the establishment of a pupillary reflex by the combined stimuli of a very strong light and a sound (bell). We found that the diameter of the pupil under constant illumination with fixation is very steady after the first five minutes; consequently it is possible to make measurements upon the pupil. To ordinary stimulations (sounds, contacts, etc.) there is a slight but not constant change in diameter (at times changes follow evidently upon intra-organic stimulation). But to such stimulation the pupil may respond either by dilation or constriction. In the short time which we had for training subjects we found two individuals in which, after fifteen to twenty minutes' training, the sound alone would produce a small constriction of the pupil in about seventy-five per cent. of the cases. In two subjects no such reflex could be built up in the time we had to devote to them." (18, 98–99). This quotation from Watson is given to show the importance which he placed in securing a conditioned pupillary reaction; and also to give a statement of the work which he did along the same line. His interest in this reflex was purely theoretical, and so also was mine.

The reader may have noticed that the statistical procedure used in Tables I. and II. is not absolutely rigid. For example, the training itself changed the absolute size of the pupil a little; so that it is not entirely correct to compare Diff. $d - c$ with Diff. $b - a$. As it is often stated, a Diff. of 80–70 is *less* than a Diff. of 40–30. At any rate, the chances are against 80–70 being *equal* to 40–30. These changes in level are comparatively small for all subjects except *C*, whose pupil was incidentally trained in both directions, 34 days apart. The 'level' of the pupil, before and after training, at the customary intensity of light but without the bell, is shown as follows:[1]

[1] A more rigorous mathematical procedure would be to express $b - a$ as a per cent. of $(a + b)/2$, and $d - c$ as a per cent. of $(c + d)/2$. If these differences were so

	Subject	Without the Bell	
		Before Training	After Training
Dilatation................	*A*	40.43	38.82
	B	41.33	41.67
	C	65.30	52.00
	D	52.50	54.25
Contraction...............	*W*	57.50	51.11
	X	41.25	42.25
	Y	41.00	40.00
	Z	43.40	38.85

A comparison can be made between the 4 measures of $(d - c) - (b - a)$ in Table I., where the training was towards dilatation, and the 4 measures of the same in Table II., where the training was towards contraction. This will give the difference between the effect of training to contract and training to dilate. We have here two distributions of 4 measures each to deal with. The average is the best measure of the central tendency in these two distributions. For dilatation, the 4 measures of $(d - c) - (b - a)$ are $+ 1.95$, $+ 1.32$, $+ 3.57$, and $+ 5.08$. The average is $+ 2.98 \pm 0.49$. For contraction, the 4 measures are $- 5.23$, $- 6.75$, $- 3.10$, and $- 6.00$. This average is $- 5.27 \pm 0.46$. The difference between these two distributions is 8.25 ± 0.67; and the conclusion is certain that a difference in training will have some effect.

Perhaps enough has been said now to satisfy the reader as to the positiveness of these results. While the absolute

expressed, the reliabilities would be even *higher* than they are given in Tables I. and II. I hope the reader will not be disturbed at this statistical fallacy; because he would be even more disturbed over my descriptive language if I should use the more rigorous statistical procedure. The only real possibility of error which I see in the statistical handling of the results is the case where, in the 4 distributions, *a*, *b*, *c*, and *d*, of Tables I. or II., *a* is correlated with *c*, but not with *d*; or where *b* is correlated with *d*, but not with *c*. The formula for the probable error of a difference assumes that these correlations do not exist. If such correlations did exist, the probable errors of the differences, $b - a$ and $d - c$, are given too small, and the reliability would be less. But I have examined the original measurements very carefully, and find no reason at all for such a belief. The correlation has in fact the same chance of being positive as it has of being negative.

change in the size of the pupil, when evoked as a conditioned reaction, was relatively small, being on the average only about half a millimeter, the conditioned effect was very definite, and was secured in the case of every subject tested. While I do not think that a great deal of weight should be given to *introspections* in an experiment of this particular nature, there was one introspective report in particular which I think is at least worthy of mention. This observation was made by subject *B*, who is a good introspectionist. At the particular time in question, he had just finished the training series. His pupil had been trained to dilate when the bell stimulus was given. The procedure in his case had been to turn the light *off* and ring the bell while the pupil was dilating. The first time I tested for a conditioned reaction, *i.e.*, left the light *on* and rang the bell, he claimed that the intensity of the light increased suddenly the instant the bell began to ring. I observed at the same time that the pupil made a rather marked dilatation. The training stimuli were then repeated several times; and the next time I left the light on and rang the bell he made the same observation the second time, namely, that the light intensity seemed to suddenly increase the instant the bell began to ring. The explanation of this subjective effect is very simple. The bell dilated the pupil. This allowed more light to reach the retina, and caused the light to seem more intense. So it appears that the subject actually observed his own conditioned reaction this time. Although there is a possibility that some other subjective effect, rhythmical effect, or contrast effect may have entered into the situation, I am inclined to believe that the above explanation is the correct one. Subject *F* (Table VII.) also gave a similar introspective report.

My general plan of procedure in the above experiments was to have the nervous impulse from the bell stimulus arrive in the central nervous system at about the same time that the impulse to change the pupil was sent. I wished the auditory stimulus to be *present* when the nervous impulse was sent to the muscles of the iris. It appeared that this would offer the best opportunity for the two impulses to become connected with each other. It would be very interesting to have some very definite information as to the best time for giving the conditioning stimulus, but this was not the main purpose of my experiment and I did not give much time to it. It would be interesting to know whether the bell would produce a greater conditioning effect if it was given a short time before the change in light, or if it was given during the change in size of the pupil; and also whether the bell would produce *any* conditioning effect if it was given after the fundamental stimulus. Although little has been said about this matter so far, data seem to be at hand on the basis of which a fairly definite conclusion may be reached. In the procedure of Table I., light caused a contraction of the pupil, and then the light was turned off and the bell stimulus was present while the pupil was dilating. In these experiments, dilatation became conditioned to the bell. Little if any connection was formed between the bell stimulus and the contraction of the pupil which immediately preceded it. At least it is certain that the dilatation which came *with* the bell had the 'right of way' over the contraction which came immediately *before* the bell, during the training series. For the 4 subjects of Table II., the light and bell were turned on at the same time, and the bell continued to ring for a little over a second while the light stimulus was contracting the pupil. After the pupil was contract*ed* the light and bell were both turned off together, and of course the pupil dilated. It is to be noted that in all 4 of these cases the contraction of the pupil, which came *with* the bell had the right of way over the dilatation which came immediately *after* the bell. From all of this it should be clear that the bell has a better opportunity of becoming conditioned to the pupillary reflex if it is repeated *with* it a large number of times, than if it is repeated *before* or *after* it.

Tables III. and IV. show the results of two experiments with the same subject, *E*, 2 months apart, which were designed to test out this matter further. In Table III., the procedure was to ring the bell 0.5 sec. before and also during the contraction of the pupil. As soon as the pupil had contracted, both light and bell were turned off together, and the pupil of course dilated. A contraction of the pupil, which came 0.5 sec. before and also with the bell had the right of way over the dilatation which came immediately after the bell. I do not think that the very definite results in this experiment should be taken as evidence that the conditioning stimulus functions better when it comes before the fundamental stimulus; because exceptionally definite results were obtained in a later experiment on the conditioned eyelid reaction with this same subject. It appears that conditioned reactions can be established in this particular subject much more readily than in most people.

Table III.

Subject *E*, Jan. 7, 8:55 A.M.—11:50 P.M.

Before Training	(*a*) Light	Meas.	20
		Av.	46.60 ± 0.40
		S.D.	2.62
	(*b*) Light and bell	Meas.	20
		Av.	51.15 ± 0.37
		S.D.	2.49
	Bell, or *b* − *a*		+4.55 ± 0.54

Training: The bell rang for 0.5 sec. before the light was turned on, and continued to ring while the pupil was contracting. So the bell rang for 1.5 sec., and the pupil was contracting during the last second of this period.

After Training	(*c*) Light	Meas.	8
		Av.	46.38 ± 0.60
		S.D.	2.50
	(*d*) Light and bell	Meas.	8
		Av.	44.00 ± 0.62
		S.D.	2.60
	Bell, or *d* − *c*		−2.38 ± 0.86
Effect of training, or (*d* − *c*) − (*b* − *a*)			−6.93 ± 1.01
Reliability			99.99/100

In the above table, the training stimuli were repeated 400 times. This experiment is an effort to test the effect of having the 'conditioning' stimulus (the sound of the bell) come partly before the fundamental stimulus (light). Although the conditioning effect here is very marked, no general conclusion can be reached as to the best time for giving the conditioning stimulus on the basis of this experiment alone.

In Table IV. the procedure was as follows: (*a*) the light was turned on, (*b*) the light remained on until the pupil had contracted, (*c*) as soon as the pupil had contract*ed* the bell was rung for about 1 sec. while the pupil was in the contract*ed* state, (*d*) after this period the light and the bell were turned off simultaneously, and (*e*) the subject kept his same position for about a second while the pupil was dilating.

TABLE IV.

SUBJECT *E*, MCH. 6, 11:40 A.M.—2:50 P.M.

Before Training	(*a*) Light	Meas.	16	
		Av.	46.7	±0.70
		S.D.	4.15	
	(*b*) Light and bell	Meas.	16	
		Av.	47.9	±0.50
		S.D.	2.96	
	Bell, or *b* − *a*		+1.2	±0.86

Training: The light was turned on, causing the pupil to contract. As soon as the pupil had contracted, the bell sounded for 1 sec.; after which the bell and the light were both turned off. Thus at every repetition the bell came *between* the contraction and the dilatation; immediately *after* the contraction, and immediately *before* the dilatation.

After Training	(*c*) Light	Meas.	16	
		Av.	41.7	±0.51
		S.D.	2.05	
	(*d*) Light and bell	Meas.	16	
		Av.	44.6	±0.47
		S.D.	2.81	
	Bell, or *d* − *c*		+2.9	±0.69
Effect of training, or (*d* − *c*) − (*b* − *a*)			+1.7	±1.10
Reliability			85.14/100	

In the above table, the training stimuli were repeated 400 times, as in all former cases. This experiment is an effort to test the effect of having the 'conditioning' stimulus (the sound of the bell) come before or after the fundamental stimulus (light). Apparently a dilatation of the pupil, which in the training came *after* the conditioning stimulus, became conditioned to the bell. No general conclusion, however, can be reached from this experiment alone.

This whole operation completed one trial. So the bell stimulus was present for about 1 sec. *between* a contraction and a dilatation of the pupil. It was present immediately *after* a contraction, and immediately *before* a dilatation. There seemed to be a tendency here for the dilatation, which came *after* the bell, to have the right of way over the contraction which came *before* the bell. I also do not consider that this experiment gives any very reliable evidence in favor of having the conditioning stimulus come before the fundamental stimulus. From the result of Table IV., it is suggested that the bell is not connected with a *state* of the pupil, but rather with a *change* in the size of the pupil. The bell stimulus was present while the pupil was *contracted*, and not while it was contracting or dilating. The bell was not connected with a contraction of the pupil, even though it was present each time the pupil was in a *contracted state*.

To sum up these various results on the question of the most advantageous time to give the conditioning stimulus, it might be stated (*a*) that *every* evidence is found for the forming of a connection when the bell stimulus is present *during* a change in size of the pupil; (*b*) no conclusive evidence is found for an increased effect when the bell comes *before* the change in the size of the pupil; and (*c*) positive evidence is found against a connection being formed when the bell stimulus is given *after* the change in size of the pupil, or, stated another way, against the forming of associations in the backward direction. The results shown in Tables V., VI., and VII. give further evidence in favor of these same conclusions.

Table V. shows the results of an experiment with subject *X*. The light was turned on and the bell stimulus was present while the pupil was contracting. As soon as the pupil had contracted, the light and bell were turned off together, and a telephone receiver sounded while the pupil was dilating. These training stimuli were repeated 400 times. The bell was the same as in former experiments. Current from the secondary coil of an induction apparatus was used to operate the telephone receiver. The receiver was fastened to the table, and made a harsh noise. The bell was not fastened down, and made a 'clear' sound. The 'pitch' of the receiver was a little higher than that of the

Table V.

Subject *X*, Mch. 3, 7:30 P.M.–11:00 P.M.

Before Training	(*a*) Light	Meas.	12	
		Av.	48.9	±0.61
		S.D.	3.15	
	(*b*) Light and bell	Meas.	12	
		Av.	49.4	±0.70
		S.D.	3.61	
	(*c*) Light and receiver	Meas.	12	
		Av.	50.7	±0.72
		S.D.	3.7	
	Bell, or *b* − *a*		+0.5	±0.93
	Receiver, or *c* − *a*		+1.8	±0.94
	Receiver—bell, or *c* − *b*		+1.3	±1.00

Training: Towards contraction (bell), and towards dilatation (receiver).

After Training	(*d*) Light	Meas.	11	
		Av.	44.2	±0.31
		S.D.	1.53	
	(*e*) Light and bell	Meas.	16	
		Av.	42.2	±0.57
		S.D.	3.40	
	(*f*) Light and receiver	Meas.	15	
		Av.	47.1	±0.42
		S.D.	2.36	
	Bell, or *e* − *d*		−2.0	±0.65
	Receiver, or *f* − *d*		+2.9	±0.52
	Receiver—bell, or *f* − *e*		+4.9	±0.71

Effect of training (bell), or (*e* − *d*) − (*b* − *a*)	−2.5	±1.13
Reliability	93.21/100	
Effect of training (receiver), or (*f* − *d*) − (*c* − *a*)	+1.1	±1.07
Reliability	75.65/100	
Effect of difference in training, or (*f* − *e*) − (*c* − *b*)	+3.6	±1.22
Reliability	97.70/100	

For each of the 400 training stimuli of the above table, the bell stimulus was present while the pupil was contracting, and immediately following this, a noise was made in a telephone receiver while the pupil was dilating. Numbers in the table represent the horizontal diameter of the subject's pupil in tenths of a millimeter. The measurements of distributions *a*, *b*, and *c*, and likewise those of distributions *d*, *e*, and *f*, were mixed up together. The effect of training towards contraction, with the bell, and towards dilatation, with the receiver, may here be compared in the same subject.

Table VI.

Subject *B*, Mch. 4, 7:30 P.M.–11:20 P.M.

Before Training	(*a*) Light	Meas.	9
		Av.	40.4 ±0.30
		S.D.	1.34
	(*b*) Light and receiver	Meas.	11
		Av.	42.0 ±0.17
		S.D.	0.85
	(*c*) Light and bell	Meas.	12
		Av.	41.4 ±0.27
		S.D.	1.38
	Receiver, or $b - a$		+1.6 ±0.34
	Bell, or $c - a$		+1.0 ±0.40
	Bell—receiver, or $c - b$		−0.6 ±0.32

Training: Towards contraction (receiver), and towards dilatation (bell).

After Training	(*d*) Light	Meas.	10
		Av.	43.4 ±0.31
		S.D.	1.43
	(*e*) Light and receiver	Meas.	12
		Av.	40.1 ±0.34
		S.D.	1.75
	(*f*) Light and bell	Meas.	10
		Av.	48.2 ±0.60
		S.D.	2.82
	Receiver, or $e - d$		−3.3 ±0.46
	Bell, or $f - d$		+4.8 ±0.68
	Bell—receiver, or $f - e$		+8.1 ±0.69

Effect of training (receiver), or $(e - d) - (b - a)$	−4.9 ±0.57
Reliability	100/100
Effect of training (bell), or $(f - d) - (c - a)$	+3.8 ±0.79
Reliability	99.94/100
Effect of difference in training, or $(f - e) - (c - b)$	+8.7 ±0.76
Reliability	100/100

For each of the 400 training stimuli of the above table, a sound was made in a telephone receiver while the pupil was contracting, and, immediately following, a bell stimulus was present while the pupil was dilating. The general plan of procedure is the same as in Table III., except that contraction is conditioned to the receiver, instead of to the bell, and dilatation is conditioned to the bell, instead of to the receiver. The measurements comprising distributions *a*, *b*, and *c*, and likewise those of distributions *d*, *e*, and *f*, were mixed up together. The effect of training towards contraction, with the receiver, and towards dilatation, with the bell, may here be compared in the same subject.

bell. The results show that the bell actually contracted the pupil after training. The receiver dilated the pupil after training even more than before. An interesting result of this experiment is that the pupil 'distinguished' between the bell and the receiver. It seems probable that 'just *any* stimulus' would not have produced this effect; because contraction was conditioned to the *bell*, and dilatation to the *receiver*. A comparison may be made between the effect of training with the bell and the effect of training with the receiver. This difference is represented by $+ 3.6 \pm 1.22$, and the chances are 97 out of 100 that some difference would be obtained again.

Table VI. shows the results of an experiment with subject *B*, in which the procedure was the same as that used in Table V., except that the stimuli were 'reversed.' These two experiments should be considered together. In Table VI., the light was turned on, and the receiver sounded while the pupil was contracting. When the pupil had contracted, the light was turned off and the bell stimulus was present while the pupil was dilating. The same bell and the same receiver were used here as in Table V. In Table VI., contraction was definitely conditioned to the receiver; and dilatation definitely conditioned to the bell. The 'neural mechanism' again distinguished between these two sounds. The results do not indicate whether it was a difference in pitch or quality or intensity of the sounds which was distinguished. It is probable that all of these elements played a part. The contracting effect showed up more markedly than the dilating effect in both cases. This same tendency, for contraction to show up more markedly than dilatation under the conditions of this experiment, had also been observed in almost all of the other experiments. Apparently the constrictor system is more readily innervated, and the dilator system more readily inhibited. The combined results of Tables V. and VI. remove all possibility of adaptation, fatigue, or the emotional factor explaining the effects produced. The results with subject *B* were more positive than those with subject *X*. This may or may not be due to the difference in procedure.

It is in all probability simply due to an individual difference in the subjects.

The combined results of Tables V. and VI. apparently furnish rather conclusive evidence that a connection was also formed between a *dilatation* of the pupil and the conditioning stimulus. The emotional factor operated just the same when the conditioning stimulus caused a contraction, as when it caused a dilatation of the pupil. Likewise, the 'emotional factor' functioned the same way in either case. A contraction of the pupil was secured *in spite of* the emotional factor; and a dilatation of the pupil was secured *in spite of* the factors of adaptation and fatigue. The effects of adaptation and fatigue were probably always real, and somewhat constant. The emotional factor was not *necessarily* present at all times; and all indications point to the conclusion that it operated irregularly, and *most of the time not at all*. It is hard to believe that the emotional factor played a very important part, when such definite contraction effects were secured in every case. It was obviously much *weaker* than the conditioned reaction which was established. Since the dilating effect had to be obtained in spite of these rather constant factors of adaptation and fatigue, and since it was obtained regularly in *every* case, I am inclined to believe that a real connection was formed between *dilatation* of the pupil and the conditioning stimulus. Further evidence in favor of this same conclusion is found in the next experiment, in which a dilatation of the pupil was conditioned to an electric shock, with a reliability of practically 100 per cent.

Table VII. gives the results of an experiment with subject *F*, in which a contraction, and later in the same experiment a dilatation, of the pupil was conditioned to an electric shock. The electrodes were two cups of salt water, and the subject kept the first two fingers of his right hand in one cup and the last two fingers of his right hand in the other cup. The procedure in the first part of this experiment was to turn on the light and give the electric shock while the

Table VII.

Subject *F*, Mch. 5, 1:30 P.M.–5:00 P.M.

Before Training	(*a*) Light	Meas.	13	
		Av.	39.5	±0.42
		S.D.	2.24	
	(*b*) Light and shock	Meas.	13	
		Av.	40.8	±0.39
		S.D.	2.08	
	(*c*) Light and bell	Meas.	13	
		Av.	40.8	±0.30
		S.D.	1.61	
	Shock, or *b* − *a*		+1.3	±0.57
	Bell, or *c* − *a*		+1.3	±0.52

Training: Shock with contraction. 200 repetitions. Bell not used.

After Training Towards Contraction	(*d*) Light	Meas.	14	
		Av.	39.9	±0.41
		S.D.	2.29	
	(*e*) Light and shock	Meas.	14	
		Av.	36.4	±0.33
		S.D.	1.84	
	(*f*) Light and bell	Meas.	4	
		Av.	40.5	±0.02
		S.D.	0.05	
	Shock, or *e* − *d*		−3.5	±0.53
	Bell, or *f* − *d*		+0.6	±0.41

Effect of training (shock), or $(e-d)-(b-a)=$ .... −4.8 ±0.78
Reliability = ... 100/100

Difference in effect of bell, or $(f-d)-(c-a)=$ .... −0.7 ±0.66
Reliability = ... 76.06/100

Training: Shock with dilatation. 200 repetitions. Bell not used.

After Training Towards Dilatation	(*g*) Light	Meas.	11	
		Av.	39.4	±0.46
		S.D.	1.15	
	(*h*) Light and shock	Meas.	13	
		Av.	45.5	±0.33
		S.D.	1.79	
	Shock, or *h* − *g*		+6.1	±0.57

Effect of training (shock), or $(h-g)-(b-a)=$ .... +4.8 ±0.81
Reliability = ... 99.997/100

Effect of difference in training, or $(h-g)-(e-d)=$ .... +9.6 ±0.78
Reliability = ... 100/100

In the above table, the conditioning stimulus was an electric shock applied to the subject's right hand. In the first training period, the shock stimulus was present while the pupil was contracting. After the first conditioned response had been 'worn out,' the bell was used to test the effect of 'just any stimulus' on the pupil. In the second training period, the shock stimulus was present while the pupil was dilating. A contraction, and later a dilatation, of the pupil was conditioned to the shock, in the same experiment. The measurements of distributions *a*, *b*, and *c* were mixed up together; likewise those of *d*, *e*, and *f*; and also *g* and *h*.

pupil was contracting. After some 200 repetitions of the training stimuli had been given, I found that the conditioned reaction was already fairly well established. So I decided to take some measurements, and these are represented by distributions *d* and *e*. The effect of training towards contraction, with the shock, is given by the figure -4.8 ± 0.78. The reliability of this difference is approximately 100 cases out of 100.

I next called out the conditioned reaction several times, and when it seemed to have temporarily disappeared, I rang the *bell* to see what effect might be produced. I rang the bell 4 different times (distribution *f*), and found that the bell now caused less dilatation of the pupil than at the beginning of the experiment. The difference is represented by -0.7 ± 0.66; which is not very reliable. This difference in effect of the bell might be explained by a difference in a state of the subject caused by the experience of going through the first part of the experiment. The bell did not contract the pupil. It simply dilated it less than it had done before. No particular conclusion can be drawn from this phase of the experiment.

During the latter part of this experiment the training was in the other direction; the light was turned off and the shock stimulus was given while the pupil was dilating. After some 200 repetitions of this compound stimulus, a *dilatation* of the pupil was definitely conditioned to the shock. The results of this whole experiment, obtained with this reversal of procedure in the same subject and at the same sitting, are very definite and convincing. The difference between training towards contraction and training towards dilatation comes to 9.6 ± 0.78; and it is positively certain that a difference in training will produce a different effect. It appears from this experiment that the pupillary reflex can be conditioned to a shock somewhat more readily than to a sound. But I have not worked with shocks enough to be very certain that this is true.

Practically nothing has been said so far about the matter of rhythm. If the pupil is caused to dilate rhythmically, for example, every 15 sec. during the training period, and, furthermore, if the light change (from bright to dark) is now omitted, it might be expected that the pupil would dilate somewhat after 15 sec. had elapsed. Although there is no positive evidence that such a result can be secured in the case of the pupillary reflex, this is a factor which should always be guarded against. Although there was a very decided rhythmic effect (as regards the *time* of giving the stimuli) in the training series, the presence of this factor would not in any way vitiate the conclusions reached. For example, in the procedure of Table I., the rhythmic effect was for the pupil to contract and then to dilate. This rhythmic effect had exactly the same opportunity of operating when the bell was ringing as when it was not ringing. The same thing is true for all other procedures used: this rhythmic effect had very little opportunity to function, and had exactly the *same* chance of operating when the bell stimulus was present as when it was not present.

From the results of the above experiments, it may not be entirely clear whether the bell was connected with a change-in-size-of-the-pupil or with a state of the pupil. In all experiments where the training was towards contraction, the light stimulus was present while the bell was ringing. It is quite possible that the constant light stimulus and the auditory stimulus became 'integrated.' We might say that the two had come by habit to form a 'situation.' *With* the auditory stimulus, the light stimulus is now more at home; *without* the auditory stimulus, the light stimulus does not work so 'smoothly.' Furthermore, the same thing is true of all experiments where the training was towards dilatation. The positive darkness stimulus was always present when the bell stimulus was present. In these cases, the bell stimulus and the darkness may have become 'integrated.' *With* the bell, the darkness now functions with fewer 'disturbances'; *without* the bell, the darkness is 'ill at ease.' This 'integration' may even have been made at a higher level in the nervous system. This interpretation, even if correct, would not influence any of the conclusions which have been reached. In Table IV. the bell was present while the pupil was in a *state* of contraction. This was the experiment in which the bell came immediately after a contraction and immediately before a dilatation of the pupil. The results of this particular experiment furnish evidence *against* the idea of the bell being associated with a *state* of the pupil; but, also, it does *not* furnish any evidence for the assumption that the bell stimulus became associated with a change in the size of the pupil. Evidence for this last assumption, however, is furnished by every other experiment which has been performed.

In the above discussion, I have included the results of *all* experiments which were conducted with the procedure as outlined in the previous section. I have not discussed the somewhat irregular results of some of the 'preliminary' experiments, because my early procedure was not satisfactory. In the first place, the training period was much too short; and, in the second place, I attempted to condition this reflex by giving the training on successive days. The irregular and contradictory results obtained were due in part to insufficient training, and in part to the fact that any association which might have been formed on one day was probably unlearned before the training was renewed on the following day. All indications are to the effect that memory for the conditioned pupillary reaction persists for a very short time. This is no doubt partly due to an active unlearning process. While the results of the preliminary experiments do not deserve a detailed discussion, these experiments do furnish some information as to the best procedure to follow in conditioning the pupillary reflex. The requirements which this procedure should meet, as I see it, are as follows:

1. The training should consist of a great many repetitions of the training stimuli, given continuously.

2. Due attention should be paid to the matters of fatigue and adaptation. The subject should be given frequent rests, to allow the pupil to function properly. Occasionally a rest of a few minutes should be given to allow the subject to relax. There will be less general fatigue of the subject if some schedule is followed rigidly.

3. The conditioning stimulus must not be given except when the fundamental stimulus is present. In other words, no 'exceptions' to the training must be allowed. It is probably advisable also not to allow the pupil to change markedly (in the direction of the training) except when the conditioning stimulus is also present, though there is no conclusive evidence favoring this assumption.

4. Disturbing light and sound stimuli especially must be avoided.

Section 6

Conclusions

The main conclusions which I feel justified in drawing from this experiment are as follows:

1. The pupillary reflex can be conditioned to the sound of an electric bell, to the buzzing sound of a telephone receiver, or to an electric shock.

2. A contraction of the pupil is more definitely conditioned than a dilatation. The evidence is conclusive that a contraction of the pupil was conditioned; and it is practically certain that a dilatation of the pupil was also conditioned.

3. The conditioning stimulus may eventually cause a change in the size of the pupil which can be readily distinguished by the observer. With the procedure used, this change generally amounted to about 0.5 mm. (in horizontal diameter); but it differed widely in different individuals and in the same individual at different times. But even after a long training series, the conditioned reaction disappeared very quickly.

4. A conditioned reaction is readily established when the bell stimulus is given *during* a change in the size of the pupil. Every evidence is found in favor of this conclusion. No very good evidence is found in favor of the supposition that a conditioned reaction may be established when the bell stimulus is given *before* the change in the size of the pupil. This possibility, however, is not eliminated. Positive evidence is found against the forming of a connection when the bell stimulus is given *after* the change in the size of the pupil.

This is to say, no evidence is found for the forming of associations in the backward direction.

5. Positive evidence is found that a change in the size of the pupil can be conditioned to a specific stimulus, and not to just any stimulus affecting the nervous system at the time. In two experiments, the sound of a bell and the buzzing sound of a telephone receiver were distinguished from each other. In one of these experiments, a contraction of the pupil was conditioned to a telephone receiver, and a dilatation of the pupil to a bell; and, in the other experiment, a contraction of the pupil was conditioned to a bell, and a dilatation of the pupil to a telephone receiver. The nervous system did not confuse these two stimuli.

6. Negative auditory adaptation, fatigue, and an emotional reaction are the principal disturbing factors in this experiment. These factors, however, do not vitiate any conclusions which have been reached, because both contracting and dilating effects were produced. These three factors operated in the same fashion when the training was towards dilatation as when the training was towards contraction. If their effects were considerable when the training was towards contraction, then their effects were small when the training was towards dilatation; and vice versa. The procedure used gives a reliable measure of the limitations of these three factors in explaining the results.

In the above discussion and summary of results, no claim has been made that the particular mechanism of the conditioned reaction secured was at all clear-cut or well-defined. While the original stimulus and the end reaction are definitely known, what intervened between the two is partly a matter of speculation. It is not known just what 'secondary' reactions are called out in normal people by an auditory stimulus. It is therefore quite possible that in the present experiment some such reaction or reactions intervened between the innervation of the cochlear nerve and the innervation of the pupillary fibers. The fibers of the pupillary system may have been affected at several different points along their course. There is also good reason to believe that in the training towards contraction and also in the training towards dilatation, the tonus of *both* the sphincter and dilator systems were affected. It may not be wise, therefore, to attempt to localize the particular part of the anatomy where the 'conditioning' took place, because the process was probably diffused and complicated. Still it will not be entirely out of order to make a guess at the structures which probably played a major part in learning this conditioned reaction. Since the bell did not contract the pupil at the beginning of the experiment, and since the bell did contract the pupil at the end of the experiment, it is evident that some terminal axones from the cochlear system have been brought into functional activity with certain cell bodies whose processes go to the muscles of the iris. This is certainly *not* speculation. To explain this result, we simply need some terminal endings from the cochlear nerve situated near some cell bodies of the efferent pupillary fibers. The oculomotor nuclei are suggested at once. This seems to be a likely place where *some* of the conditioning may have taken place, since some of these nuclei have to do with the reflex turning of the eyes to sound, while other oculomotor nuclei are ordinarily involved in contracting the pupil. There may be cross connections in these nuclei, but this is not definitely known. Furthermore, it is likely that a large part of the 'conditioning' was done by the cerebrum, especially the temporal lobe. But these two suggestions may prove misleading, since various secondary reactions may have been produced.

The results of this experiment are very suggestive in view of the fact that the pupillary reflex is not subject to voluntary control, and also in view of the fact that this reflex occurs in the human fœtus of 6 months (12, 241), and in sharks several hours after death (10, 39–40). The experiment, taken as a whole, emphasizes the reflex nature of the human organism. It is generally known that many reflexes are 'nonconscious,' and that various acts may easily become nonconscious with repetition. For example, one quickly learns to write 'the' on the typewriter without thinking of the three separate elements, t, h, and e. It is soon reacted to as a whole. In *learning*, however, it is often assumed in one way or another that conscious processes are *necessarily* involved. In the above experiments, the subjects may have been conscious sometimes of the conditioning stimulus, but they were not conscious of the changes which were taking place in their pupils. Ordinarily, we are conscious of the movement we are learning; but in this experiment the subjects were not conscious of the *act* which was learned.

It would not be correct to assume that a natural 'tendency' for the bell to *contract* the pupil was present already. The only 'tendency' present was for the bell to dilate the pupil. The nerve fibers involved in learning this conditioned reaction were already present, to be sure, but they were not already in functional activity. The experiment is suggestive of the fundamental mechanism of all associations. Just 'fulfill the requirements' of the training period of a given conditioned reaction, and there is no a priori reason for believing that a connection will not be formed. This experiment is suggestive of the way all learning may take place.

Bibliography

1. Bach, Ludwig. Pupillenlehre. Anatomie, Physiologie und Pathologie. Methodik der Untersuchung, 1908.
2. Dontchef-Dezeuze, Mme. Marcelle. L'image et Les Réflexes Conditionnels dans les travaux de Pawlow. Paris, 1914.
3. François-Franck. Cours du Collège de France, p. 42. Quoted from Magitot.
4. Gifford, Harold. An orbicularis pupillary reaction. *Arch. of Ophthalmology*, 1895, vol. 24, p. 402.
5. Gley, E., et Mendelssohn, M. Quelques expériences sur le réflexe salivaire conditionnel chez l'homme. *Comptes-rendus de Soc. de Biol.*, 1915, 78, 645–649.
6. Graef, C. The Pupil in Health and Disease. *Medical Times*, 41, June, 1913.
7. Haab, O. Der Hirnrindenreflex der Pupille (Züricher Festschrift für Naegeli und Koelliker, 1891).
8. Heddaeus, Ernst. Reflexempfindlichkeit, Reflextaubheit und reflectorisch Pupillenstarre. *Berl. Klin. Wochenschrift*, No. 17, 1888.
9. Howell, W. H. A Text-Book of Physiology, 7th ed., 1919.
10. Loeb, Jacques. Comparative Physiology of the Brain and Comparative Psychology, 1900.
11. Lutz, Anton. The Light Pupillary Reflex. . . . *Archives of Ophthalmology*, 47, 1918, 266–294, 370–400.

12. MAGITOT, A. L'iris. Etude physiologique sur la pupille et ses centres moteur. Paris, 1921.

13. SCHIRMER, OTTO. Theoretical and Practical Considerations Concerning the Pupillary Reflex. *The Post Graduate*, 28, 1913, 20–27.

14. SHERRINGTON, C. S. The Integrative Action of the Nervous System, 1920.

15. VERVOORT, H. La réaction de la pupille lors de l'accommodation et de la convergence. *Graefe's Archiv. f. Ophth.*, t. 49, p. 348–375, 1899. *Archives d'ophth.*, mars, 1901.

16. VIDAL, CHARLES-ALPHONSE. Étude sur les Réflexes Pupillares. Paris, Theses, 1900–01.

17. WALLER, A. Physiologie, French ed., 1898.

18. WATSON, JOHN B. The Place of the Conditioned-Reflex in Psychology. *Psy. Rev.*, 1916, 23, 89–116.

19. WEILER, KARL. Untersuchung der Pupille und der Irisbewegungen beim Menschen. *Zeitschrift für die gesamte Neurologie und Psychiatrie*, 1910, 2, 101–274.

THE EFFECTS OF REPETITIONS UPON RETENTION

BY SUGI MIBAI

University of Michigan

At the suggestion of Professor Pillsbury and under his guidance, I repeated a part of the experiments on memory done by Ebbinghaus. Consequently the brief summary of the part of his work is the necessary introduction to my work.

Ebbinghaus,[1] while he was doing very extensive experiments on memory, noticed the fact that as the number of readings increases, the series are engraved more and more deeply and indelibly; if the number of readings is small, the inscription is but surface deep and only fleeting glimpses of the tracery can be caught; with a somewhat greater number of readings, the inscription can, for a time at least, be read at will; as the number of readings is still further increased the deeply cut picture of the series fades out only after ever longer intervals. Then he thought that the inner stability of a series of ideas, or degree of its retainability means the greater or less readiness with which it is reproduced at some definite time subsequent to its first memorization. This readiness is measured by the amount of work saved in the relearning of any series as compared with the work necessary for memorizing similar but entirely new series.

From this conclusion he devised the Savings Method. For material, he used series of nonsense syllables, each series consisting of sixteen syllables. These series are impressed to definite extents as a result of different numbers of readings, and then twenty-four hours later they are learned to the point of the first possible reproduction by heart, and he found how resulting savings in work are related to each other and to the corresponding number of the former readings.

[1] Ebbinghaus—memory 1913.

The first reading was repeated 8, 16, 24, 32, 42, 53, 64 times, and for each, ten tests were made. He worked as subject and experimenter himself. The learning curve in this method is almost a straight line within this limit,—zero on the one hand and on the other 64, which is almost double the number of readings that is sufficient for him to learn the series and beyond which he found himself unable to go, owing to headache and fatigue caused by too long concentration.

Ebbinghaus used such big intervals in numbers of the first readings as 8, 16, 24, and so on, and did not see what would happen between 8 and 16, 16 and 24, and so on, granting that the lines between these points are straight. My experiment is to find whether the line between 0 and 10 repetitions is straight or not.

My material consisted of series of nonsense syllables, each series consisting of twelve nonsense syllables and each syllable consisting of two consonants and one vowel between the consonants, avoiding syllables difficult to pronounce.

Subjects[1] are all trained subjects, four graduate students and two seniors in psychology, four of whom were women and the rest men.

Each series is put on the drum and is shown to the subject one syllable at a time through the window of the screen before the drum, the rate being approximately two seconds per syllable. The subject is told to concentrate on the syllable shown in the window at the time and to try to memorize it but not to make any special associations between syllables or with any thing else. The learner was free to read aloud or silently or to move head or hands as syllables were read. In this way the series is shown twice successively and then there was one hour intermission, during which he is free to do anything but recall the syllables just read. At the end of one hour he comes back to me and recalls as many syllables as he can and I take the records of them. Then the

[1] I wish to express my great appreciation to the subjects who worked so earnestly and patiently in memorizing nonsense syllables for me. If the result of the work is a contribution to psychology, all the compliments should be given to my diligent subjects, Miss Flynn, Miss Gordon, Miss Guernsey, Miss Palmer, Mr. Page, and Mr. Skaggs.

same series is shown in the same way as before but this time it is shown as many times as he needs in order to give the first possible reproduction of the series without a mistake in spelling and in order.

In order not to give more readings than adequate for the first possible reproduction, I let the subject recite the series after every three readings, provided the subject should tell me if he thinks he memorized the series before the three readings were over. I adopted this method because I did not see how a subject could tell whether or not he had just adequate readings for a perfect reproduction if no chance of trial is given, and even when he thinks he memorized all syllables, he may not be able to recite perfectly when he tries, then I must give him more readings, and that trial of reciting will engrave the already acquired syllables and thus influences the number of readings. Some subjects may be too conscientious and do not dare to recite until they feel sure of them, thus they may have more readings than necessary. To avoid this irregularity I took the methods of three readings and one recitation. Later I found that I did not need to fear this irregularity so much because when subjects were practiced in the work they could tell when they memorized the series pretty definitely. But it was too late to change the method when I learned it so I continued to use it all through the experiments for every subject. I took records of numbers of readings and recitations and syllables at each recitation. The time for recitation was not limited.

The same test was repeated with different series each time and with different numbers of first readings. The number of first readings varied as 2, 4, 6, 8, 10. The time of experiment for the same subject was almost constant each day. The place of experiment was my private laboratory where external disturbance was at a minimum.

The results for each individual are gathered in the following table. At the top of each double column is the designation of the worker. Under Rep. for learning is given the number of readings originally made. Under Readings and Recitations stands the number of readings plus attempted recitations required to learn the series after one hour. In the column headed 'saved' is given the difference in number required between the material read twice and the material read four times.

TABLE I

Rep. for Learning	F.		S.E.		S.O.		Pg.		Pl.		S.	
	Readings and Recitations	Saved	Readings and Recitations	Saved	Readings and Recitations	Saved	Readings and Recitations	Saved	Readings and Recitations	Saved	Readings and Recitations	Saved
2	6.9		8.2		11.3		10.2		8.4		14.8	
4	7.3	−0.4	7.9	0.3	9.2	2.1	8.6	1.6	6.2	2.2	9.6	6.2
6	5.5	1.8	5.4	2.5	7.2	2.0	8.0	0.6	6.9	−0.7	10.7	−0.9
8	5.5	0	4.8	0.6	5.2	2.0	7.9	−0.1	5.2	1.7	11.3	−0.6
10	5.9	−0.4	2.8	2.0	5.2	0	7.4	0.5	4.2	1.0	9.7	1.3
No. of series	10		10		6		10		10		7	

As will be seen, the records are for the most part highly irregular. In S. and F. we have individuals who find it difficult to give attention to nonsense syllables, and who object to the work. They are most irregular, but no one shows any approximation to a straight line or to the type of curve reported by Williams. Any attempt to treat them mathematically would not be worth the time expended. Instead we attempted to smooth out the curve by averaging the results from all subjects, bad as that may be in theory. The results give a regularity nearly equal to those obtained by Ebbinghaus, although the combination from so many diverse individual reports is obviously open to objections.

Bringing all the means of numbers of readings in relearning of all subjects together the following table is obtained.

The increases of the numbers of the first readings (Column I.) and of the work saved (Column II.) in each step are in the same ratio, and the number of readings saved for each first reading (Column IV.) is constant except at one point and consequently the curve is almost a straight line.

Sometimes some subjects had more recitations than the proportion of one to three readings and so adding those

TABLE II

I	II		III		IV
No. of First Readings	No. of Readings in Relearning	P.E.	No. of Readings Saved	P.E.	No. of Readings Saved for Every First Reading
2	6.89	.15			
4	5.70	.11	1.19	.13	.30
6	4.91	.12	1.98	.06	.33
8	4.51	.17	2.38	.07	.30
10	3.89	.16	3.00	.13	.30

$M = .31$

recitations to readings will be fairer. The following table will show the result of readings and recitations required together.

TABLE III

I	II		III		IV
No. of First Readings	No. of Readings and Recitations in Relearning	P.E.	No. of Readings and Recitations Saved	P.E.	No. of Readings and Recitations Saved for Every First Reading
2	9.60	.18			
4	8.00	.14	1.60	.16	.40
6	7.09	.17	2.51	.07	.42
8	6.49	.22	3.11	.10	.39
10	5.70	.23	3.90	.18	.39

$M = .40$

In this table the result is almost the same as in the first, with slightly smaller deviations.

Our results indicate that the straight line relation which Ebbinghaus found to hold for the moderate and large number of repetitions holds also for the earlier repetitions, the values below his initial value. Ebbinghaus found that each three repetitions saved approximately one repetition after twenty-four hours. Our values are approximately the same.

REFERENCES

EBBINGHAUS. Das Gedächtniss, 1885.

H. D. WILLIAMS. On the Calculation of an Associative Limen. *Am. J. of Psych.* 29, 219.

151

Journal of
Experimental Psychology

VOL. V., NO. 3 JUNE, 1922

THE CONDITIONED EYELID REACTION

BY HULSEY CASON

COLUMBIA UNIVERSITY

SECTION I

Introduction

The present experiment on the conditioned eyelid reaction should probably be considered in connection with a former one on the conditioned pupillary reaction (4). An effort was made in both cases to eliminate voluntary factors as far as possible. In the pupillary experiment, no measure was taken of the time which elapsed between the reception of the conditioning stimulus and the pupillary reaction. This was hardly necessary since the pupillary reaction is not subject to voluntary control in a strict sense. The pupillary reaction is a relative affair. The eyelid reaction, on the other hand, can be controlled voluntarily to some extent, and is generally 'all or none.' In the present experiment a measure was taken of the time which elapsed between the reception of the conditioning stimulus and the eyelid reaction. The ability of the subject to react somewhat after the manner of the usual reaction-time experiment was also measured. These latter measurements were taken in order to determine the speed with which the subject was *able* to wink when the sound stimulus was given. While the time factor is not a sure criterion for distinguishing a 'reflex,' it is clear that if the time of the conditioned eyelid reaction is made faster than voluntary factors can make it, there is good reason for believing that 'voluntary' factors do not by themselves account for the result.

In the conditioning process, winking was effected by the electrical stimulation method, the 'active' electrode being placed just under the right eye of the subject. These electric shocks produced winks in a very regular manner. The conditioning stimulus was a sharp, sudden sound, of *low* intensity, which was found by trial not to be sufficient to call out a reflex wink at the beginning of the experiment. The subject was given the electric shock and made to wink simultaneously with the sound a large number of times. The sound stimulus was then given alone, without the electric shock, and the time of the wink, if such occurred, was measured. In most cases the conditioned eyelid reaction which was secured was faster than the normal reaction time to the same stimulus. After the sound stimulus and the shock (which caused the wink) had been repeated together a thousand or so times, it was found that the difference between the time of winking to the compound sound-shock stimulus and to the shock stimulus alone was greater than at the beginning of the experiment. The presence of the sound now apparently increased the speed of winking more than it had done before. When the shock was now given alone, there were irregularities in the winking, which were not present at the beginning of the experiment. Apparently some connection had been formed between the sound stimulus and the winking reaction.[1]

Section 2

The Eyelid Reflex

The following is a brief statement of some general features of the eyelid reflex, together with a summary review of the methods which have been employed in investigating this reflex. This will aid perhaps in a better understanding of the various factors encountered in this particular experiment.

Of the two lids or palpebræ of the eye, the upper is the larger and more movable, and has its own levator muscle, the levator palpebræ superioris, which is supplied by a branch

[1] The writer is under many obligations to those who were kind enough to serve as subjects, especially to Miss Eloise Boeker. I wish also to express my appreciation to Professors R. S. Woodworth and A. T. Poffenberger, and to Mr. Arthur L. H. Ruben for much valuable assistance.

of the 3d nerve. The lower lid is not very mobile and lacks a depressor. Closure of the lids is effected by a ring of smooth muscle fibers, the orbicularis palpebrarum, which is common to both lids, and is innervated by the 7th nerve (3, 6).

Closure of the lids may occur: if a very bright light enters the eyes, by the sudden approach of some foreign body, by contact of a foreign body with the lashes, and by irritation of the cornea or conjunctiva (21, 514). "The reflex can be initiated by the stimulation of any of the branches of the ophthalmic (1st) division of the 5th (trigeminal) nerve. From the nucleus of this nerve in the pons Varolii fresh fibers take the impulses, it is believed, to the upper part of the facial nuclei of both sides (7th nerve), and from these to the orbicularis palpebrarum. This reflex is one of the last to be abolished by anæsthetics . . . " (21, 514). It is variously named the palpebral reflex, the conjunctival reflex, or the corneal reflex (17, 12). According to Potter (19, 294), the nerves concerned are the 3d, facial, and sympathetic to the muscles; and the 5th to the skin and conjunctiva.

The wink reflex is one of the most constant and fastest of human reflexes, and it has been studied by a variety of methods. In 1874, Exner (11) measured the time of winking by means of a long, light lever, which was fastened to a frame and held firmly between the teeth. A thread was attached to the lever near its pivoted end. This thread passed over two pulleys and was glued to the eyelid. The lever wrote on a revolving drum. An electrical shock was applied to the lower eyelid to cause reflex winking. The current was varied by changing the distance of the secondary coil of an induction apparatus from the primary coil. Exner found that the reflex could be made faster by increasing the current. With the secondary coil at a distance of 9 cm., the average time was 0.0662 sec., when at a distance of 5 cm., the average time was 0.0578 sec. This time includes the time of transmission of the nerve impulse, the latent period of the muscle, and also the time taken by the nerve cells in the central nervous system. Exner made an estimate of the 'central time,' but it does not seem advisable to give his figures since the speed of nerve transmission and also the latent periods of the muscles involved are not definitely known (16, 35–36; 15). Exner does not state what strength of electric current was used, and this is known to influence the speed of the wink. According to Stiles, "When the lower eyelid receives an electric shock of sufficient strength there is a wink which is quite involuntary. This is a reflex and is accomplished within 0.05 second of the stimulus. With a somewhat weaker shock it is often found that the delay is three times as long. Such a case is a voluntary reaction; there is a sensation, a purpose to wink, and then the movement" (22, 58–59).

Mayhew's method (16) had some points of superiority over that used by Exner. He had a fine copper wire connected with a small piece of platinum foil, which was itself fastened to the eyelid and also insulated from it. On a light aluminium frame fitted around the subject's head was mounted an adjustable bit of platinum wire, in such a way as to touch the foil on the eyelid when the latter was open. A slight movement of the eyelid broke the electrical circuit. The lid reflex was called out by applying a light blow to the face about an inch below and to the outer side of the outer canthus of the eye, on the side from which the record was taken. This was executed by means of a small spring hammer controlled electrically. Mayhew does not say with what force the blow was struck. As the eyelid follows the vertical movements of the eye to some extent, Mayhew had his subjects fixate on a certain object. The record was taken on a kymograph, and timed with a tuning fork. From Mayhew's results of 450 'experiments' on 16 subjects, with the same stimulus but under different conditions, I find the average to be .0413 ± .0001 sec. The S.D. is .005 sec. This average figure represents total time, and is less than that given by Exner. The two experiments, however, were carried out under somewhat different conditions.

Mayhew attempted to test the effect of different mental states on the wink reflex. Some subjects were required to do mental multiplication; some were required to keep the mercury in a manometer at a certain height; while others were warned a few seconds before the blow that it was about to fall. He did not get very regular results. The plotted curves of different subjects did not correspond to one another; but for any given individual there was a close correspondence between the averages of sets of experiments obtained under similar conditions. In those experiments, however, in which apprehension was felt, the average time was found to be shorter (16, 46). The reflex time varied in different individuals, the average times (he does not state for how many cases) falling between 0.0351 and 0.0491 sec.

Garten (13) was the first to use the photographic registration method in studying the wink. He studied the voluntary wink and also the reflex wink. Weiss (25) also used the photographic registration method, but only studied the voluntary wink. Weiss found that the course of a single voluntary wink, in his own case, was divided as follows: closing the eyelid took on the average .060 sec., ranging from .043 to .107 sec.; the lid remained closed an average time of .031 sec., ranging from .014 to .055 sec.; opening the eyelid took on the average .111 sec., ranging from .077 to .160 sec. (cf. 12; and 13).

Dodge (7, 41–45) used his photographic registration method to measure the latent time of the protective wink. The head of the subject was placed in such a position, between an arc light and the photographic recording apparatus, that a shadow of the eyelash fell perpendicularly across the slit. Dodge employed the sound of a clapper on a sounding-board to call out the reflex wink. For 4 of his human subjects, the latent times were 36, 42, 42, and 56σ. Dodge's method is to be recommended wherever very accurate measurements are desired.

Judging from 26 cases, Mayhew concluded that a natural wink, occurring very shortly before a second wink was called out, had no influence (16, 44). According to Sherrington, "In the eyelid-reflex for nearly a full second after initiation of a reflex, the chance that a second stimulus then delivered will, though otherwise appropriate, excite the reflex, is fifty per cent less than it is one second later. The refractory phase, therefore, is marked though not absolute: it operates longer for a visual stimulus than for a tactual or thermal" (20, 45).

Dodge has studied the refractory phase of the wink reaction. He states that "Stimuli from 500σ to 700σ apart gave double reactions in 34% of the cases. Stimuli from 750σ to 1000σ apart gave double reactions in 67% of the cases. Stimuli from 1000σ to 1250σ apart gave double reactions in 100% of the cases" (8, 6; see also 26).

A brief review of Partridge's work will aid us in understanding some further factors which are encountered in this experiment. Partridge (18) studied the ability of normal people to inhibit the reflex wink. Close before the subject's face was placed a framed piece of thick plate-glass. "On the back of this glass and attached to the lower side of the frame was a small rubber-faced wooden-headed hammer which, when released from a catch under the control of the experimenter, was swung suddenly upward, and struck the glass about the level of the eyes of the subject. . . . " (18, 244). In the experiments with the university men, Partridge found the greatest difference in the power to inhibit the wink. In the case of one subject, who was given a series of 20 tests of 50 trials each, on alternate days, the number of winks inhibited each day was as follows: 0, 0, 0, 0, 5, 6, 0, 6, 11, 5, 10, 10, 17, 17, 32, 21, 21, 15, 26, 16, 41, 16, 22, 23, 15, 27, 21, 32, and 35. It seemed to this investigator that a change in the rhythm of releasing the lever or a short interruption in the series lessened the control in the case of this subject. Fatigue also seemed to lessen the power to inhibit the wink. Partridge similarly tested 1100 pupils of the Worcester schools (18, 247). Improvement was more marked and also more uniform for boys than for girls. A boy described by his teachers as 'nervous' was more apt to have difficulty in controlling the wink, but a girl so described was "but little more likely to have difficulty than one described as normal" (18, 249). "Momentary changes in attention frequently inhibited a few winks in a series long before control was finally gained. Smiling seemed to be especially effective in temporarily inhibiting the reflex. Changes of any kind in the environment, as of some one entering the room, opening a door, an unexpected sound, usually increased the tendency to wink, though in some cases such a distraction seemed to have the opposite effect" (18, 250).

Section 3

Electrical Stimulation of Muscles and Nerves

There are a few facts about the electrical stimulation of muscles and nerves which should be kept in mind; and while these are being briefly stated I should like to add along with them certain facts about the procedure which was used in this experiment.

The eyelid reflex was evoked by electrical stimulation, the current being furnished by an induction coil. The 'indifferent' electrode consisted of a cup of salt water, into which the three middle fingers of the subject's right hand were placed. This made a very constant contact. The 'active' electrode, where the effect of the current was more pronounced, consisted of a thin sheet of aluminium, 1.5 by 2.5 cm., placed over the levator labii superioris muscle and about 1 cm. above the zygomaticus major muscle of the subject's face (see the charts of motor points of Erb, 10, 289; Castex, 5, 371; and especially Tousey, 24, 386). The active electrode was held in place by adhesive plaster. I tested out various positions around the eye for this electrode, and it seemed that this particular position was most effective in calling out the wink, although a certain amount of squinting was also present.

The electric shock used in calling out the wink reflex may have stimulated the facial nerve directly, and it may have also had an indirect reinforcing effect through stimulation of the ophthalmic division of the trigeminus nerve. According to Du Bois-Reymond, the effect of such stimulation is proportional to the amperage which actually reaches the muscle or nerve, and to the rapidity of the change in the strength of the current (24, 303). This law is not rigidly correct (14, 130–131), but is useful for most practical purposes. Certain further facts about electrical stimulation have a bearing on our experiment, and they might be very briefly mentioned. The best effect is produced by placing the active electrode directly over the motor point of the muscle. A striated muscle is affected most at the cathode when the current begins, and at the anode when the current is turned off. The wave of stimulation in a nerve starts from the cathode when the current is turned on or its strength is increased, and from the anode when the current is turned off or its strength is diminished. The cathodal closure effect is stronger than the anodal closure effect. In this experiment with the eyelid reflex, the cathode was the active electrode; but it makes little difference when a faradic current is used. The current I employed was not strong enough to stimulate the orbicularis palpebrarum muscle directly, since a period of only about 2σ elapses between the direct application of an electrical stimulus to a muscle and the muscular contraction (24, 329). The time measures I obtained were considerably longer than this figure. If the stimulus is applied to the nerve, in order to affect some muscle, then the time of nerve transmission must be added. Mayhew gives the distance from the eyelid to the center controlling the wink and return as roughly 35 cm. (16, 35–36). The fatigue effect of continued stimulation, which has not been accurately determined, was probably due more to nervous than to muscular fatigue (24, 333). This is especially true when an induced current is used (24, 343). Any reduced response to the successive stimuli is probably entirely due to the phenomenon of fatigue rather than to the stimulation of inhibitory nerves which do not exist in the ordinary voluntary muscles.

Section 4

Apparatus

All time measurements were taken with the Bergstrom chronoscope. This apparatus will not be described since a good description can be easily obtained (see 1 and 2). All measurements were made with the weight at the upper end of the pendulum taken completely off. While this instrument is not as precise as the Hipp chronoscope, it was better adapted to the purposes of the present experiment. The Bergstrom chronoscope had the following advantages over the Hipp chronoscope: (*a*) measurements could be made much faster, (*b*) the disturbing sound of the Hipp chronoscope was not present, and (*c*) the electrical wiring was greatly facilitated.[1]

[1] As a check on the accuracy of this Bergstrom chronoscope, it was decided to connect it with a properly timed Hipp chronoscope and make several measurements on the two simultaneously. For assistance in making the electrical connections I am indebted to Mr. H. K. Nixon. I have also to thank Mr. J. L. Holmes for the use of the Hipp chronoscope which he was using in a reaction-time experiment. The main features of the electrical wiring were as follows:

Pressure on a telegraph key closed the circuit passing through the pendulum release of the Bergstrom chronoscope, starting the pendulum; and at the same time operated a relay. This relay broke the circuit in the upper coils of the Hipp chronoscope, which allowed the armature to fall, and started the clock. The telegraph key was quickly released, so as to put the current back in the upper coils of the Hipp chronoscope. The two clocks were stopped by pressing a second telegraph key. This was used to close the circuit passing through the electromagnets of the Bergstrom chronoscope, which stopped the swinging pointer; and at the same time operated a second relay. This relay broke the circuit passing through the lower coils of the Hipp chronoscope, stopping the clock. In this way both chronoscopes were started and stopped together. The only latent time in the arrangement, which was not ordinarily present in the two chronoscopes, was that of the relays. The armature connections of both relays were normally closed by the light springs. A fraction of a sigma may have elapsed before the inertia of the relay armature was overcome, and before it could be pulled away from the contact. The latent time of the relay in the starting system, however, almost exactly balanced the latent time of the relay in the stopping system. It was not necessary to interfere in the least with the 4-volt current ordinarily used for the Hipp chronoscope; and the currents employed to start and stop the Bergstrom chronoscope were so regulated that they were exactly the same as during the wink experiment.

By this arrangement I started and stopped both chronoscopes 500 times, taking readings on each chronoscope at each trial. I tested out the scale on the Bergstrom chronoscope from 50σ to 1,000σ, taking most of the readings in the range from 80σ to 500σ. The measures showed a constant error on the Bergstrom chronoscope of about 15σ. Its readings were too large when compared with the Hipp chronoscope. I next made a scatter diagram of the 500 pairs of readings, representing each pair by a dot. A curvilinear line was drawn freehand through this very long and narrow distribution, an effort being made to keep the sums of the squares of all the perpendicular distances from the dots to the curvilinear line as small as possible. Then I made a table of the coördinates of this curvilinear line, at intervals of 1σ, and used this table

I used three different sound stimuli at various times. The loudest sound was produced by a telegraph sounder, which was mounted on a box and situated about 6 in. from the subject's head. This sound was intended to call out the wink reflexly, but it did not succeed in all cases. This sound was only used for a few of the first experiments performed. Definite mention is made of the fact in this report whenever it was used, and the measures of the time of winking made with it are referred to specifically as 'sound reflex' times. It was never used as the conditioning stimulus. Another sound stimulus was produced by a second telegraph sounder, located at a distance of some 5 ft. from the subject. The amplitude of this sounder was made less than the other, and the sound intensity thereby decreased. The third sound, which was used in the case of most of the subjects, was the click of a relay, whose armature had an amplitude of some 2 mm. In the case of both the relay and the sounder, one side was deadened so that only a single click, instead of a double one, could be heard.

The device used for registering the time of the wink was a thin aluminium lever attached at one end to the eyelid, and supported just to the right of the subject's right eye by a pair of 'spectacles.' The 'spectacles,' which were made of heavy wire, 3 mm. in diameter, could be bent to conform to different subjects. Once properly fitted, they remained in the same relative position, and were not uncomfortable.

to transmute Bergstrom readings into Hipp readings. The S.D. of the Bergstrom readings, taking the Hipp as standard, was found to be about .004 sec. for that part of the scale with which we are most concerned. The S.D. of the Hipp readings, however, when compared with a fall apparatus, was known to be about .0007 sec. Hence the S.D. of the Bergstrom, if compared with a standard which did not itself vary (*i.e.*, if compared with the 'true') would be *less* than .004 sec., being given by $\sqrt{(.004)^2 - (.0007)^2}$, or .0039 sec. If a time function is now measured by the Bergstrom the obtained S.D. of the distribution will be too high, because the Bergstrom is itself varying while the measurements are being made. If the obtained S.D. of a distribution of wink times, for example, measured by the Bergstrom, is found to be .009 sec., then the true S.D. of this distribution would be given by $\sqrt{(.009)^2 - (.004)^2 - (.0007)^2}$, or .008 sec. This is mentioned to show that my obtained variabilities are, if anything, given too high. I will not make use of this correction, however, in the computations to follow, because I have no good reason for believing that variations in either of these chronoscopes follow the normal curve of distribution. Still, it must be admitted that the *most probable* variability would be obtained only by applying the correction.

The aluminium lever was 0.6 mm. in diameter and 10 cm. in length. One end was drawn to a point and attached to the subject's eyelid by means of a piece of sticking plaster, 2 by 7 mm. The sticking plaster was placed on the subject's right eyelid, just above the free margin. The point of the lever was passed through the central portion of this strip of sticking plaster and back again. There was no danger of the lever coming loose, and very little chance for the sticking plaster to come off of the eyelid. The lever was free to slide to the right and left in a fulcrum which could be adjusted to a favorable position. The lever was very delicate, and so balanced that after a few minutes the subject was no longer conscious that anything was moving with his eyelid. To the outer end of the lever (the end away from the subject's right eye) was attached a piece of German silver wire. The outer end of the lever was about 2 mm. below a second piece of German silver wire when the eyelid was in the normal raised position. Whenever the subject winked the outer end of the lever was raised and was brought in contact with the second piece of German silver wire. This caused an electrical contact to be made, every time the subject winked. The distance between the lever and the stationary wire was maintained practically constant, so that the contact was made with the eyelid always in the same 'phase.' This last adjustment was necessary in order to secure a fairly high degree of accuracy. This apparatus has very distinct advantages over the other devices which have been used. It is very simple in mechanism, and does not often get out of order. It allows the subject to move his head freely, without any interference in the functioning of the device. It can be worn for a continuous period of 4 hours without any particular discomfort. It is very accurate, and enables the wink to be registered when the eyelid is in practically the same position each time.

The electrical contact, which was made when the eyes were closed, was used to operate a relay, which had a very small 'amplitude.' When the relay was closed by a wink the current from 6 dry cells was circuited through the coils of the relay, keeping it closed, and at the same time stopping the pointer on the Bergstrom chronoscope. The pointer was held in the clamped position even though the contact at the eye was almost immediately broken. After noting the time, I would quickly release the pointer by a convenient switch, the purpose being to keep the 6 dry cells from 'running down.'

The apparatus used in registering the time of the wink, the manner of evoking it reflexly, and the various sound stimuli employed, have now been described in detail. Certain electrical connections and combinations of stimuli used at different times are as follows:

1. One side of a double knife switch was employed to release the pendulum, the other side being used to give the shock. Thus the pendulum could be released and the shock given at the same time, and the time which elapsed between the *shock* stimulus and the wink reaction could be measured.

2. Closure of a telegraph key could release the pendulum, produce any one of the sound stimuli, and give the shock, all at the same time. In this way the speed of winking to the *shock and sound* together could be measured.

3. The shock could be cut out of combination 2 by opening a single switch; and the time of winking (if such occurred) to the *sound alone* could be measured.

4. During the 'training series,' which consisted of a great many repetitions of the shock and sound together, I did not care to take the time measurements of every one of the winks to this combination. While no time measurements were being taken the electrical connection to the pendulum was short-circuited, and the pendulum remained in the set position. To measure the time of winking to the sound alone, it was now only necessary to break the circuit involved in the shock stimulus, open the switch short-circuiting the pendulum, and press the key. The time of winking to this sound, if a wink occurred, would represent conditioned reaction time. These switches were manipulated noiselessly, and at irregular times; so that the subject was almost always taken by surprise.

Section 5

Procedure

The procedure was varied somewhat at different times, but the general plan used in the case of the first 7 subjects was about as follows:

1. The wink-registering apparatus was placed on the subject, and properly adjusted. He was next introduced to the shock, which was made just strong enough to call out the wink reflexly.

2. Without giving any instructions I repeated the sound stimulus alone several times. I made sufficient trials to be sure that the sound did *not* cause a wink at the beginning of the experiment.

3. I next alternated with measurements of the time of winking (*a*) to the shock alone and (*b*) to the shock and sound together, taking 20 or 25 of each. I thus had a measure of the effect produced by the shock, both with and without the sound, before the long training period was begun.

4. The training series consisted of from 500 to 3,000 repetitions of the shock and sound together, given at intervals of from 3 to 6 seconds. After this compound stimulus had been repeated from 15 to 35 times (the number was purposely varied, but averaged about 25), I would cut out the shock and give the sound alone one time, to see if a conditioned reaction could be evoked. Proceeding in this way, I obtained several measures of the conditioned reaction time.

5. During the training series, I would also take an occasional measure of the speed of winking to the shock and sound together, for use in comparisons.

6. After the training period was over, and after the measures of the conditioned reaction had been made, I again mixed up measurements of the time of winking (*a*) to the shock alone and (*b*) to the shock and sound together. I had now a measure of the speed of winking to the shock, *with* and *without* the sound stimulus, both *before* and *after* the 'training.'

7. The connection which had been established between

the eyelid reflex and the sound stimulus was 'unlearned' by repeating the sound alone a few times.

8. The subject's voluntary reaction time to the sound stimulus was now measured. He was instructed to listen attentively for the sound stimulus and to wink as quickly as possible when it was heard. No 'ready' signal was given before each reaction. This 'ready' signal was omitted so that the conditions here would be more nearly like those under which the conditioned reaction time was measured. In this way a comparison could be made between the conditioned reaction time and the voluntary reaction time. The voluntary reaction time showed the speed with which the subject was *able* to wink to this same sound. It should be pointed out that no instructions at all were given when a 'conditioned reaction' was called out. In the voluntary reaction time, however, he was definitely instructed to wink as quickly as possible.

9. In a few cases, winks were called out reflexly by the rather loud stimulus, which was produced by a telegraph sounder, mounted on a box, 6 in. from the subject's head. I was trying to make a fair comparison between this 'sound reflex' time and the conditioned reaction time. But it seems that a just comparison cannot be made; since the reflex winks to the loud sound were somewhat irregular for all subjects; and, besides, the two stimuli were themselves *different*.

There is one difficulty with the procedure, as outlined above, which apparently cannot be surmounted. The subject was of course winking voluntarily off and on during the experiment. An effort was made to give the stimuli at times when the subject was not in the act of winking; and also when he had not just completed a wink. Every time the subject winked a relay made a faint sound, distinguishable by me, but not heard by the subject. This was much better I think than looking at the subject's eye all the time. I could tell whenever his eye closed, and gave the stimuli at what seemed to be appropriate times. I could easily avoid giving the stimuli immediately after a natural wink; but could not avoid giving the stimuli sometimes when the wink had already been initiated. If the wink was already 'on the way,' the registered time would be too short; and if a wink had just been completed, the second wink would probably be delayed.

Section 6

Results

Since the detailed procedure was changed from time to time, the results from all subjects cannot be readily combined. The changes in procedure had to do principally with the kind of sound used for the conditioning stimulus, with the number of repetitions of the training stimuli between each measure of the conditioned reaction, and with the particular time of taking the reaction time to the sound stimulus. The results of the different subjects will be described separately, in the order in which the trials were made. This will show the different effects which may be produced by certain changes in procedure.

Table I. gives a summary of the results obtained with subject *J. H.*[1] The main comparison we wish to make in this particular experiment is between the 'conditioned' reaction time and the voluntary reaction time:

	C.R.	R.T.
Av.	173.5 ± 3.07	228.7 ± 10.47
Med.	160.0 ± 3.85	195.0 ± 13.12

R.T.—C.R. (Av.)	+ 55.2 ± 10.91	Reliability	= 99.996/100
" (Med.)	+ 35.0 ± 13.67	"	= 95.73/100

These figures suggest that the conditioned reaction time was faster than this subject's voluntary reaction time. Al-

[1] The number of measures taken in each of the 9 distributions is given in this table, together with the Av., P.E. of the Av., S.D., Q_1, Med., P.E. of the Med., and Q_3, for each distribution. The S.D., which was calculated from the Av., was used throughout in calculating the P.E. of the Med. The medians of the various distributions are given along with the averages, but the average is the better measure in practically all cases both because of the mathematical development of certain formulæ and also because there is no very good reason why the extreme measures in this particular case should not be weighted. The extreme measures were mostly accidental. The median has only a limited meaning; whereas the average is more of a mathematical expression. In all of these computations, I make no correction for the variability in the instruments, in calculating the standard deviations. My variabilities are obviously given somewhat too high in all cases.

Table I

Subject *J. H.*, Jan. 28, 2:30 P.M.–5:00 P.M.

	Sound Reflex (Involuntary)	Shock	Shock and Sound	C. R. (Sound)	Shock and Sound	'Unlearning'	R. T. (Voluntary)	Sound Reflex (Involuntary)	Shock
Meas.	21	21	21	13 (last)	26	15	19	11	11
Av.	127.3±1.25	128.6±1.50	132.6±2.06	173.5±3.07	140.1±1.38	250.9±18.97	228.7±10.47	213.6± 9.78	158.2±3.09
S.D.	8.5	10.2	14.0	16.4	10.4	108.8	67.7	48.1	15.2
Q_1	119	120	121	143	136	173	179	172	147
Q_2	126.0±1.57	128.0±1.88	128.0±2.58	160.0±3.85	140.0±1.73	189.0±23.78	195.0±13.12	206.0±12.26	153.0±3.87
Q_3	134	138	140	170	147	303	290	260	170

The numbers in the above table represent the time in sigma which elapsed between the reception of the stimulus and the winking reaction. The order of the experiment was from left to right, that is to say, the 'sound reflex' times were taken first (1st column), next followed the time of winking to the shock alone (2d column), and so on. The subject did not wink to the 'conditioning' stimulus, the sound of a *relay*, at the beginning of the experiment. The sound reflex (1st column) was called out by a loud telegraph sounder. Column 2 in the table shows the time of winking to the shock alone; while the third column shows the time of winking to the sound of the relay and the shock together. Each conditioned reaction (C.R.), *i.e.*, winking to the sound alone, was preceded on the average by 25 repetitions of the 'training stimuli,' the shock and sound together. The conditioned reaction was evoked 17 times out of 24 attempts. There were 600 repetitions of the training stimuli; and the time of winking to each of these was not measured. To *unlearn* the conditioned reaction, the sound was simply repeated alone several times. The first four measures of the 'Unlearning' time were 166σ, 173σ, 174σ, and 189σ. After this point the reactions were long and irregular. The time required to react voluntarily to the relay was taken without a 'ready' signal: the subject was simply instructed to wink as quickly as possible every time the sound was heard. This subject was not feeling at all well during the experiment.

though there was a general slowing up of all times during the experiment, the reaction time of this subject was probably *faster* at the end than it 'would have been' at the beginning. This matter was tested out in the case of the next three subjects to be considered, where the conclusion is confirmed. This experiment therefore seems to indicate that the eyelid reflex *can* be conditioned to a sound, and that the speed of the conditioned reaction may be made fast enough to rule out the possibility of voluntary reaction. In voluntarily winking to a sound, it is possible that a real inhibitory effect might have been present, due to the subject holding the lid in a 'state of readiness.' That is to say, the subject might have been 'holding' his eyelid in the open position, in waiting for the stimulus, so that a short interval elapsed before he could release this 'hold' and innervate the antagonistic muscles. The *same* consideration, however, would obtain where a measure is being made of the conditioned reaction time, in case the subject did try to react voluntarily to the 'conditioning' stimulus. There is no good reason for believing that this 'inhibitory effect' behaved differently in the two cases,

It seems plausible that the subject might have been in a state of continued apprehension during the training series, due to receiving the electrical shock so many times and in such an irregular way. Some measure of this factor, however, was obtained with later subjects as it does not seem that it is sufficient to explain the results.

The sound reflex time is a hybrid measure, mainly because the loud sounder did not produce a reflex wink in all cases.

Table II. shows the results in full for subject *A. R.* This subject's voluntary reaction (winking) time to the sound stimulus was definitely faster at the end than at the beginning of the experiment. The average reaction time at the beginning of the experiment was 285σ; and at the end it was 203σ. It is probable that the speed of winking voluntarily was continuously improved throughout the experiment, but the exact shape of the curve is not known. There is of course a possibility that the curve took a rather sudden drop when

TABLE II

SUBJECT *A. R.*, JAN. 30, 10:30 A.M.–2:30 P.M.

Sound	R. T.	Shock	Sound and Shock	C. R. (Sound)				Sound and Shock	R. T.			Shock
—	224	126	123	—	—	234	176	126	348	147	152	126
—	211	122	117	—	333	89	209	126	160	296	195	130
—	293	121	127	333	422	72	176	134	143	147	217	130
—	268	114	122	—	224	234	194	122	103	184	217	130
—	186	117	122	—	277	204	184	127	130	151	242	122
—	381	130	125	715	72	—	234	123	143	214	202	130
—	366	122	119	—	—	209	176	—	107	355	155	136
139	122	127	122	—	—	277	138	122	155	160	155	143
—	245	105	122	—	277	107	305	112	151	199	157	135
—	333	118	121	—	234	—	72	122	225	172	252	143
—	352	126	124	—	184	214	348	134	153	240	160	130
—	167	107	120	862	42	296	160	119	167	160	179	—
—	252	118	111	287	—	214	315	117	270	194	151	122
—	651	118	118	—	—	209	204	126	260	168	192	135
—	298	113	119	42	—	292	211	112	210	194	176	—
—	315	112	117	266	199	84	154	111	237	176	168	—
—	324	122	125	395	138	189	—	112	160	216		144
—	196	126	122	—	204	168	234	108	122	224		136
168	224	106	120	209	194	124	103	109	219	204		134
—	297		130	—	84	434	—	112	250	298		130

TABLE II. (*Continued.*)

SUBJECT *A. R.*, JAN. 30, 10:30 A.M.–2:30 P.M.

Sound	R. T.	Shock	Sound and Shock	C. R. (Sound)				Sound and Shock	R. T.		Shock
—	251			—	199	151	147		209	324	
—	239			—	277	126	199		172	184	
—	265			—	42	204	310		191	178	
—	422			324	310	199	168		273	164	
—	250			209	—	199	138		134	170	
				395	260	229	199		144	277	
				42	84	72	385		225	204	
				266	348	348	214		161	324	
				58	—	209	234		160	296	
				72	189	—			168	244	
				—	194	184			127	155	
				255	—	266			155	202	
Meas. . .	25	19	20	79 (last)				19	48 (last)		17
Av.	285.3 ±13.42	118.4 ±1.13	121.3 ±0.51	205.2 ±6.28				119.7 ±1.22	203.0 ±5.09		132.7 ±1.05
S.D.	99.5	7.3	4.0	82.8				7.9	52.3		6.4
Q_1	224	113	119	154				112	162		130
Q_2	265 ±16.82	118 ±1.42	122 ±0.64	199 ±7.87				122 ±1.53	193 ±6.38		130 ±1.32
Q_3	324	126	123.5	234				126	220.5		136

In the above table the experiment proceeded from top to bottom in the columns and from left to right on the page. The conditioning stimulus was a telegraph sounder, which did not call out a reflex wink at the beginning of the experiment. There were 3,125 repetitions of the training stimuli (shock and sound together), given at intervals of from 3 to 10 sec. 125 attempts were made to call out the conditioned reaction; each one being preceded on the average by 25 repetitions of the shock and sound together. A dash in the table indicates that the subject did not react at all within 1 sec.

the electric shocks were first given, so I am on the safe side in using the voluntary reaction time which was obtained at the *end* of the experiment in all comparisons.

This is the only case where the conditioned reaction time is at all *slower* than the voluntary reaction time:

	C.R.	R.T.
Av.	205.2 ± 6.28	203.0 ± 5.09
Med.	199.0 ± 7.87	193.0 ± 6.38
R.T.–C.R. (Av.)	– 2.2 ± 8.08	Reliability = 57.62/100
" (Med.)	– 6.0 ± 10.13	" = 65.50/100

The difference, C.R.–R.T., however, is not reliable. The possibility of a voluntary reaction is not eliminated in this case. It is well to point out that the subject was told to react as quickly as possible when the voluntary reaction time was measured; whereas he was given no instructions whatever when a measure of the conditioned reaction time was taken, and he had no particular 'will to wink' in the latter case. He did not wink reflexly to the sound stimulus at the beginning of the experiment, as the first column of the table shows. The conditioned reaction naturally did not function well at the beginning of the training series (see columns under C.R. in Table II.), but as the training was continued the conditioned reaction became more and more regular, and eventually seldom failed to function. The very marked irregularities in the time of the conditioned reaction can only partly be explained by the fact that the subject was winking off and on during the experiment. There are various factors which would naturally tend to make the conditioned reaction irregular and halting. For example, it is a newly acquired reaction, and the stimuli were given in an irregular manner. These factors, however, do not excuse the rather marked irregularity in the conditioned reaction times, which is for the most part very little less than that of the voluntary reaction time. Our conditioned reaction seems to have little of the regularity and constancy which is very characteristic of some 'reflexes.' It is obviously impossible to treat statistically those trials when the subject did not react at all.

In the case of subject *E. B.*, Table III., the voluntary

TABLE III

SUBJECT *E. B.*, JAN. 31, 10:10 A.M.–1:30 P.M.

Sound	R. T.	Shock	Sound and Shock	C. R. (Sound)			Sound and Shock	R. T.	Shock
—	333	122	130	—	—	234	130	170	123
—	385	122	126	277	—	—	—	—	122
—	336	122	131	—	—	191	138	—	—
—	260	116	138	—	—	—	132	355	140
—	334	118	130	—	234	—	114	—	—
—	195	132	132	—	199	—	130	260	148
—	455	124	136	—	—	233	131	252	142
—	290	124	140	—	—	176	—	—	146
—	284	117	117	310	203	58	113	244	147
—	230	113	135	—	—	—	115	235	145
—	278	130	135	—	—	42	—	296	139
—	261	136	134	—	—	—	117	220	165
—	311	118	130	—	—	—	121	277	121
—	302	116	127	271	—		125	245	155
—	309	122	125	—	—		130	300	144
—	240	122	120	—	42		115	218	143
	266	136	134	—	168			194	124
	199	112	134	—	465			168	122
	349	139	115	—	—			—	161

TABLE III. (*Continued.*)

SUBJECT *E. B.*, JAN. 31, 10:10 A.M.–1:30 P.M.

Sound	R. T.	Shock	Sound and Shock	C. R. (Sound)		Sound and Shock	R. T.	Shock
	232	131	123	—	260		271	—
	227	123	115	—	—		—	168
	326			—	—		258	131
	297			58	—		258	121
	215			255	385		247	
	252			—	—		233	
				—	—		189	
				229	229		266	
				135	—		200	
				—	40		306	
				130	—		265	
				—	42		277	
Meas.	25	21	21	22 (last)		13	25	20
Av.	286.6± 8.00	123.6±1.12	128.4±1.09	182.2±15.66		123.9±1.48	248.2±5.80	140.3±2.20
S. D.	59.3	7.6	7.4	108.9		7.9	43.0	14.6
Q_1	240	118	125	58		115	220	123.5
Q_2	284 ±10.03	122 ±1.40	130 ±1.37	195 ±19.63		125 ±1.85	252 ±7.27	142.5±2.76
Q_3	326	130	134	234		130	271	147.5

In the above table the conditioning stimulus was the click of a relay. This did not call out a reflex wink at the beginning of the experiment. There were 1,875 repetitions of the shock and sound together, given at intervals of from 4 to 10 sec. 75 attempts were made to call out the conditioned reaction, each one being preceded on the average by 25 repetitions of the shock and sound together. The very irregular functioning of the conditioned reaction is shown under 'C.R. (sound).' A dash in the table indicates that the subject did not react within 1 sec.

Table IV

Subject *B. C.*, Feb. 2, 1:00 P.M.–3:00 P.M.

Sound	R. T.	Shock	Sound and Shock	C. R. (Sound)		Sound and Shock	Shock	R. T.
—	184	118	106	—	362	—	—	229
—	632	112	114	208	199	150	161	—
—	348	120	108	199	—	144	—	289
—	320	117	113	164	190	186	161	211
—	216	118	117	160	245	168	179	236
—	345	120	109	234	199	153	144	222
—	397	118	109	126	214	—	203	232
287	172	112	108	143	169	—	—	232
—	303	108	114	155	172	161	168	251
—	153	108	111	174	—	145	144	171
475	355	108	108	235	172	—	142	187
301	199	108	108	204	188	132	166	230
333	194	112	112	251	282	174	150	216
372	210	110	114	160		169	155	216
783	191	117	111	160		149	150	283
—	206	114	108	95		145	149	305
—	349	111	106	197		—	153	156
—	200	116	112	203		174	130	—

Table IV. (*Continued.*)

Subject *B. C.*, Feb. 2, 1:00 P.M.–3:00 P.M.

Sound	R. T.	Shock	Sound and Shock	C. R. (Sound)	Sound and Shock	Shock	R. T.
—	113	108	117	208	173	130	201
—	209	108	110	205	—	139	236
—	218		113	—	—		176
—	253		115	171	170		216
—	215		116	204	152		215
	223		114	—	—		235
	255			240	156		253
				213			206
				40			194
Meas.	25	20	24	35	17	17	25
Av.	258.4 ±14.08	113.1 ±2.08	111.4 ±0.48	192.6 ±6.03	158.9 ±2.27	154.4 ±2.88	223.9 ±4.68
S.D.	104.4	13.8	3.5	52.9	13.9	17.6	34.7
Q_1	199	108	108	164	149	144	206
Q_2	216 ±17.65	112 ±2.61	111.5 ±0.60	199 ±7.56	156 ±2.84	150 ±3.61	222 ±5.87
Q_3	320	117.5	114	213	170	161	236

The conditioning stimulus in the above table was the click of relay, which did not cause reflex winking at the beginning of the experiment. My procedure in measuring the time of winking to the *shock alone* and to the *compound shock-sound stimulus* (columns 3 and 4) was as follows. I would take 1 or 2 measures of the shock time; then 1 or 2 measures of shock-sound time; then 2 or 3 of shock time; then 2 or 3 of the other; and so on, in an irregular manner. The two distributions of measures (columns 3 and 4) were therefore taken under the same conditions, and they may be fairly compared with each other. The same procedure was followed in making the measures shown in columns 7 and 8, so that they may also be fairly compared with each other. During the 'training series' proper, no measures were taken of the time of winking to the compound shock-sound stimulus. While columns 4 and 7 (headed 'Sound and Shock') represent some 'training,' these measures were used mainly in comparisons. When referring to the 'training series,' I always have in mind those compound shock-sound stimuli which immediately preceded the various measures of the conditioned reaction, and whose times were not measured. There were 1000 repetitions of the training stimuli. 40 tests for the conditioned reaction were made, and each was preceded on the average by 25 repetitions of the training stimuli.

reaction time to the sound was *faster* at the end than at the beginning of the experiment. The conditioned reaction was very halting and irregular. The conditioned reaction time, however, considering the last 22 measures, is faster than the reaction time:

	C.R.	R.T.
Av.	182.2 ± 15.66	248.2 ± 5.80
Med.	195.0 ± 19.63	252.0 ± 7.29
R.T.–C.R. (Av.)	+ 66.0 ± 16.70	Reliability = 99.61/100
" (Med.)	+ 57.0 ± 20.94	" = 97.31/100

Table IV. shows the results for subject *B. C.* The reaction time speeds up a little during the experiment. The conditioned reaction functioned from the beginning, and was faster than the voluntary reaction time.

	C.R.	R.T.
Av.	192.6 ± 6.03	223.9 ± 4.68
Med.	199.0 ± 7.56	222.0 ± 5.87
R.T.–C.R. (Av.)	+ 31.30 ± 7.63	Reliability = 99.71/100
" (Med.)	+ 23.00 ± 9.57	" = 94.74/100

A comparison was made in the case of this subject between the speed of winking to the shock alone and to the shock and sound together, both before and after the training series. Before training, the averages are 113σ and 111σ (23). The presence of the sound, after training, seems to slow up the speed of winking a little, though the difference is not reliable. Just the opposite result was obtained in the case of every other subject, where anything like a fair comparison could be made. In all other cases except one this, the presence of the sound seemed to speed up the winking *more* after the training series than before.

Table V. is a summary of the results obtained with subject *D. K.* The conditioned eyelid reaction obtained in this experiment was very unstable, and was quickly lost. The conditioned reaction time, however, was shorter than the voluntary reaction time.

	C.R.	R.T.
Av.	216.9 ± 30.60	288.5 ± 18.89
Med.	151.0 ± 38.35	284.0 ± 23.67
R.T.–C.R. (Av.)	+ 71.6 ± 35.96	Reliability = 91.04/100
" (Med.)	+ 133.0 ± 45.06	" = 97.67/100

TABLE V

SUBJECT *D. K.*, FEB. 5, 4:15 P.M.–6:00 P.M.

	Shock	Sound and Shock	C. R. (Sound)	Sound and Shock (in C. R. Series)	Sound and Shock	Shock	'Unlearning'	R. T.
Meas.	22	19	9 (last)	12	18	10		25
Av.	103.5 ±0.52	102.7 ±0.71	216.9 ±30.60	128.4 ±1.62	124.1 ±1.51	131.3 ±1.56	C. R.	288.5 ±18.89
S.D.	3.6	4.6	136.1	8.2	9.5	7.3	was	140.0
Q_1	100	98	131	125	115	123	quickly	159
Q_2	104.5 ±0.65	105.0 ±0.89	151.0 ±38.35	126.5 ±2.03	124.5 ±1.89	131.5 ±1.96	lost	284.0 ±23.67
Q_3	107	106	359	131.5	132	138		395

In the above table, the conditioning stimulus was the sound of a relay, which did not cause a wink at the beginning of the experiment. The measures of the time of winking to the shock alone and to the compound shock-sound stimulus (columns 1 and 2, also 5 and 6) were taken as described in connection with Table IV. During the training series, occasional measurements of the time of winking to the shock-sound stimulus were made (column 5). Each test for the conditioned reaction was preceded on the average by 25 repetitions of the training stimuli. The conditioned reaction was called out 12 times, out of 20 attempts. This subject was highly sensitive to the shock, and the experiment had to be discontinued sooner than would otherwise have been desired.

In the experiment with this subject, a comparison could be made between the time of winking to the shock stimulus with the sound and without the sound, both before and after training. The *shock and sound* had been given together some 500 times, so it now seemed plausible that the presence of the sound would have some effect on the speed of winking, when the shock was given. This indeed was the case, as the following figures show:

Before Training:	(*a*) Sound and Shock	102.7 ± 0.71
	(*b*) Shock alone	103.5 ± 0.53
	Effect of sound, $a - b$	-0.8 ± 0.886
	$\dfrac{a-b}{\frac{a+b}{2}}$	-0.00776 ± 0.00859
After Training:	(*c*) Sound and Shock	124.1 ± 1.51
	(*d*) Shock alone	131.3 ± 1.56
	Effect of sound, $c - d$	-7.2 ± 2.171
	$\dfrac{c-d}{\frac{c+d}{2}}$	-0.05638 ± 0.01700

Effect of training, $\dfrac{c-d}{\frac{c+d}{2}} - \dfrac{a-b}{\frac{a+b}{2}}$, $\quad -0.04862 \pm 0.019$

Reliability $= 95.74/100$

The chances are 95 out of 100 that $\dfrac{c-d}{\frac{c+d}{2}} - \dfrac{a-b}{\frac{a+b}{2}}$ will be equal to or less than zero. Or, stated in another and perhaps a better way, the chances are 95 out of 100 that, if the experiment is repeated under the same conditions, the sound will have more of a speeding-up effect on the wink time (called out by the shock) *after* the training series than *before*. While this conclusion seems to be correct as stated, it does not logically follow that it may be taken 100 per cent. as evidence of a conditioned reaction. Still it seems difficult to explain the matter any other way. Negative adaptation and fatigue would apparently operate to make the effect of the sound *less after* the training than *before*. It seems that this result has been attained *in spite of the effects of adaptation and fatigue*. I think it may be safely stated that the shock → wink situation has been definitely 'associated' with the sound stimulus, so that it now functions more 'smoothly' when the conditioning (sound) stimulus is present.

One further possibility should not be overlooked. This is the possibility that adaptation to the shock proceeds faster than adaptation to the sound. If such were the case, then the calculations mentioned just above would be considerably vitiated. However, the changes in procedure at different times afford some data on the basis of which the conclusion may be fairly reached that adaptation is faster to the *sound* than to the *shock*. The results of Tables I., II., and III. show clearly that adaptation to the shock (as measured by the speed of winking) is rather small. The results of these three experiments show that the shock continues to call out a wink readily, after a very large number of repetitions. In contrast with this, Tables VIII. and IX. show that the adaptation factor in a reflex wink (to a sound) is fairly rapid. In column 1 of Table VIII., it is to be noted that the sound ceases to call out a wink reflexly after only 26 repetitions. In the 1st 3 columns of Table IX., the reflex (?) winking rapidly becomes very irregular. I made an effort to call out winks reflexly by a very loud sound in the case of 6 subjects. It did not work satisfactorily for more than a dozen or so repetitions! There is no reason at all for *assuming* that adaptation to the shock is greater than adaptation to the sound. As a matter of fact, *the reverse seems to be true*, and the conclusion that "the sound has more of a speeding up effect on the wink time (called out by the shock) after the training series than before" seems to be quite justified.

One possible factor in the 'fatigue' to the shock was the matter of electrical polarization, which was of course present in every experiment (see 24). Although I did not obtain a quantitative measure of the effect of the electrical polarization, its effect (on the speed of winking to the shock) was doubtless less than that of the general fatigue factor. That is to say, the polarization factor was probably less efficacious than the fatigue of the muscles, end plates, nerve cells, and so on.

One objective measure of the total fatigue is shown in the various tables of results in the increased time of winking to the shock.

TABLE VI

SUBJECT *E. B.*, FEB. 7, 10:00 A.M.–1:00 P.M.

	Shock	Sound and Shock	C. R. (Sound)	Sound and Shock (in C. R. Series)	Sound and Shock	Shock	R. T.
Meas.	24	24	C. R.	10	30	29	38
Av.	118.4±1.20	115.8±0.95	was	128.6±2.07	118.7±0.86	125.6±1.00	218.8±10.57
S.D.	8.7	6.9	not	9.7	7.0	8.0	96.6
Q_1	109.5	110	secured	120	112	118	155
Q_2	117.0±1.50	114.5±1.19		126.0±2.59	117.0±1.08	126.0±1.25	199.0±13.25
Q_3	126	122		138	127	128	273

In the above table, the conditioning stimulus was the sound of a relay, which did not cause a wink at the beginning of the experiment. The measures of the time of winking to the shock alone and to the compound shock-sound stimulus were taken as described in connection with Table IV. Each test for the conditioned reaction was preceded by 50 repetitions of the training stimuli. The training stimuli were repeated 1200 times, but no conditioned reaction at all was obtained.

Tables III. and VI. give results which were obtained with the same subject, *E.B.* In the case of Table III., the conditioned reaction was very irregular; and in the present trial (Table VI.) no conditioned reaction at all was established. But in the present experiment some conditioned effect *may* have been present, even though it was not strong enough to call out a wink reaction. Considering shock and sound-shock times, the *presence of the sound* in this case had a speeding-up effect of $2.6\sigma \pm 1.52\sigma$ before training, and a speeding-up effect of $6.9\sigma \pm 1.32\sigma$ after training. The difference between the two, or the *effect of training*, is represented by the figure, 4.3 ± 2.01; and the chances are 93 out of 100 that, if the experiment is repeated under the same conditions, the presence of the sound will speed up the time of winking more *after* the training series than *before*. Several time measures of the sound-shock stimuli were made in the course of the training series; and these figures give a fair indication I think of a lower limit to be expected for the conditioned reaction time.

Table VII. is a summary of the results for subject *E. G.* One peculiarity in this experiment, which is not shown in the table, is the fact that the conditioned reaction functioned slightly better near the beginning of the training series than at the end. It did not function regularly at any time, but in the 13 cases where it did function the time is rather fast.

	C.R.	R.T.
Av.	122.3 ± 16.22	244.5 ± 10.10
Med.	104.0 ± 20.33	222.5 ± 12.66

R.T.–C.R. (Av.)	122.2 ± 19.11	Reliability = approx. 100/100
" (Med.)	118.5 ± 23.95	" = " 100/100

In this case, the average of the conditioned reaction time (122.3σ) came near reaching the time of winking to the sound and shock together in the C.R. series (110.8σ). A comparison of the shock and sound-shock times before and after training shows that the presence of the sound increased the speed of winking (to the shock) $0.4\sigma \pm 1.23\sigma$ before training, and $5.0\sigma \pm 1.94\sigma$ after training. The effect of

Table VII

Subject *E. G.*, Feb. 7, 3:30 P.M.–5:30 P.M.

	Shock	Sound and Shock	C. R. (Sound)	Sound and Shock (in C. R. Series)	Sound and Shock	Shock	R. T.
Meas.	28	28	13 (all)	20	24	19	30
Av.	115.6±1.01	115.2±0.71	122.3±16.22	110.8±1.57	118.4±0.99	123.4±1.67	244.5±10.10
S.D	7.9	5.6	86.7	10.4	7.2	10.8	82.0
Q_1	111.5	111	62	98.5	115.5	112	200
Q_2	113.5±1.27	115.0±0.89	104.0±20.33	111.0±1.97	120.5±1.24	125.0±2.09	222.5±12.66
Q_3	122	119	135	118.5	124	135	283

In the above table, the conditioning stimulus was the sound of a relay, which did not cause a wink at the beginning of the experiment. Each of the 31 tests for the conditioned reaction was preceded (on the average) by 25 repetitions of the training stimuli. The training stimuli were repeated 775 times in all. The conditioned reaction functioned only 13 times, out of the 31 trials. The time measure of the conditioned reaction above represents the time for those cases when it did function.

TABLE VIII

SUBJECT *M. W.*, FEB. 7, 8:00 P.M.–10:45 P.M.

Sound (Loud)	Shock	Sound and Shock	C. R. (Sound)	Shock Alone	C. R. (Sound)	Shock Alone	Sound and Shock (in C. R. Series)	Sound and Shock	Shock	'Unlearning'	R. T.
—	99	102	106	104	113	101	102	89	108	230	252
—	98	88	—	90	193	—	103	102	112	—	352
103	95	99	102	97	112	110	106	104	104	200	284
106	89	89	211	109	175	305	91	99	106	223	257
95	89	99	—	148	176	161	88	98	103	511	264
78	90	85	—	—	202	111	89	103	103	305	138
80	99	86	—	110	146	113	88	99	209	349	321
83	101	84	—	115	123	122	97	105	105	478	320
98	99	85	163	147	176	114	102	93	106	278	348
99	98	89	—	—	167	102	106	84	422	194	335
93	97	84	—	78	218	114	102	108	266	209	382
100	102	84	—	85	172	113	94	91	213	98	395
—	97	89	88	92	214	122	98	96	112	196	198
122	96	94	—	95	198	287	100	105	105	245	245
97	103	84	36	94	210	106		105	229	275	221
89	98	86	28	110	168	116		91	106	199	262
106	102	98	—	77				99	212	247	258
104	96	90	111	122				99	362	236	402
99	96	92	144	112				84	96	461	197
100	97	88	174	109				101	355	135	419
100	99	83	143	101				93	106	497	402
95	95	83	105	109				101	103	333	350
103	96	91	171	130				96		542	220

TABLE VIII. (*Continued.*)

SUBJECT *M. W.*, FEB. 7, 8:00 P.M.–10:45 P.M.

Sound (Loud)	Shock	Sound and Shock	C. R. (Sound)	Shock Alone	C. R. (Sound)	Shock Alone	Sound and Shock (in C. R. Series)	Sound and Shock	Shock	'Unlearning'	R. T.
89	99	89	190	109				105		144	316
97	93	93	113	113				97		113	152
125	94	94	126	108				95		—	423
98	98	96	134	110				94		—	
122	94	92	184	109				104		—	
113	98	89	189	99				98		293	
—	99	94	223	104				94		—	
—	103	80	169	106				104		—	
132	97	94	128	102				106		—	
—	95	88	105	103				95		155	
—	94	92	146	103				101		—	
—	97	90	106	101				105		—	
108		92	156	83				105		725	
—		89	155	107						744	
—		95	422	113						520	
475		94	305	120						531	
Meas. . 26 (1st)	35	39			36 (last)	36 (last)	14	36	22	25 (last)	26
Av. . . . 99.8±1.52	96.9±0.39	90.0±0.54			172.4±6.63	120.8±4.96	97.1±1.13	98.6±0.67	170.1±14.08	327.9±24.39	296.7±10.67
S.D. . . 11.5	3.4	5.0			59.0	44.1	6.3	6.0	97.9	180.8	80.7
Q_1 95	95	86			131	103.5	91	94.5	105	196	245
Q_2 . . . . 99.0±1.90	97 ±0.49	89 ±0.68			170 ±8.32	109.5±6.22	99 ±1.42	99 ±0.84	107 ±17.65	275 ±30.51	300 ±13.37
Q_3 104	99	94			191.5	114	102	104	213	478	352

In the above table, the conditioning stimulus was a loud telegraph sounder, which caused winks at the beginning of the experiment as shown in column 1. There were 1,650 repetitions of the training stimuli. The measures of 'C. R. (Sound)' and 'Shock Alone,' in columns 4 and 5 (*continued* in columns 6 and 7) were taken along with each other (the same procedure as described in connection with Table IV. 55 attempts were made to call out the conditioned recation, and it functioned 45 times.

training is given by $4.60\sigma \pm 2.30\sigma$, and the conclusion that the sound speeds up the winking more after the training than before is true with a reliability of 91 cases out of 100.

Table VIII. gives the complete results for subject *M. W.* In this particular experiment I wished to try out the general effect of a sound which at the beginning of the experiment *did* cause a reflex wink. Column 1 shows the effect of this sound in calling out a wink reflexly at the beginning of the experiment. After some 26 repetitions of the sound, it no longer called out a wink reflexly. Measurements of the time of winking to the shock alone and to the shock-sound stimulus were made as described in connection with Table IV. These two distributions have very small variabilities. As a further variation in the procedure, I took measures of the shock time, designated 'shock alone,' during the training series (columns 5 and 7). I also took the usual measures of the time of winking to the compound sound-shock stimulus during the training series (column 8). The sound stimulus not only lost its power of evoking a wink reflexly after 26 repetitions at the beginning of the experiment (column 1 of the table), but also did not call out a wink even after the first 400 or so repetitions of the training stimuli had been given. The sound stimulus did not begin to call out a wink regularly again until about 500 repetitions of the training stimuli had been given. This winking was a *conditioned reaction.* It was secured *in spite of* the auditory adaptation or the fatigue factor plainly evident at the beginning of the experiment. Whatever factor or factors caused the eyelid reflex to *cease* when the sound was repeated at the *beginning* of the experiment were now in all probability *still more* active. But the tendency towards the forming of a connection between winking, caused by the shock, and the sound, was probably *stronger* than the above factor, and had the 'right of way' to such an extent that the conditioned reaction did not fail to function a single time during the last 38 trials! It cannot be claimed that this is a reflex wink caused naturally by the sound, because it has already been demonstrated with even stronger stimuli that this functions irregularly, and most of the time not at all. The 'unlearning' of this connection is shown in next to the last column of the table.

A comparison between the voluntary reaction time and the conditioned reaction time to the *same* sound shows that the latter is definitely faster.

	C.R.	R.T.
Av.	172.4 ± 6.63	296.7 ± 10.67
Med.	170.0 ± 8.32	300.0 ± 13.37

R.T.–C.R. (Av.)	+ 124.3 ± 12.56	Reliability = approx. 100/100
" (Med.)	+ 130.0 ± 15.75	" = " 100/000

Any factors which tended to make the voluntary reaction time *slow* had exactly the same chance I think of making the conditioned reaction times *slow*. The voluntary factors seem to be definitely ruled out in this case, because of the *speed* of the conditioned reaction.

A comparison between the shock time and the shock-sound time shows that the presence of the sound increased the speed of winking (to the shock stimulus) $6.9\sigma \pm 0.67\sigma$ before training, and $71.5\sigma \pm 14.09\sigma$ after training. The effect of training was $64.6\sigma \pm 14.11\sigma$, and the conclusion that the training had some effect is true with a reliability of 99 out of 100. The measures of *shock-sound* times slow up a little during the experiment. They are 90.0σ, before training; 97.1σ, during training; and 98.6σ, after training. The variability remains very nearly the same: the S.D.'s being 5.0, 6.3, and 6.0. The measures of *shock* alone, on the other hand, slow up much faster. They are 96.9σ, before training; 120.8σ, during training; and 170.1σ, after training. The variability increases rapidly; the S.D'.s being 3.4, 44.1, and 97.9. The conclusion which may be fairly drawn from this unusual behavior is that the sound and the wink were 'connected' with each other, so that immediately after the training series was over the presence of the sound helped the shock to call out a wink very regularly (S.D. = 6.0), whereas the absence of the sound now caused marked irregularities in the distribution (S.D. = 97.9, see third to last

Table IX

Subject *M. A.*, Feb. 11, 9:15 A.M.–12:00 M.

Sound (Loud)			Shock		Sound and Shock	C. R. (Sound)	Shock Alone	Sound and Shock (in C.R. Series)	Shock	Sound and Shock	'Unlearning'	R. T.
—	—	305	107	106	95	132			104	143	156	148
—	508	—	104	105	93	144			199	84	117	272
379	309	—	102	115	85	125			105	97	163	450
—	379	440	108	105	88	77			74	106	—	226
349	—	—	107	115	92	104			99	97	205	281
604	296	—	99	103	99	149			102	93	278	460
—	—	199	91	96	102	45			116	99	—	210
—	—	—	97	123	102	253			106	94	255	256
198	419	—	105	129	95	113			107	110	311	277
320	652	324	113	105	84	130			95	106	207	247
222	—	—	100	115	96	132			224	122	143	252
—	764	—	125	118	103	192		99	122	96	225	342
298	225	508	145		103	103		107	257	115	388	232
571	—	—	97		93	260		80	102	103	149	253
—	—	—	136		94	116		85	193	96	226	324
471	—	804	97		102	124	95	97	—	115	112	144
446	265	—	109		96	139	110	106	117	97	514	199
316	305	—	144		97	150	72	99	475	126	144	271
376	442	—	117		106	176	130	113	485	105	134	294
188	382	—	113		95	343	229	109	205	105	162	565
—	—	589	111		92	155	104	108	642	113	188	191
349	403	298	125		85	51	103	103	116	104	240	169
322	359	—	118		104	34	122	94	123	118	205	196
278	—	—	109		106	121	215	102	—	95	183	171
356	—		134		105	189	103	101	783	114	411	—
—	—		111		108	111	125	108	—	125	107	268

TABLE IX. (*Continued.*)

SUBJECT *M. A.*, FEB. 11, 9:15 A.M.–12:00 M.

Sound (Loud)		Shock	Sound and Shock	C. R. (Sound)	Shock Alone	Sound and Shock (in C.R. Series)	Shock	Sound and Shock	'Unlearning'	R. T.
509	—	99	99	154	122	109	114	99	194	155
249	302	109	99	253	108	121	465	131	220	
233	—	108	110	112	102	99	135	124	—	
410	—	122	103	122	110	112	127	127	—	
329	—	132	94	164	244	107	116	113	349	
508	—	129	97			110	97	97	237	
297	—	127	95				116	119	148	
373	319	74	84				194	—	385	
328	—	106	94				246	110	—	
329	—	108	106				196	101	289	
395	—	105	106				107	111	220	
—	—	125	110				138	95	210	
Meas. . .	24 (last)	50	38	31	16	21	35	37	25 (last)	26
Av.	408.2 ±21.38	112.1 ±1.30	97.8 ±0.78	144.3 ±2.47	130.9 ±8.31	103.3 ±1.34	200.1 ±18.76	108.2 ±1.41	231.3 ±15.97	263.6 ±13.03
S.D.	155.3	13.6	7.1	20.4	49.3	9.1	164.5	12.7	118.4	98.5
Q_1	303.5	105	94	112	103	99	106	97	149	196
Q_2	369.0 ±26.8	109 ±1.63	97 ±0.98	132 ±3.1	110 ±10.41	106 ±1.68	117 ±23.51	106 ±1.77	210 ±20.02	252.5 ±16.33
Q_3	475	122	103	164	111	109	205	115	240	281

In the above table, the conditioning stimulus was a telegraph sounder, which called out winks at the beginning of the experiment as shown in the first three columns. Each of the 31 measures of the conditioned reaction time, and also each of the 16 measures of the time of winking to the shock alone (column 8) was preceded on the average by 25 repetitions of the training stimuli. There were 1,175 repetitions of the training stimuli in all.

column)! After a certain point in the training series the conditioned reaction time actually slowed up somewhat. This may or may not have been due to fatigue.

The complete results obtained with subject *M. A.* are shown in Table IX. These results are of the same general nature as those just described in connection with Table VIII.; and the same general trends are again positively indicated. The subject winked quite often to the sound stimulus (which was not as loud as that used for the 'sound reflex' time earlier), at the beginning of the experiment, but the winks were delayed and irregular. The time of a *reflex* wink to sound, which has been discussed in section 2, is exceptionally fast. It seems therefore that the long times, as shown in the first three columns of Table IX., could not properly be called 'reflex' times. Very few of these times are less than 200σ, and a 'reflex' wink to sound is certainly less than 100σ. This same sound was *repeated* some 1,200 times in the course of the experiment; and there is every reason to believe that the slight tendency to wink reflexly to this sound, present at the start, became weaker as the experiment proceeded. Certainly the tendency would not become stronger. A comparison between the voluntary reaction time and the conditioned reaction time shows the former to be *definitely* slower.

	C.R.	R.T.
Av.	144.3 ± 2.47	263.6 ± 13.03
Med.	132.0 ± 3.10	252.5 ± 16.33

R.T.–C.R. (Av.)	+ 119.3 ± 13.26	Reliability = approx. 100/100
" (Med.)	+ 120.5 ± 16.62	" = " 100/100

It is noticeable that the conditioned reaction functioned perfectly from the very beginning of the training series (*i.e.*, there were no 'blanks' in the series), and that the time of the conditioned reaction is *very fast*. The average conditioned reaction time is 144.3 ± 2.47σ, while the median time is 132.0 ± 3.10σ. This time is comparable to the speed with which a *trained* subject can react to a sound by raising the finger from a telegraph key, in the usual reaction time experiment. Reacting with the finger in this way is also known to be faster and more constant than reactions made with other parts of the body.

A comparison between the effect of the sound on the speed of winking to the shock, before and after the training series, shows that the sound had a speeding-up effect before training of $14.3 \pm 1.51\sigma$, and that after training the sound had a speeding-up effect of $91.9 \pm 18.81\sigma$. The effect of training is given by $77.6 \pm 18.87\sigma$; and the chances that this difference is equal to or greater than zero are 99.71 out of 100. But if the medians of the distributions are considered, instead of the averages, very different results will be obtained.

Considering the results of all these experiments together, the conclusion is certain that the conditioned reaction time *may* be made faster than voluntary factors alone can make it. The conditioned reaction was established when the conditioning stimulus (sound) and the fundamental stimulus (shock) were given at the *same* time.

A very simple experiment was arranged to test the effect of giving the conditioning stimulus a short time *after* the wink occurred. A telegraph sounder was connected with the wink apparatus, so that every time the subject (*H. S.*) winked naturally, the sound was given automatically. The subject was allowed to wink naturally for a continuous period of 5 hours. The sound stimulus was given a short time *after* the impulse to wink had been sent each time, in fact when the natural wink was practically completed. The subject surely winked several thousand times during this period. The apparatus was so connected that I could give the same sound stimulus myself at any time, and measure the time of winking, if such occurred. I gave the sound stimulus myself some 40 times during the latter part of this 5-hour period. The sound did *not* cause a wink at any time. No connection was formed between the wink reaction and the sound stimulus which, during the 5-hour 'training period,' had been given after the natural impulse to wink had been sent each time. This experiment furnishes some further evidence I think against the forming of associations in the backward direction.

Very little has been said so far about introspections. Although our most important results do not depend on introspective reports, the introspections which were taken are worthy of consideration. After a conditioned eyelid reaction had been called out, I would occasionally ask the subject about this conditioned wink. Whenever the time of the wink was at all fast, the subject would *insist* that he was not *aware* of the wink until *after* it had occurred. This was the introspection of every subject where the conditioned reaction was established. Voluntary factors apparently were not the 'cause' of the conditioned reactions. I asked several subjects about any 'images' and 'affective elements,' which might have been present, and they were unanimous in affirming that no such were hanging around, at least not *before* the wink occurred. This is offered as evidence against Mme. Dontchef-Dezeuze's interpretation (9) of the conditioned reaction in terms of images and affective elements. I insisted in the case of some of the subjects that they *should* have images and affective elements *with* the conditioned reaction,—but they were unable to get them even then! I think that this may also be taken as evidence that images and affective elements did not play a *conspicuous* rôle in these conditioned reactions.

The other introspective evidence which I wish to mention is as follows. At the beginning of the second experiment with subject *E. B.* (Table VI.), I requested *E. B.* to judge whether the shock coming alone was stronger or weaker than the shock which was present in the compound shock-sound stimulus. The strength of the electric shocks were as a matter of fact exactly the same in both cases, and this subject judged them to be of equal strength at the beginning of the experiment. But after the compound shock-sound stimulus had been repeated together some 1,200 times *E. B.* insisted that the shock coming alone was decidedly *weaker* than the shock which was present in the shock-sound stimulus, and yet only one shock was used throughout this experiment. When the shock stimulus was now given *alone* (see next to last column, Table VI.), this subject would say: "I got the shock that time all right, but it was much weaker than usual. Something must be the matter!" The shock coming by itself felt 'weak,' 'creepy,' 'strange,' 'lacking in something'; and it always seemed to contain an element of surprise. This introspective report corresponds somewhat with the objective result, *i.e.*, in respect to the slower time of winking to the shock alone, after the training series. All later subjects were questioned on the same matter, and their introspective reports agree very closely with each other. While the shock alone seemed queer to all of these subjects (after the 'training' had been given), the sound alone seemed quite 'natural' at all times. The explanation of this matter is not very clear to me. This is a rather interesting observation, however, especially in view of the fact that the *speed* of winking to the shock alone or to the shock-sound stimulus is very much faster than the voluntary reaction time. These considerations would suggest I think that the conscious aspects in these phenomena were 'results' or 'by-products' of the *process*, as the process moved along.

At the beginning of all of the above experiments, with the exception of one, the sound stimulus was not effective in causing a reflex wink. The sound would have been an effective stimulus, however, even without any conditioned reaction training, if it had been made intense enough. At the beginning of an experiment, the nervous connections between the sound and the winking reaction were already present, but the threshold was too high for the sound to be effective. In those experiments in which a definite conditioned reaction was obtained, this threshold was *lowered*: sound had now been made an effective stimulus.

Section 7

Conclusions

1. The conditioned reaction time, or the time of winking when the sound stimulus is given alone, may be made so much faster than the voluntary reaction time, or the speed with which the subject is able to wink when the same sound

is heard, that voluntary factors in themselves do not explain the result. In the case of one subject (*M. A.*) the average *conditioned* reaction time was 144.3 ±2.47 sigma, and the median conditioned reaction time was 132.0 ± 3.10 sigma. His average *voluntary* reaction time was 263.6 ± 13.03 sigma; whereas the median voluntary reaction time was 252.5 ± 16.33 sigma. Such a result as this could not have been obtained 'by chance.'

Fatigue to the sound stimulus proceeds faster than fatigue to the shock stimulus. The results of this experiment were obtained in spite of the effects of fatigue and negative auditory adaptation.

Any inhibitory effect which may have been present while the subject was reacting voluntarily to the sound (at the end of the experiment), would have also been functioning in a similar manner for the conditioned reaction if voluntary factors had played an important rôle in speeding up the conditioned reaction. Apprehension causes only a very slight increase in the speed of winking, and certainly not enough to account for the results of this experiment.

2. In almost all of the experiments described above there were very marked irregularities in the time of the conditioned (eyelid) reaction. In most subjects the conditioned reaction times had a rather large variability. In a very few subjects it was difficult to establish any conditioned reaction at all. In a few other subjects, however, the conditioned reaction was easily established and it was also fairly regular.

3. The electric shock was kept practically constant for each subject throughout a single experiment. At the beginning of the experiments, all of the subjects judged the shock coming alone to be equal in strength to the shock present in the compound shock-sound stimulus. *After* the training period, *i.e.*, after the compound shock-sound stimulus had been repeated a thousand or so times, the shock coming alone now seemed to the subject decidedly *weaker* than the shock which was present in the shock-sound stimulus. The objective measures of the time of winking to the shock

alone and to the shock-sound stimulus, after training, show that the latter are faster and more regular. The sound has more of a speeding up and regulating effect on the wink (called out by the shock) after the training series than before. This result cannot be accounted for on the assumption that fatigue to the shock proceeded at a much faster rate than fatigue to the sound. There is every evidence that just the reverse is true. These results were obtained in spite of the effects of negative auditory adaptation and fatigue.

4. The results of the experiment indicate some of the conditions which are favorable for securing a conditioned eyelid reaction. The intensity of the sound should be such that it just fails to call out a natural reflex wink at the beginning of the experiment. When the conditioning stimulus (sound) is given *with* the fundamental stimulus (shock) a large number of times, a conditioned reaction will probably be established. Some evidence is found that a conditioned reaction will *not* be established when the sound stimulus is given shortly *after* the shock stimulus. Or, stated more correctly, giving the sound *after* the wink a large number of times does not cause the wink to become conditioned to the sound.

5. The above experiments have the effect of speeding up the subject's voluntary reaction time, so that he is able to voluntarily wink faster at the end than at the beginning of the experiment.

Bibliography

1. Bergström, John A. A Type of Pendulum Chronoscope and Attention Apparatus. *Psychol. Rev.*, 1900, 7, 483–489.
2. Bergström, John A. Pendulum Chronoscopes and Accessories for Psychological Experimentation. *Psychol. Rev.*, 1910, 17, 1–18.
3. Bland-Sutton (Sir J.). The Third Eyelid. *Lancet*, London, 1919, 2, 673–675.
4. Cason, Hulsey. The Conditioned Pupillary Reaction. Journal of Exper. Psychol., 5, 1922, 108–146.
5. Castex, Edmond. Précis d'electricité médicale. Technique. 3d ed. Paris, 1916.
6. Charpy, M. et Clermont. Structure topographique des paupières et épanchements intra-palpébraus. Bibliog. anat., Paris & Nancy, 1910, 21, 65–85.
7. Dodge, Raymond. A systematic exploration of a normal knee jerk. *Zeitschrift für allgemeine Physiologie*, 1911, 12, 1–58.
8. Dodge, Raymond. The Refractory Phase of the Protective Reflex Wink. *Amer. J. of Psychol.*, 1913, 24, 1–7.

9. DONTCHEF-DEZEUZE, MME. MARCELLE. L'image et Les Réflexes Conditionnels dans les travaux de Pawlow. Paris, 1914.
10. ERB, WILHELM. Electrotherapeutics. Trans. by A. de Watteville. N. Y., 1887.
11. EXNER, SIGM. Experimentelle Untersuchung der einfachsten psychischen Processe. *Pfluger. Archiv. für Physiologie*, 8, 1874, 526–537.
12. FRANCK, C. Über die zeitlichen Verhältnisse des reflektorischen Lidschlusses. Inauguraldissertation. Königsberg, 1889. Quoted from Weiss (25).
13. GARTEN, S. Zur Zenntnis des zeitlichen Ablaufes der Lidschläge. *Pflügers Archiv.*, 71, 477–491, Taf. 2, 3, 1898.
14. LADD, G. T., AND WOODWORTH, R. S. Elements of Physiological Psychology, 1911.
15. LAMANSKY, G. Bestimmung der Winkelgeschwindigheit der Blickbewegung, resp. Augenbewegung. *Pflügers Archiv.*, 2, 1869, S. 418–422.
16. MAYHEW, DAVID P. On the Time of Reflex Winking. *J. Exper. M.*, N. Y., 1897, 2, 35–47.
17. MEUNIER, CHARLES. Le réflexe palpébral ou mieux le réflexe cornéo-conjonctival dans l'hemiplegie. Paris Theses, 1909–10.
18. PARTRIDGE, GEORGE E. Experiments upon the Control of the Reflex Wink. *Am. J. Psychol.*, 1899–1900, 11, 244–250.
19. POTTER, S. O. L. Compend of Human Anatomy. Revised by D. G. Metheny, 1915.
20. SHERRINGTON, C. S. The Integrative Action of the Nervous System, 1920.
21. STARLING, E. H. Principles of Human Physiology, 1920.
22. STILES, P. G. The Nervous System and Its Conservation, 1917.
23. TODD, J. W. Reaction to Multiple Stimuli. *Archives of Psychol.*, No. 25, 1912.
24. TOUSEY, SINCLAIR. Electricity, Röntgen Rays, and Radium, 1921.
25. WEISS, OTTO. Die zeitliche Dauer des Lidschlages. *Ztschr. f. Psychol. u. Physiol. d. Sinnesorg.*, 1911, 45, 1, Abt., 307–312.
26. ZWAARDEMAKER, VON H. AND LANS, L. J. Ueber ein Stadium relativer Unerregbarkeit als Ursache des intermittirenden Charakters des Lidschlagreflexes. *Centralblatt für Physiologie*, 13, 1899, 325–329.

AN INSTRUMENT FOR MEASURING CERTAIN ASPECTS OF INTELLIGENCE IN RELATION TO GROWTH, PRACTICE, FATIGUE, AND OTHER INFLUENCES

BY E. L. THORNDIKE,

Institute of Educational Research, Teachers College

It seems likely that psychologists will wish to make repeated measurements of the same individuals in such aspects of intelligence as tests of the type of the Army Alpha measure. Growth, practice and fatigue, are three notable cases. It is convenient and time-saving in such work to have an instrument represented in many alternative forms which are nearly equal in difficulty and whose differences in difficulty are rather exactly known.

In an earlier paper[1] I presented provisional facts for a thirty-minute test[2] (which is essentially a harder Alpha, plus four new tests and one of the old Examination A) prepared in fifteen alternative forms. I am now able to state the relative difficulty of these fifteen forms with much greater precision, precision sufficient probably for any purposes to which such tests are likely to be put.

In the earlier paper it was shown that $A = B + 5$, $J = G + 2\frac{1}{2}$ and $J = C - 1$, with high reliability. I have shown elsewhere that when two forms are given after ten minutes' fore-exercise, the median gain from practice is 8. By using the data just mentioned for J, F, G, and C, and the following data for forms given after ten minutes' fore-exercise, we may equate all forms except K.

[1] *Journal of Applied Psychology*, 1920, vol. 4, pp. 283–288.

[2] Part I of the 1919–1923 Series of the Thorndike Intelligence Examination for High-School Graduates.

Order of Forms	Difference	Amount of Difference: Allowing 8 for Practice	Number of Individuals
B C..............	*C* − *B*	+ 2	86
B G..............	*G* − *B*	+ 4	126
C E..............	*E* − *C*	− 3	254
D C..............	*D* − *C*	+ 2	168
D E..............	*E* − *D*	− 3½	312
D E..............	*E* − *D*	− 4	134
D E..............	*E* − *D*	− 4½	99
E D..............	*E* − *D*	− 4	72
E H..............	*H* − *E*	+ 3½	222
E M..............	*M* − *E*	− 3	86
E M..............	*M* − *E*	+ 1½	34
F M..............	*M* − *F*	0	28
J M..............	*M* − *J*	− 4½	28
F N..............	*N* − *F*	− 4	167
N F..............	*N* − *F*	− ½	70
G H..............	*H* − *G*	− 2	105
G H..............	*H* − *G*	+ 1	28
H O..............	*O* − *H*	− 8	68
I M..............	*M* − *I*	− 4½	320
I M..............	*M* − *I*	− 9	69
I M..............	*M* − *I*	− 6	24
M I..............	*M* − *I*	− 4½	409
M I..............	*M* − *I*	− 3½	419
J L..............	*L* − *J*	− 2½	392
L O..............	*O* − *L*	0	273
L O..............	*O* − *L*	− 1	286
O L..............	*O* − *L*	− ½	54

From the above we have, in comparison with the average, − meaning harder and + meaning easier:

A + 2
B − 3
C + 2
D + 3
E − 1
F 0
G − 1½
H + 1½
I + 2½
J + 1
L − 1½
M − 2
N − 2
O − 2

The computations were made as follows:

$C = J + 1 \qquad F = J - 1 \qquad G = J - 2\tfrac{1}{2}$ are given.

$E - C = -3 \qquad E - J - 1 = -3 \qquad E = J - 2$

$L - J = -2\tfrac{1}{2} \qquad L = J - 2\tfrac{1}{2}$

$O - L = -\tfrac{1}{2} \qquad O - J + 2\tfrac{1}{2} = -\tfrac{1}{2} \qquad O = J - 3$

$N - F = -4$ from 167 cases

$N - F = -\tfrac{1}{2}$ from 70 cases

So $N - F = -3 \qquad N - J + 1 = -3 \qquad N = J - 4$

But since from other scattered data we have $N = J$, we may estimate N as $J - 3$.

$D - C = 2 \quad C = J + 1 \quad D - J - 1 = 2 \qquad D = J + 3$
$n = 168$

$E - D = 4 \quad E = J - 2 \quad J - 2 - D = -4 \quad D = J + 2$
$n = 312 + 134 + 99 + 72$

$M - E = -3 \qquad M - J + 2 = -3 \qquad M = J - 5 \qquad n = 86$
$M - E = 1\tfrac{1}{2} \qquad M - J + 2 = 1\tfrac{1}{2} \qquad M = J - \tfrac{1}{2} \qquad n = 34$
$M - F = 0 \qquad M - J + 1 = 0 \qquad M = J - 1 \qquad n = 28$
$M - J = -4\tfrac{1}{2} \qquad M = J - 4\tfrac{1}{2} \qquad n = 28$

So $M = J - 3$
$M - I = -4\tfrac{1}{2}$ for $n = 729$
$M - I = -3\tfrac{1}{2}$ for $n = 419$
$M - I = -9$ for $n = 69$
$M - I = -6$ for $n = 24$

So $M - I = -4$
$M = J - 3$
$J - 3 - I = -4\tfrac{1}{2}$
$I = J + 1\tfrac{1}{2}$

$H - E = 3\tfrac{1}{2} \qquad H - J + 2 = 3\tfrac{1}{2} \qquad H = J + 1\tfrac{1}{2} \qquad n = 222$
$H - G = -2 \qquad H - J - 2\tfrac{1}{2} = -2 \qquad H = J + \tfrac{1}{2} \qquad n = 105$
$H - G = 1 \qquad H - J - 2\tfrac{1}{2} = 1 \qquad H = J + 3\tfrac{1}{2} \qquad n = 28$
$H - O = -8 \qquad H - J - 2\tfrac{1}{2} = 8 \qquad H = J - 5\tfrac{1}{2} \qquad n = 68$
So $H = J + \tfrac{1}{2}$

$C - B = 2 \qquad J + 1 - B = 2 \qquad B = J - 1 \qquad n = 86$
$G - B = 4 \qquad J - 2\tfrac{1}{2} - B = 4 \qquad B = J - 6\tfrac{1}{2} \qquad n = 126$
So $B = J - 4$

Replacing J by (Average $+ 1$) in the above, we have A, B, C, etc., in terms of the average, as shown above.

Since the average score for college entrants in these tests after fore-exercise is about 100, the numbers $+ 2$, $- 3$, $+ 2$, etc., may be considered as percents.

K was equated as $J - 0.3/4$ in an earlier paper on the assumption that the practice effect from a second to a third full trial was 3. If this practice effect is counted as 4, $K = J - 1\frac{3}{4}$. K is thus very near the average. We have called it 0.

Using these values we determine the practice effect from first to second trial when no fore-exercise is given as follows:

Order	Difference	Raw Difference	Difference Corrected for Difficulty of Forms	Number of Cases
$A\ F$	$F - A$	10	12	160
$A\ G$	$G - A$	9	$12\frac{1}{2}$	113
$A\ J$	$J - A$	12	13	154
$B\ F$	$F - B$	15	12	200
$B\ G$	$G - B$	$9\frac{1}{2}$	8	20
$F\ G$	$G - F$	10	$11\frac{1}{2}$	27
$F\ J$	$J - F$	$13\frac{1}{2}$	$12\frac{1}{2}$ } 12	36 } 144
$G\ J$	$J - G$	15	$17\frac{1}{2}$	33
$J\ L$	$L - J$	11	$13\frac{1}{2}$	28

This practice effect is thus $12\frac{1}{4}$.

Using the same values we determine the practice effect from second to third of three full trials preceded by no fore-exercise as follows:

Order	Difference	Raw Difference	Difference Corrected for Difficulty of Forms	Number of Cases
$G\ H$	$H - G$	9	6	113
$F\ K$	$K - F$	$2\frac{1}{2}$	$2\frac{1}{2}$	351
$J\ K$	$K - J$	3	4	144
$H\ L$	$L - H$	-1	2	36
$J\ M$	$N - J$	$2\frac{1}{2}$	$5\frac{1}{2}$ } $3\frac{1}{2}$	40 } 102
$G\ J$	$J - G$	5	$\frac{1}{2}$	26

The practice effect from a second to a third trial thus seems to be about $3\frac{1}{2}$, but is less reliably determined than the $12\frac{1}{4}$.

Using $12\frac{1}{4}$ and $3\frac{1}{2}$, we may check the determinations of the relative difficulty of the forms from the cases where no fore-exercise was given.

Thus $J - A = 12 - 12\frac{1}{4}$ or $-\frac{1}{4}$ for $n = 154$. We had $J - A = -1$, a difference of $\frac{3}{4}$. $F - A = 10 - 12\frac{1}{4}$ or $-2\frac{1}{4}$ for $n = 160$. We had $F - A = -2$, a difference of $\frac{1}{4}$. $F - B = 15 - 12\frac{1}{4}$ or $2\frac{3}{4}$ for $n = 200$. We had $F - B = 3$, a difference of $\frac{1}{4}$. $G - A = 9 - 12\frac{1}{4}$ or $-3\frac{1}{4}$ for $n = 113$. We had $-3\frac{1}{2}$, a difference of $\frac{1}{4}$. $K - F = 2\frac{1}{2} - 3\frac{1}{2}$ or -1 for $n = 351$. We had $K - F = 0$, a difference of 1. $K - J = 3 - 3\frac{1}{2}$ or $-\frac{1}{2}$ for $n = 144$. We had $K - J = -1$, a difference of $\frac{1}{2}$. The average discrepancy between these differential determinations by the two methods is thus $\frac{1}{2}$.[1] The determinations are thus sufficiently reliable for any use that is likely to be made of the tests, since the median difference between two trials by the same individual is about fifteen times the probable error of the estimate of difficulty.

[1] I have checked further with recently obtained data from tests given after fore-exercise as follows:

The order being Practice	$I\,L$,	the median	$L - I = 8$,	$n = 44$
" " " "	$I\,M$,	" "	$M - I = 6$,	$n = 40$
" " " "	$I\,N$,	" "	$N - I = 5.5$,	$n = 41$
" " " "	$J\,L$,	" "	$L - J = 8$,	$n = 41$
" " " "	$J\,M$,	" "	$M - J = 7.5$,	$n = 41$
" " " "	$J\,N$,	" "	$N - J = 8.5$,	$n = 45$

Equating I with J by their differences from L M and N practice we have $I - J = 1.8$ By our standard values $I - J$ was 1.5.

The order being Practice	$J\,E$,	the median	$E - J$	was 5.5,	$n = 73$
" " " "	$J\,I$,	" "	$I - J$	" 8,	$n = 64$
" " " "	$J\,L$,	" "	$L - J$	" 5,	$n = 63$
" " " "	$J\,O$,	" "	$O - J$	" 3.5,	$n = 56$

Using the standard values for E, I, L and O, and reversing the 8 allowance for practice, we have $J = +2$. By our standard values $J = +1$.

The order being Practice	$E\,H$,	the median	$H - E$	was 8.5,	$n = 32$
" " " "	$E\,M$,	" "	$M - E$	" 9.5,	$n = 34$
" " " "	$E\,O$,	" "	$O - E$	" 6,	$n = 33$

Using the standard values for H, M and O, and reversing the 8 allowance for practice, $E = -0.8$. By our standard it was -1. The discrepancies in these cases are about the same as before.

These differences in difficulty are of course from medians and will fit fairly well the sort of person who scores about 100 in the test. It cannot be in any way guaranteed that, say, Form *B* is 3 per cent. harder than the average form for very low scorers and for very high scorers. I have data from which the differences in difficulty between forms could be computer, for high scorers and low scorers separately, but the matter does not at present seem important enough to justify the great expense of time required.

RELIABILITY OF SCORES IN STEADINESS TESTS

BY JAMES QUINTER HOLSOPPLE

Psychological Laboratory, The Johns Hopkins University

If significant results are to be obtained from steadiness tests a sufficiently large number of cases must be examined with enough accuracy to validate the conclusions. If too fine measurements are used the work on each individual is unnecessarily increased, the total amount of work done is very great, and the number of cases which one experimenter can examine becomes correspondingly small. However, if the method is simplified too far serious errors occur because of different interpretations which may be drawn from the same data. When a steadiness plate is used the touches are usually recorded either by the kymograph or by an observer with a telephone receiver. In both cases the number of holes used is arbitrarily limited.

When a telephone receiver is used in counting the touches the experiment can be performed more quickly and it need not be done in the laboratory. On the other hand, the difficulties inherent in the recording of light electrical contacts by mechanical means appear to invalidate conclusions drawn from kymographic records. For example; if the plate and stylet are connected in a power circuit (110 volts), with resistance small enough to be certain of the passage of the current when a light contact is made, the arc interferes with the steadiness of the subject and in a short time changes the size of the hole. But if the amperage is low enough to eliminate the arc there is danger that the current which passes through in case of a very light touch will be insufficient to operate the marker which is necessarily wound with low resistance.

If we assume that a record of touches made by a subject in all eleven holes of a plate is a fair index of his steadiness, but that such a record can not be practicably obtained, several questions arise.

1. Do the obvious disadvantages of the telephone receiver (the difficulty in identifying double touches, the failure to record the duration of contacts, and the introduction of the 'personal equation' as a source of error) outweigh its advantages?

2. If so, is this true only for the smaller holes where a great many touches are made?

3. If the receiver is satisfactory to certain minimal limit of hole-size do the touches made before this limit is reached give a fair index of the subject's steadiness?

4. Or is the variability of the telephonic record independent of the number of touches per hole?

5. If so, is this variability ever negligible?

6. What method of scoring is least affected by these disturbing factors?

The method adopted for the solution of these problems which seemed to be the simplest, was to obtain and compare the records of several subjects' steadiness as recorded by the kymograph and counted simultaneously by two independent observers. In order that no touches could be recorded by the kymograph which were not heard by the observers, one switch threw the motor, marker, and telephone receivers into the circuit at the same time. The Johns Hopkins steadiness plate was used (1). Ten subjects; four undergraduates, four graduate students, and two instructors in psychology, were tested. Schachne Isaacs, instructor in the department of Psychology, and Mildred E. Day, graduate student, acted as observers. Results of the test on ten reactors are given in Tables I. and II.

Summary of Results

Total number of touches recorded by Kymograph	4846
" " " " " " Observer 1	4656
" " " " " " Observer 2	3872

In Table III. are the rankings of the subjects by the kymograph and two observers based on the total number of touches made by each subject in all of the eleven holes.

Table I

Table of Records

Right Hand

Subject	Hole Number											
	1	2	3	4	5	6	7	8	9	10	11	Recorded by
A..........	1	0	1	4	6	7	15	25	31	47	75	Kym
	1	0	1	4	6	6	14	26	34	41	58	O–1
	1	0	1	4	6	5	13	21	24	26	46	O–2
B..........	0	0	0	0	2	4	9	25	24	43	52	Kym
	0	0	0	0	2	3	9	31	32	49	61	O–1
	0	0	0	0	2	3	9	20	22	38	41	O–2
C..........	0	0	0	0	4	4	13	19	27	42	63	Kym
	0	0	0	0	5	4	13	21	28	44	62	O–1
	0	0	0	0	3	4	13	17	24	32	45	O–2
D..........	8	18	9	12	18	37	56	79	102	110	108	Kym
	7	18	10	10	20	33	61	72	89	85	105	O–1
	6	13	10	8	17	24	30	42	51	50	70	O–2
E..........	0	4	2	1	1	12	15	22	29	50	60	Kym
	0	4	3	1	1	11	18	23	33	38	65	O–1
	0	4	2	1	2	12	16	20	29	44	52	O–2
F..........	0	0	0	1	0	1	6	19	15	29	37	Kym
	0	0	0	1	0	1	5	16	13	28	30	O–1
	0	0	0	1	0	1	6	16	12	21	33	O–2
G..........	2	1	2	8	7	11	24	36	45	45	61	Kym
	2	1	2	6	8	10	22	34	37	47	64	O–1
	2	1	2	6	7	10	22	36	40	38	64	O–2
H..........	1	2	0	1	3	1	7	7	29	45	38	Kym
	0	1	0	0	1	1	5	6	34	41	36	O–1
	0	1	0	0	1	1	6	8	35	33	28	O–2
I..........	3	3	0	15	9	7	18	24	36	42	58	Kym
	2	3	0	11	8	7	18	25	30	31	49	O–1
	2	3	1	10	8	7	17	22	28	36	62	O–2
J..........	6	0	0	1	1	2	18	9	20	29	67	Kym
	3	0	0	1	1	2	17	10	19	27	51	O–1
	3	0	0	1	1	2	14	8	20	27	38	O–2

Legend: Kym indicates Kymographic record
O–1 " Observer 1's "
O–2 " " 2's "

The subjects may be ranked on the basis of the last hole previous to the one in which an arbitrarily selected number of touches were made. Table IV. contains a ranking of this type, taking fifteen touches in twenty seconds as a limit. This is the method used by Woolley and Fischer (2), except that they chose twelve touches in fifteen seconds as a limit rather than fifteen touches in twenty seconds. Whipple (3) also suggests this method. Dewey, Child, and Ruml (4) report the test as given in the same way, but in scoring the separate values for hole and contacts were combined. This makes possible a distribution which can be easily correlated with others but involves the danger of weighting the holes more or less than their difference in size justifies.

Table II

Table of Records

Left Hand

Subject	Hole Number											
	1	2	3	4	5	6	7	8	9	10	11	Recorded by
A..........	1	4	5	1	9	10	30	27	37	43	78	Kym
	1	4	5	1	9	11	30	32	37	43	64	O–1
	1	4	5	1	9	10	24	20	25	35	70	O–2
B..........	0	0	1	2	4	8	28	29	34	57	74	Kym
	0	0	1	2	3	8	30	35	40	62	85	O–1
	0	0	1	2	4	8	25	24	28	44	70	O–2
C..........	3	1	0	2	8	14	14	35	36	43	66	Kym
	1	1	1	1	7	11	13	36	41	40	70	O–1
	1	1	1	1	6	10	10	26	31	37	52	O–2
D..........	3	7	2	4	10	8	25	60	58	58	79	Kym
	2	8	2	5	9	9	38	58	68	78	110	O–1
	2	7	2	4	8	9	17	32	38	44	56	O–2
E..........	16	22	21	41	43	40	62	55	57	66	63	Kym
	12	26	27	39	38	39	50	61	45	48	58	O–1
	12	19	21	29	29	28	45	47	44	58	64	O–2
F..........	0	1	3	1	3	5	17	21	34	40	57	Kym
	0	1	2	1	4	4	15	19	31	35	45	O–1
	0	2	2	1	4	5	15	18	28	37	42	O–2
G..........	6	8	7	9	13	32	42	41	45	55	57	Kym
	5	7	6	8	12	28	33	33	41	51	68	O–1
	6	8	6	8	13	23	40	45	48	52	60	O–2
H..........	3	2	1	12	16	9	22	32	25	36	57	Kym
	2	1	1	8	11	7	16	32	27	38	45	O–1
	2	1	1	7	9	8	17	29	27	20	42	O–2
I..........	6	0	1	4	4	10	10	31	29	46	49	Kym
	6	1	1	2	4	10	9	28	29	36	48	O–1
	6	1	1	3	4	7	9	22	22	46	44	O–2
J..........	2	3	3	8	7	22	24	33	49	52	90	Kym
	2	2	4	4	7	21	23	35	39	42	59	O–1
	2	3	4	5	7	18	22	26	36	36	57	O–2

Legend: Kym indicates Kymographic record
O–1 " Observer 1's "
O–2 " " 2's "

Johnson (5) modified the Cincinnati method so that the cases whose limiting hole was the same were differentiated by their total number of touches rather than by the number made in the last hole. Table IV. contains a ranking on this basis.

Table III

Ranking on Basis of Total Number of Contacts Made in All Eleven Holes of the Plate

	Ranking by Kymograph	Ranking by Observer 1	Ranking by Observer 2
Right Hand.........	*F* 108	*F* 94	*F* 90
	H 134	*H* 125	*H* 113
	J 153	*J* 131	*J* 114
	B 159	*C* 177	*B* 135
	C 172	*I* 184	*C* 138
	E 196	*B* 187	*A* 147
	A 212	*A* 191	*E* 182
	I 215	*E* 220	*I* 196
	G 241	*G* 233	*G* 228
	D 557	*D* 510	*D* 321
Left Hand..........	*F* 182	*F* 157	*F* 154
	I 190	*I* 174	*H* 163
	H 215	*H* 188	*I* 165
	C 222	*C* 222	*C* 176
	B 237	*A* 237	*A* 204
	A 245	*J* 238	*B* 206
	J 293	*B* 266	*J* 216
	D 314	*G* 292	*D* 219
	G 315	*D* 387	*G* 309
	E 486	*E* 443	*E* 396

Legend: Every subject is represented by a letter. The score following each is the total number of his contacts in the eleven holes.

The number of touches in a selected group of holes may be used as a score. With our subjects the sum of the contacts in holes six, seven, and eight gave a very high correlation with the scores based on the total number of touches in all eleven holes, as shown in Table IV. However, with other groups of holes this was not true. Hence this method would hardly be satisfactory for general use since a group of subjects with a different range of steadiness, such as young children, would have to be tested in an entirely different set of holes.

A combination of the above methods might eliminate some of the disadvantages of each. If a maximum total number of contacts, such as fifteen, is allowed and the rating based on the last hole before this limit was reached, one obtains a ranking which is almost identical with that based on the total number of contacts in all eleven holes. An additional advantage is that the records of an observer using a telephone

TABLE IV

	I	II	III	IV
Right Hand . .	*F*	*H* 8–7	*H* 15	*H* 8–22
	H	*F* 7–6	*F* 26	*F* 7–8
	J	*B* 7–9	*J* 29	*B* 7–15
	B	*C* 7–13	*C* 36	*C* 7–21
	C	*J* 6–2	*B* 38	*E* 7–33
	E	*A* 6–7	*A* 47	*A* 7–34
	A	*G* 6–11	*E* 49	*J* 6–10
	I	*E* 6–12	*I* 49	*G* 6–31
	G	*I* 3–0	*G* 71	*I* 6–37
	D	*D* 1–8	*D* 172	*D* 4–8
Left Hand . . .	*F*	*I* 7–10	*F* 43	*I* 7–35
	I	*C* 7–14	*I* 51	*C* 7–42
	H	*F* 6–5	*H* 63	*F* 6–13
	C	*B* 6–8	*C* 63	*B* 6–15
	B	*D* 6–8	*B* 65	*A* 6–30
	A	*A* 6–10	*A* 67	*D* 6–34
	J	*J* 5–7	*J* 79	*J* 5–23
	D	*G* 5–13	*D* 93	*G* 5–43
	G	*H* 4–12	*G* 115	*H* 4–18
	E	*E* 0 (1–16)	*E* 157	*E* 0 (1–16)

Legend:

Column I—Ranking based on total number of touches in all eleven holes as recorded by kymograph.

Column II—Rank and score by the method of Wooley and Fischer except that 15 touches in 20 seconds were selected as a limit.

Example: *H* 8–7
H subject
8 hole reached
7 number of contacts in hole 8.

Column III—Ranking and score based on the total number of touches made in holes six, seven, and eight.

Column IV—Ranking by Johnson's method.

Example: *H* 8–22
H subject
8 hole reached (the last in which not more than 15 touches were made)
22 the total number of contacts up to and including hole 8.

receiver will be practically the same as the kymographic record because of the fewness of contacts which must be counted in any one hole.

Example of this method:

Subject *A* made one contact in the first hole, four in the second, five in the third, one in the fourth, and nine in the fifth. When he has finished with the fourth hole his total is eleven; in the fifth, twenty. Here he has passed his limit of fifteen so his score is 4–11 (4 representing the hole and 11 the number of his contacts).

TABLE V

RANKINGS WHEN A TOTAL OF FIFTEEN CONTACTS IS SELECTED AS A LIMIT COMPARED WITH THE RANKING BASED ON THE TOTAL NUMBER OF CONTACTS IN ALL ELEVEN HOLES

	Kymograph Ranking Based on Total Contacts in All Holes	Rankings on Basis of Fifteen Total Contacts by		
		Kymograph	Observer 1	Observer 2
Right Hand . .	*F*	*F* 7–8	*H* 8–14	*F* 7–8
	H	*B* 6–6	*F* 7–7	*H* 7–9
	J	*H* 6–8	*B* 7–14	*B* 7–14
	B	*C* 6–8	*J* 6–7	*J* 6–7
	C	*J* 6–10	*C* 6–9	*C* 6–7
	E	*E* 5–8	*E* 5–9	*E* 5–9
	A	*A* 5–12	*A* 5–12	*A* 5–12
	I	*G* 4–13	*G* 4–11	*G* 4–11
	G	*I* 3–6	*I* 3–5	*I* 3–6
	D	*D* 1–8	*D* 1–7	*D* 1–6
Left Hand. . .	*F*	*F* 6–13	*F* 6–12	*F* 6–14
	I	*B* 5–7	*B* 5–7	*B* 5–7
	H	*C* 5–14	*C* 5–11	*C* 5–10
	C	*I* 4–11	*I* 5–14	*I* 4–11
	B	*A* 4–11	*A* 4–11	*A* 4–11
	A	*H* 3–6	*H* 4–12	*H* 4–11
	J	*J* 3–8	*J* 4–12	*J* 3–9
	D	*D* 3–12	*D* 3–12	*D* 3–11
	G	*G* 2–14	*G* 2–12	*G* 2–14
	E	*E* 0 (1–16)	*E* 1–12	*E* 1–12

Example: *F* 7–8
F subject
7 hole
8 contacts made up to and including hole 7.

Table V. contains rankings by the kymograph and observers on this basis compared with rankings based on the total contacts made in eleven holes.

Figures 1 and 2 show typical variations of the observers from the kymographic records. Figure 3 contains extracts from several kymographic records indicating that the variations of the observers may depend on the frequency per second

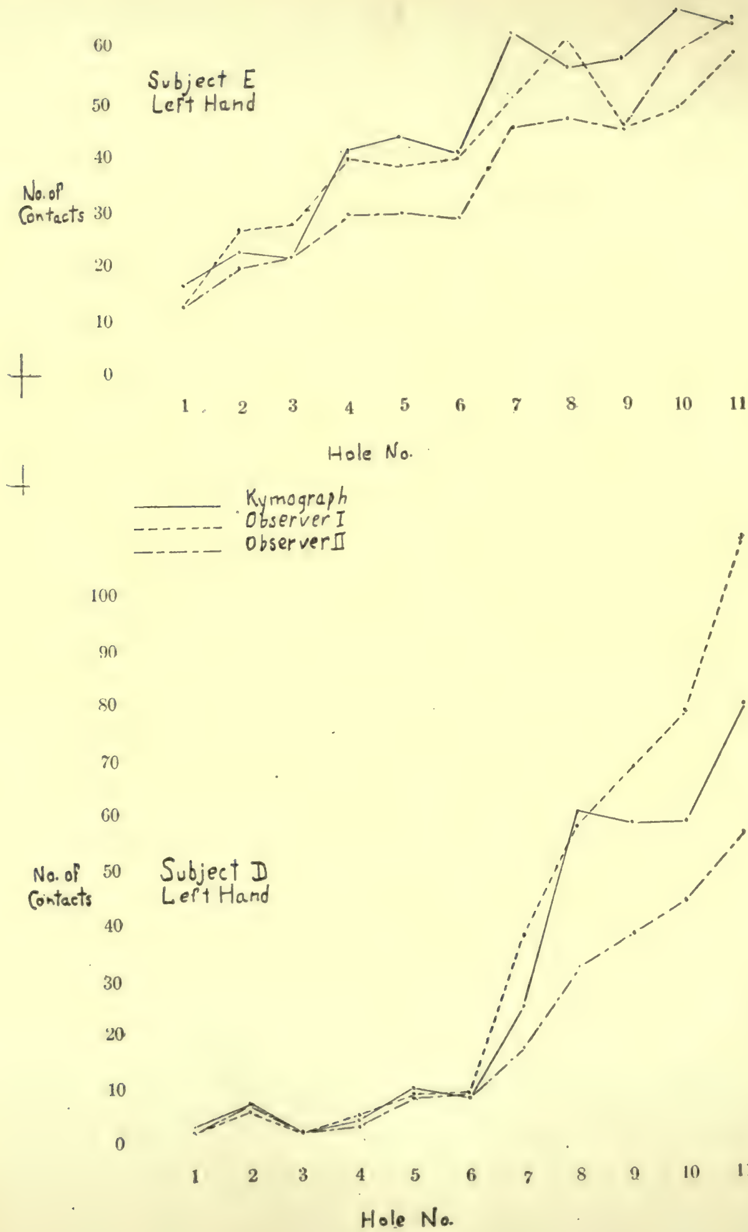

Subject E
Left Hand
No. of Contacts
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11
Hole No.
Kymograph
Observer I
Observer II
Subject D
Left Hand
No. of Contacts
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11
Hole No.

FIG. 1.

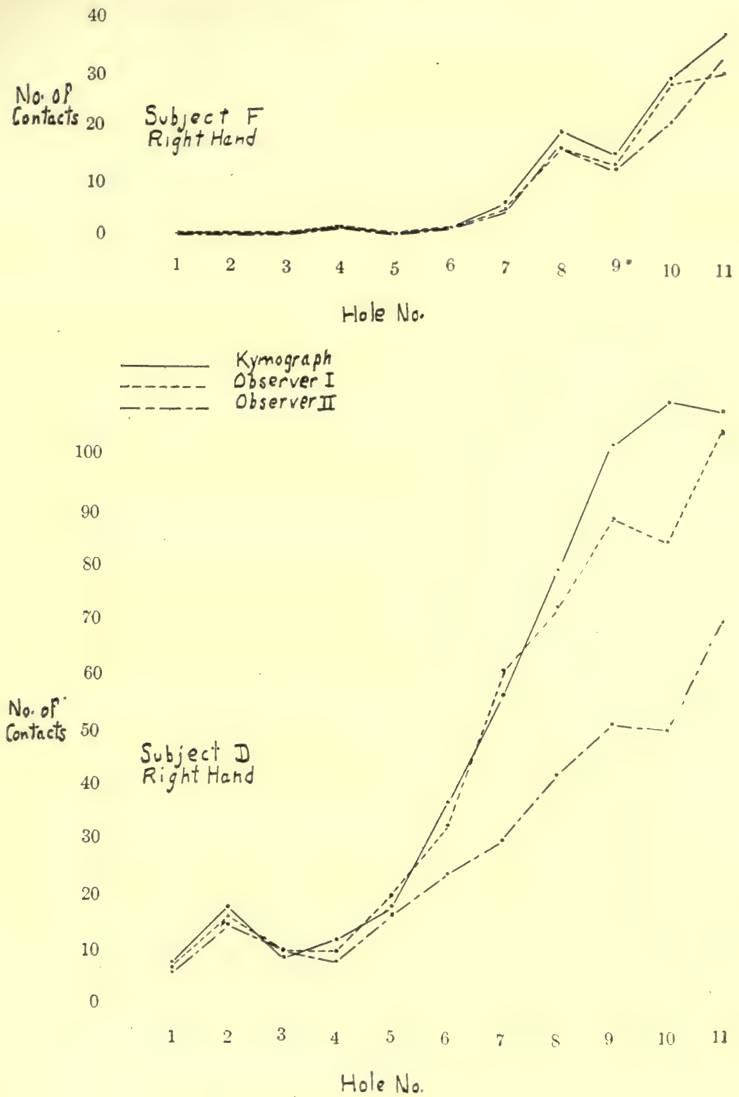
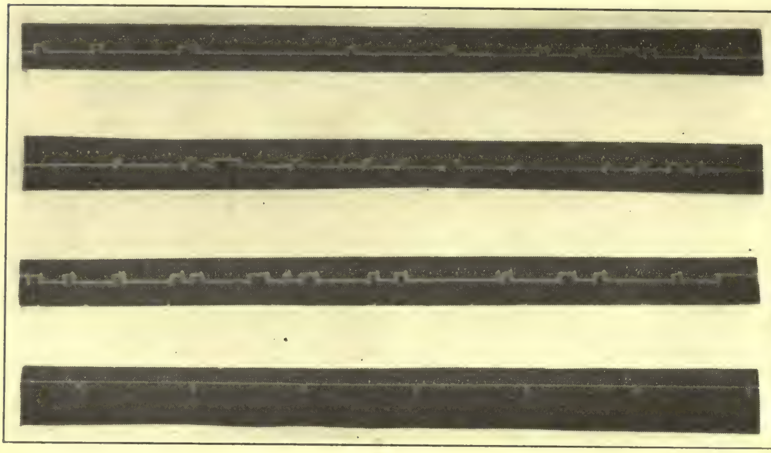


Fig. 2.

of the contacts rather than on the number made per hole. From this data several inferences might be drawn.

1. The variations of the observers from the kymograph are not uniform. This is obvious from the curves shown in Figs. 1 and 2 which were not selected as examples of especially wide variability.

2. There is a direct relation between the number of touches made in any hole and the reliability of the observers at that point.



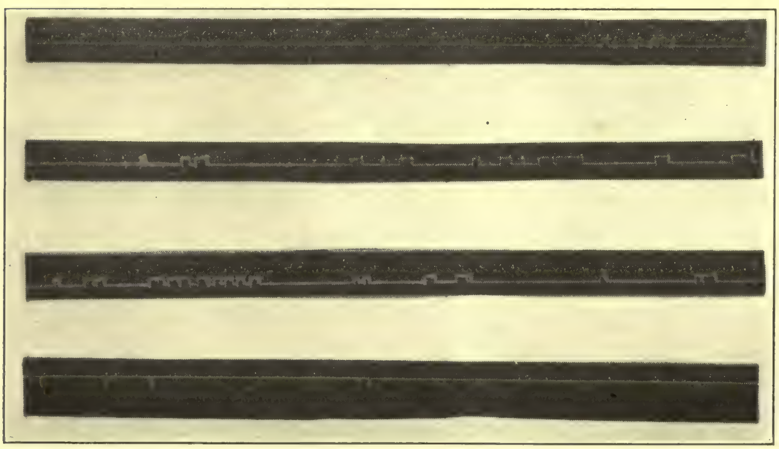
Time Line
In Seconds

E-4-left

E-5-left

J-1-right

Examples of cases in which results of observers tallied closely with kymograph.



Time Line
In Seconds

F-11-right

A-8-right

G-8-right

The total number of touches is practically identical in the two sets of records.

Capital letter indicates the subject, right or left, the hand used, and the number is that of the hole in which the contacts were made.

A comparison of these records with the table of data will show the actual variation in each case.

FIG. 3. Typical kymographic records showing frequency of contacts in cases where results of observers varied widely

3. But there is no single number of touches that can be arbitrarily given as the limit of the observer because different observers vary in accuracy and the same observer varies with different subjects, as shown in Figs. 1 and 2.

4. Since different methods of scoring based on the same data give different rankings some constant method should be used.

5. Although the data is insufficient for drawing any conclusions relative to a satisfactory method the indications are that:

No single hole or arbitrarily selected group of holes offers a fair index of steadiness.

The method of scoring shown last in the summary of data probably simplifies the test without interfering seriously with its accuracy.

If this method is used the advantages of the telephone receiver do appear to outweigh its disadvantages.

References

1. Dunlap, Knight. Improved Forms of Steadiness Tester and Tapping Plate. Journal of Experimental Psychology, 1921, Vol. IV., pp. 430–437.
2. Woolley, H. B., and Fischer, C. R. Mental and Physical Measurements of Working Children. Psychological Monographs, 1914, Vol. XVIII., No. 1.
3. Whipple, G. M. Manual of Mental and Physical Tests, Part I., pp. 155–160. Baltimore, 1914.
4. Dewey, E., Child, E., and Ruml, B. Methods and Results of Testing Schoo Children, Part I., pp. 60–63. New York, 1920.
5. Hunt, J. L., Johnson, B. J., and Lincoln, E. M. Health Education and the Nutrition Class, Chap. VII., p. 165. New York, 1921.

THE RELIABILITY OF THE SEASHORE PHONOGRAPH RECORD FOR THE MEASUREMENT OF PITCH DISCRIMINATION

BY A. R. GILLILAND AND C. R. JENSEN

Lafayette College *Dartmouth College*

This experiment is in no way concerned with the value of the measurement of pitch discrimination as a factor in the investigation of musical talent. It was conducted with a sympathetic appreciation of such measurements and is merely a criticism of the phonograph as a means of obtaining the records.

While giving the Seashore tests for pitch, intensity, time, consonance, and for tonal memory as arranged for the Columbia graphophone records certain distracting factors seemed to vitiate the results. One of these distracting factors was the scratch of the needle on the phonograph record, another was the always more or less noticeable metallic overtones common to all tones reproduced on the phonograph. It seems reasonable to think that these two factors might seriously affect the student's judgment of pitch. In order to overcome these difficulties it was necessary to construct a suitable apparatus for the production of relatively pure tones of the same pitch as those used by Professor Seashore. This made it possible to directly compare the records obtained by each method.

The Apparatus[1]

The method used for producing relatively pure tones is not so essentially different from that used by Seashore in his experiment described in Psychological Monograph No. 53. The accompanying photograph gives an idea of the apparatus used in this experiment. It consists of eight tuning forks, a standard of 435 v.s. and seven others varying by increments

[1] Much credit is due Mr. E. C. Robes, of the Dartmouth workshop, for valuable suggestions and help in the construction of the apparatus.

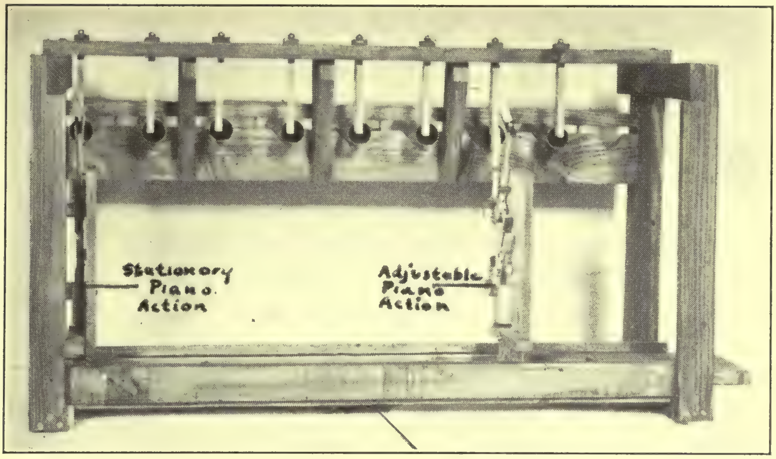


FIG. I.

of ½, 1, 2, 3, 5, 8, and 12 v.d. respectively. These forks are mounted with resonators made of closed brass cylinders. The forks are struck by piano hammers actuated by ordinary upright piano actions. One piano action was attached to a stationary board for striking the standard fork. Another action was attached to a sliding board in such a way that the hammer could be made to strike any one of the seven other forks. The piano actions and the resonators were very carefully adjusted so that with ordinary precaution in striking the keys the tones produced would vary imperceptibly in intensity. Professor Seashore objects to mounting the forks, saying that they then tend to assume individual peculiarities. The authors encountered the same difficulties at first, but these difficulties were surmounted by the methods described above, together with a very exact sidewise adjustment of the movable hammer. By these means noticeable differences in intensity and overtones were almost entirely eliminated. Seashore's further argument that slight intensity differences are desirable in order to minimize any deliberate or subconscious attempt to discriminate the tones on the basis of intensity differences does not seem especially convincing.

The Method

The subjects were 43 unselected members of a class in elementary psychology. Most of them had taken the Seashore tests some months before and so were familiar with the method. The subjects were given no indication of the purpose of the experiment and they did not know that a comparison of the records was to be made, or even that the same tones were to be used with the apparatus as with the phonograph. The observers were seated in the center of a recitation room 30 feet x 40 feet, their backs to the apparatus, so that they could see none of the procedure. The phonograph test was given first. Since only seven comparisons were to be made with the forks—the seven corresponding to the last seven sets of comparisons on the graphophone—the time during which the first three sets of tones were being run off was used to explain and illustrate how the comparisons were to be made. The directions for recording were given as found in the 'Manual of Instructions for Measures of Musical Talent.' At the beginning of the fourth series on the graphophone, the increment of 12 as given in the key, the recording was started and continued for the rest of the record. Immediately following this the tones were given with the fork apparatus again starting with the increment of 12 and following the key exactly for the tones given by the phonograph record. As has been said the observers did not know that the same order was being followed in the two experiments, furthermore the first records were not in sight while the second experiment was being conducted. The whole time taken for the two experiments was about 20 minutes.

It will be remembered that the three practice series were given first and that the experiment had already been taken once before. Any improvement in results, therefore, could not be attributed to practice. On the other hand because of the short time taken no objection of practical significance could be raised that the results of the second part would be affected by fatigue.

Results

The results are given in the accompanying table. The judgments of each observer as made from the phonograph and as made from the forks are placed side by side. These judgments are calculated according to the per cent. of correct responses. Since only seven series instead of ten were taken with each method these results cannot be directly compared with the table given by Professor Seashore for the relative standing on the basis of one hundred unselected individuals. They are valuable only for the comparisons made in this case. It will be noted that the average phonograph judgments are 75.1 per cent right and that of the forks are 83.0 per cent right, a difference of 7.9 per cent in favor of the forks.

The graph offers another interpretation of the results and probably the more significant one. Not only does it show the difference in the total number of errors but also their distribution. The errors made with the forks are concentrated at the side representing the smallest vibration differences. The total number of errors made with the graphophone was 757 as compared with 510 for the forks; yet the number of errors for the ½ increment was 236 on the forks as compared with 207 for the graphophone. The dotted line on the graph representing the fork errors gradually drops off showing the least errors at 12 as should be expected, whereas the solid line representing the phonograph errors comes to its lowest point at 8 and rises slightly at 12. It will, therefore, be noted from the two curves that the one representing the results for the forks is the much more ideal curve.

The Apparatus Compared with Unmounted Forks

Professor Seashore states that he has discontinued the use of all striking devices and uses only an ordinary hammer with or without a resonator. In order to compare the effects of a relatively constant intensity and quality of tones as produced by the apparatus already described in this article, with relatively pure tones but of varying intensity as obtained by striking forks in the usual way, a further test was made. 30 subjects were used in this part of the experiment.

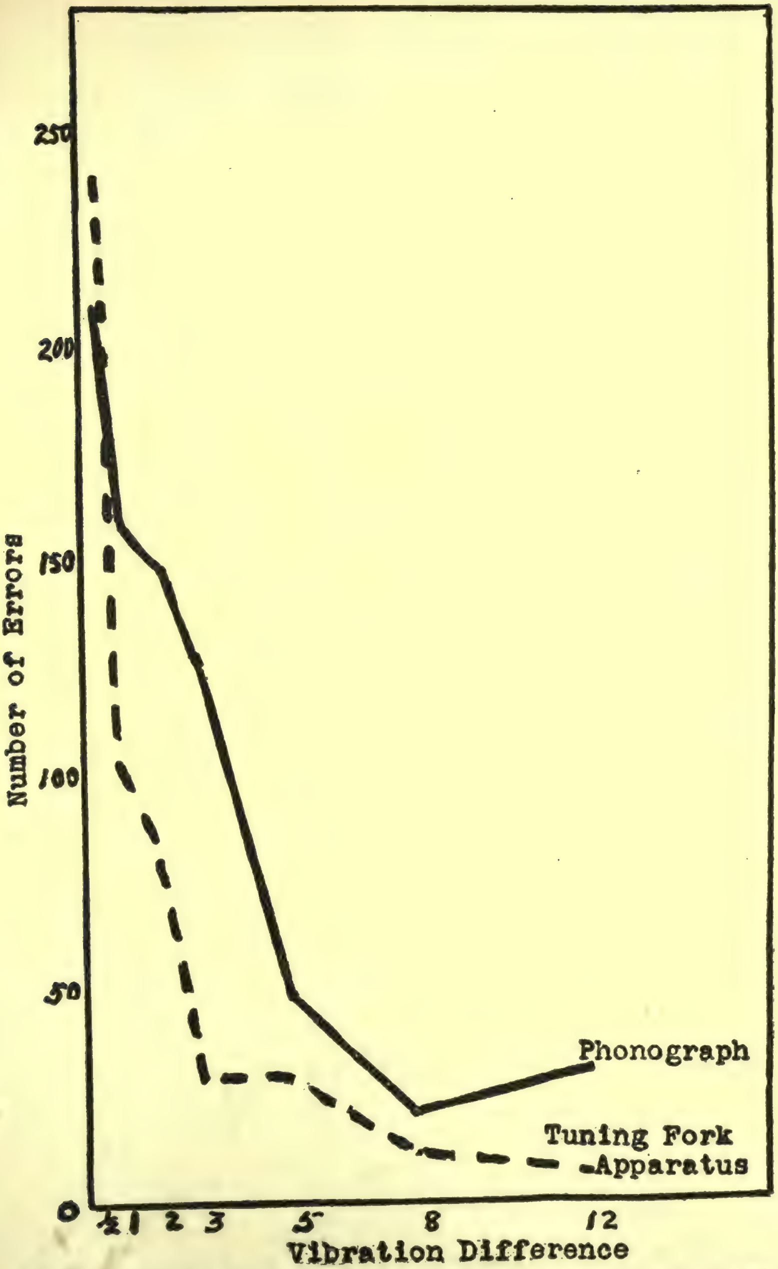
Table I

Showing Per Cent of Correct Responses out of 70 Judgments Each by 43 College Students (1) when the Seashore Record for Pitch was Used as the Source of the Tones, and (2) when Mounted Tuning Forks were Used

No.	Phono-graph	Forks	Advantage Forks	No.	Phono-graph	Forks	Advantage Forks
1	80	84	4	23	61	83	22
2	70	81	11	24	84	89	5
3	73	86	13	25	86	87	1
4	86	83	− 3	26	69	83	14
5	77	77	0	27	77	81	4
6	76	77	1	28	86	86	0
7	84	99	15	29	71	97	26
8	61	80	19	30	83	86	3
9	69	79	10	31	74	91	17
10	89	83	− 6	32	81	96	15
11	66	90	24	33	64	54	−10
12	86	64	−22	34	74	94	20
13	70	90	20	35	79	80	1
14	70	89	19	36	83	86	3
15	83	73	−10	37	70	93	23
16	77	86	9	38	81	81	0
17	74	80	6	39	77	87	10
18	44	80	36	40	81	91	10
19	60	50	−10	41	70	70	0
20	79	87	8	42	74	91	17
21	77	80	3	43	69	86	17
22	84	70	−14				
		Average			75.1	83.0	7.9

Method

From a second set of forks the eight forks corresponding to those set up in the apparatus were selected, that is, a standard of 435 v.s. and seven others differing from the standard by increments of ½, 1, 2, 3, 5, 8, and 12 v.d. In general the same method was used as in the first experiment. The subjects were asked to make the same kind of judgments, that is, whether the last of each pair of tones sounded was higher or lower than the first. The same order of sounding the tones was used as before. Each of the seven forks was, therefore, compared ten times with the standard, making seventy judgments in all with the apparatus and the same number with the unmounted forks. In this experiment the unmounted forks were struck upon the knee of the experimenter and then immediately placed in the socket of the



GRAPH I. Showing the number of errors and their relation to the vibration difference of the tones. 43 subjects.

small common resonance box as furnished with the forks by the C. H. Stoelting Co. Following this the same differences were sounded with the apparatus and the subjects were asked to make seventy more judgments. The subjects did not know that the same differences were being used and could not see their previous records.

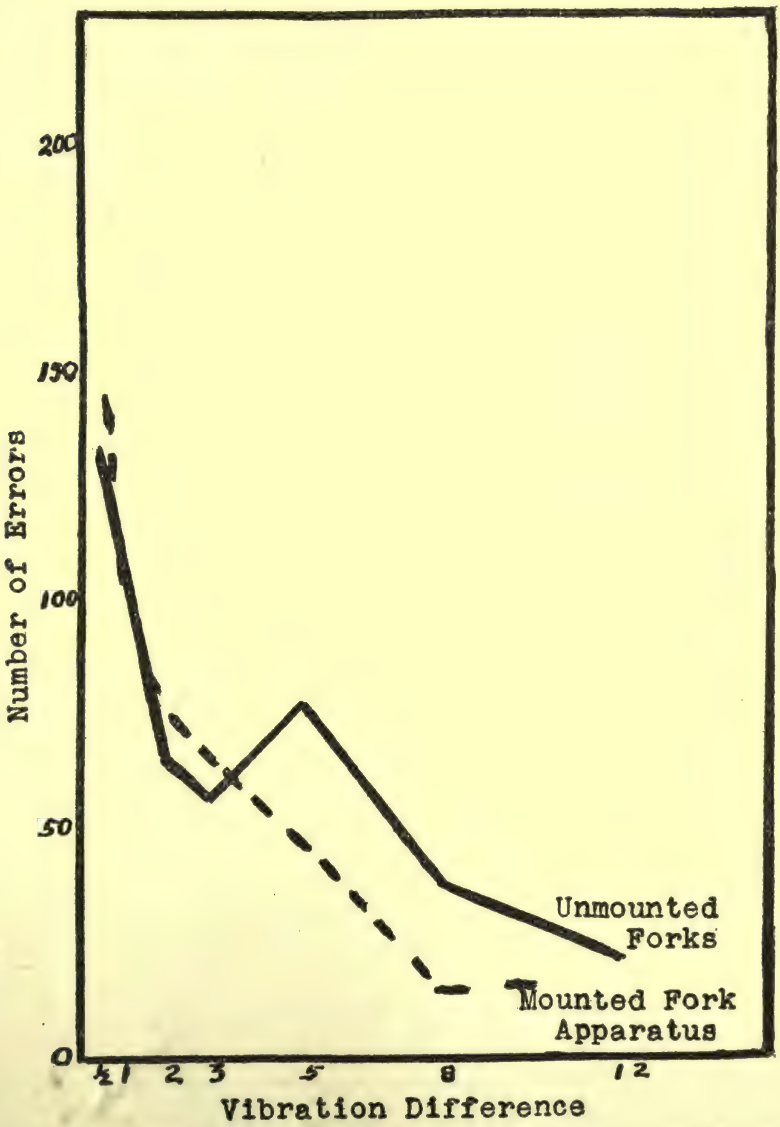
Table II

Showing Per Cent of Correct Judgments for 30 Subjects in Pitch Discrimination Using Mounted Fork Apparatus and Unmounted Forks

No. of Cases	Mounted Forks	Unmounted Forks	Advantage Mounted
1	60	66	− 6
2	77	81½	− 4½
3	76	54	21
4	57	66	− 9
5	81½	68½	13
6	79	70	9
7	79	70	9
8	83	70	13
9	70	73	− 3
10	88½	87	1½
11	64	58½	5½
12	81½	81½	0
13	80	91½	−11½
14	86½	83	3½
15	88½	80	8½
16	68½	73	− 5½
17	78½	86	− 8½
18	73	61½	11½
19	47	51½	− 4½
20	84	78½	5½
21	67	70	− 3
22	86	80	6
23	94	90	4
24	93	90	3
25	81½	74	7½
26	93	88½	4½
27	67	75½	− 8½
28	80	81½	− 1½
29	87	90	− 3
30	93	78	15
	78.1	75.6	2.5

Results

Table II. and Graph 2 give the scores for each method. There were 511 errors when the unmounted forks were used and 471 when the apparatus was used, an advantage of 8.3 per cent in favor of the apparatus. But this difference in favor of the apparatus is not as significant as the regularity of the records with the apparatus as compared with the records for the unmounted forks. The difference in loudness must have been mistaken for pitch difference many times with the unmounted forks. This occurred many times when the difference in pitch was relatively great and must account for



GRAPH II. Showing the number of errors and their relation to the vibration difference of the two tones. 30 subjects.

the greater number of errors in the 5, 8, and 12 v.d. column. This probably sometimes occurred with the apparatus. But practice in striking piano keys soon develops a very regular uniform stroke and hence not so many errors in the columns for the greater vibration differences with the apparatus.

These results prove very conclusively that for exact work, tones for pitch discrimination should be relatively pure tones, with as little quality and intensity difference as possible, that is, the one factor of pitch difference should be isolated for the comparison. The Seashore records have been standardized and are no doubt very satisfactory for the purpose for which they are used in school-room tests. They serve a real purpose in selecting those subjects who can, from those who cannot, discriminate between relatively large pitch differences. But for laboratory experiments on pitch differences the phonograph records are not satisfactory. The unmounted forks set loosely in a small sounding box and struck at different intensities are relatively satisfactory. A much more satisfactory apparatus no doubt could be constructed similar to the one here described with an electro-magnetic device for actuating the hammers. But for practical laboratory tests the results indicate that our apparatus is very satisfactory.

References

Seashore, C. E. The Measurement of Pitch Discrimination, a Preliminary Report. Psychological Monograph No. 53, 1910, Vol. XIII., pp. 21–60.

Seashore, C. E. The Psychology of Musical Talent. Silver Burdett and Co., 1919.

Seashore, C. E. Manual of Instructions, Measures of Musical Talent. Columbia Graphophone Co., New York City.

DISCUSSION

THE STIMULUS ERROR

BY A. P. WEISS
Ohio State University

"This experiment seems to show the futility of properly interpreting the statistical results of a purely behavioristic study without the control of the introspective report" (p. 76), concludes Fernberger[1] in an article that analyzes the types of reactions in a weight discrimination experiment. He bases his conclusions on the fact that: "In the case of lifted weight stimuli, the following instructions are capable of fulfilment; namely, (*A*) judging the intensity of the pressure sensations on the tips of the fingers; (*B*) the intensity of the kinesthetic sensations localized in the wrist; and (*C*) judging, by assuming a stimulus attitude, the relative intensities of the weights themselves. This latter is distinctly a complex perceptual process on the 'meaning' level" (p. 76).

Fernberger implies that the results secured when instructions (*C*) are given are 'purely behavioristic,' and that the evidence for the futility of the behavioristic method is found in the fact that the statistical results from instructions (*A*) and (*B*) differ from each other and from (*C*), and that this difference can only be 'interpreted' through the introspective report. By interpretation Fernberger seems to imply that he has evaluated, from the introspective reports, the different potencies of such factors as, practice, kinesthesis, pressure, visual imagery, widespread tactual and kinesthetic experiences, various degrees of subjective assurance, etc. If only those factors mentioned are considered, his experiment includes the following variables, all of which have an effect on the nature of the discriminating reactions that are released: (1) Pressure stimuli; (2) kinesthetic stimuli; (3) an auditory stimulus "to judge pressure sensations only"; (4) an auditory stimulus to "judge kinesthetic sensations only"; (5) an auditory stimulus to "judge the weights themselves"; (6) practice effects; (7) visual imagery; (8) widespread tactual experiences; (9) widespread kinesthetic

[1] Fernberger, S. W., 'An Experimental Study of the "Stimulus Error,"' J. Exper. Psychol., 1921, *4*, 63–76.

experiences; (10) various degrees of subjective assurance. When a behaviorist tries to investigate such a phenomenon made up of ten variable factors, he tries to keep nine of them constant while varying one factor and measuring the effects on the phenomenon in question. Such an ideal technique is, of course, not possible in a weight discrimination experiment as conducted by Fernberger. The claim that the introspective method makes it possible to approach this ideal, needs to be substantiated. The behaviorist would say that the introspective method merely adds further variables and instead of approaching an ideal technique we get further away from it.

In a behavioristic experiment on weight discrimination the first problem is that of determining what is to be included under the terms "weight discrimination." There are hundreds of different conditions under which weights may release the verbal reactions of *lighter*, *equal*, *heavier*. Which combinations of conditions are to be investigated? Unless special theoretical considerations are involved, the behaviorist would probably select those conditions under which weights are *usually* discriminated by scientists, post-office clerks, store-keepers, housewives, etc. This would give a scientists' limen, a post-office clerks' limen, a store-keepers' limen, a housewives' limen, etc. Which is the *true* or psychological limen? This question can only mean: in which of these limens is the investigator most interested? One is just as true or just as psychological as the other. No behaviorist would deny that a number of different limens may be secured even from the same individual at different times. Further, these limens may be given different names such as stimulus attitude, pressure attitude, kinesthetic attitude, etc., but this only means that the ten factors in weight discrimination mentioned in a preceding paragraph do not all have the same potencies even in the same individual at different times. Instead of dividing the weight-discriminating limens into scientists', post-office clerks', store-keepers', and housewives' categories, the behaviorist might have set up the special condition that the subject must be a graduate student in a psychological laboratory trained in introspection, and trained to discriminate *attitudes* in addition to discriminating weights. This would of course give a still different set of limens. To meet certain theoretical demands one might also reserve the term *psychological limen* for that limen secured under the conditions known as the 'process attitude.' To imply however that this is the limen which the behaviorist *must* regard as the *true* limen of weight discrimination, will hardly be taken seriously by any behaviorist. To maintain that this limen can only be interpreted from a consideration of the introspective reports is, of course, perfectly true, since the introspections are part of the total reaction that is being investigated. If in addition to requiring introspections the experimenter had required the subject to sing an anthem while the experiment was in progress, the way in which these anthems were sung would also have to be considered in the interpretations.

From the behaviorist's standpoint Fernberger's experiment in which *attitudes* are introduced, is not an experiment in weight discrimination at all. It is an experiment to determine whether or not a subject can or cannot "abstract sensation," whatever this may mean. The weight discrimination reactions are merely used as a device for measuring the extent to which *abstraction* is possible. Many other types of reactions (visual, auditory, olfactory, etc.) would have served equally well. The behaviorist is fully aware of the fact that to get some specific limen, it is necessary to maintain the stimulus conditions and the physiological conditions as constant as possible. He does not believe that introspection reveals the extent to which these conditions have remained constant.

A behavioristic investigation is usually concerned with such things as the classification of behavior, the effect of inheritance or training on behavior, the effect of the physical or social environment on behavior; behavior being regarded as those movements of the organism that establish its social (anthropological) status. If introspections are going to prove valuable in investigations directed along these lines, the behaviorist will use them. The present writer regards introspections as verbal reactions that have been developed under conditions quite foreign to the conditions under which normal human reactions occur. They are abnormal supplementary speech reactions of the same nature as illusions, dreams, etc. That the conditions under which they occur may be standardized and the reactions themselves made relatively stable, he would not deny.

The direct response to Professor Fernberger's charge of the 'futility of properly interpreting the statistical results of a purely behavioristic study,' would be:

1. None of Fernberger's experiments are 'purely' behavioristic in the sense that he is trying to establish the laws according to which weight discriminations are made by some social or anthro-

pological unit. They are behavioristic in the sense that he has investigated the manual and speech reactions of three individuals reacting under conditions involving at least ten variables.

2. By introducing the modification of establishing an *attitude* he has made the attempt of measuring two simultaneous variables (pressure and kinesthesis) in terms of a third variable (introspection) much less uniform than either the pressure or the kinesthesis.

3. The introduction of the *attitude* conditions changes the experiment from one of normal weight discrimination to what he would call an experiment on the 'abstraction of sensations.'

4. The behaviorist would like to know just what is to be included under 'to properly interpret?' If to interpret means that the introspections reveal the presence of mysterious *mental* processes that are not physiological, the behaviorist frankly admits his futility. If to interpret is to mean that the introspective reactions are to be regarded as supplementary speech reactions which occur in all adults and as such form part of the total reaction to the weights, the behaviorist will admit this also. He would affirm, however, that the introspective reaction as an end result is hopelessly inadequate to reveal the inheritance factors, the environmental factors, and the training factors, which are the causes of a given set of reactions.

5. Finally, the habit of capping off an experiment in introspection with the dogmatic statement that "therefore the futility of the behavioristic method is clearly shown" is somewhat unfair. The behavioristic method is futile only when it fails to solve *its own problems*. It is not futile because it does not seem to solve the special problems of the introspectionist.

A. P. Weiss

Ohio State University

Journal of
Experimental Psychology

Vol. V., No. 4 August, 1922

INDIVIDUAL VARIATIONS IN RETINAL SENSITIVITY, AND THEIR CORRELATION WITH OPHTHALMOLOGIC FINDINGS[1]

BY CAPTAIN PERCY W. COBB, M.C.

Medical Research Laboratory, Air Service, U. S. Army, Mitchel Field, Garden City, L. I., N. Y.

Acknowledgment

The author wishes to acknowledge the help of Dr. Mildred W. Loring and Miss Elizabeth K. Stark in the performance of most of the technical work of accumulating these results; the help of Miss Marion L. Smith in performing the computations; and the courtesy of Nela Research Laboratories in facilitating the completion of the manuscript.

The method by which the measurements of retinal sensitivity were made, which are the subject of this communication, has, in its essential details, already been described.[2] The purpose of the present paper is to investigate primarily the significance of the results obtained by the method, and secondarily its applicability as a test in the routine examination of candidates for flying status in the Air Service. Measurements were made upon a group of 101 subjects. Visual acuity was measured in nearly all of these, for each eye separately and for both eyes together by means of the Snellen letters under the actual conditions of the experiment; and in addition, comparison was made with the complete findings of the routine ophthalmological examination for flying status upon 59 of these subjects.

The group consisted of such individuals as were available,

[1] Published with permission of the Surgeon General, U. S. Army, who is not responsible for any opinion expressed or conclusions reached herein.

[2] Cobb and Loring, Jour. Exp. Psych., Vol. IV., 1921, pp. 175 to 197.

and without specific enumeration, may be described by noting the classes into which they might be divided, in the order of their relative numbers: (1) Enlisted men of the army, a few from the navy, some recruits and some civil candidates for the cadet-school of aviation; (2) a number of flying officers; (3) a number of other officers, chiefly of the medical corps, of the staff of the Medical Research Laboratory or there present for instruction; (4) a few civil employes of the Laboratory (including four women), technical, clinical and one skilled mechanic.

The present method of estimating retinal sensitivity is, essentially, to determine the duration of exposure of a black dot at the center of a white screen, necessary in order that the dot may just be seen. The actual dimensions of the dot, and the distance of the subject from the screen are kept constant, as is also the photometric brightness of the screen. The dot is formed by a small circular aperture in the screen, through which a secondary screen, behind the other, is seen; and the screen behind is so illuminated that its appearance is an exact match with the margins of the aperture. When this is accomplished the aperture disappears. It becomes in appearance a black dot when an opaque object cuts off the light coming through it, and the necessary time of darkening, in order that the dot may be perceived with 50 per cent. certainty, is the quantity measured.

The usefulness of a new test may depend upon one of two circumstances. It may be supplementary to tests already in use (*a*) in a confirmatory sense, or (*b*) it may measure something not already taken into account by tests currently used. It may be useful in either case. Intrinsically, a test must show, by the results that it gives,—(*c*) that it actually does give a measure of something worth knowing, and (*d*) it must give a fairly reliable result. This last (*d*) has been accomplished, if measurements on the same individual, under conditions as nearly the same as practical considerations warrant, will, at different times, give not too widely divergent results. The condition (*c*) is met (granting that we aim to test individuals) if results from different individuals prove to be different by variations of a larger order of magnitude than the variations in the same individual under identical conditions. If we aim to detect changes in the sensitivity of the individual, the condition (*c*) will have to be met somewhat differently.

The first point to investigate is the reproducibility of the result. It was shown in the previous communication upon this method, that for the three subjects, each having had some pactice as such, the variability of the result of a single series, representing about fifteen minutes' expenditure of time, was represented by a probable error of one unit or less on the arbitrary scale, equivalent to not more than about 8 or 9 per cent. in actual time.[1] The figures are given here in Table I.

This would be considered good reproducibility for such work as the present. But it must be remembered that these subjects were all more or less practiced in the method before the above results were obtained.

Before considering the results of the group, it may be well to examine the measurements taken upon the above three subjects in another respect. The question arises in connection with a test of this sort, not only what will happen when a measurement is attempted upon a new and unpracticed subject, but also what happens in the case of any subject, practiced or not, when several series are run in

[1] The relation between T (stimulus number) and t, the time of exposure (thousandths of a second) is given by the formula: $t = 200(\frac{1}{2})^{T/8}$. An extract from the original table embodying this relation is given here:

T	Time	T	Time	T	Time	T	Time
20	35.4	22	29.7	24	25.0	26	21.0
21	32.4	23	27.3	25	22.9	27	19.3

From these can be obtained the time values corresponding to the other numbers, if we remember that an increase of 8 units in T means that the time is divided by 2 and *vice versa*.

Furthermore, owing to the fact that the screen opening (area, 14.421 sq. mm.) was somewhat different from its designed size ($10\sqrt{2}$ sq. mm.) a slight correction is necessary to reduce the values for the group of subjects treated in this paper to standard conditions. This correction (to be perhaps unnecessarily accurate) consists in *decreasing* T by 0.22; or *increasing* the time by 2.0 per cent.

succession. There is the possibility of a fatigue effect, or of an effect of learning or practice. These two may be expected to affect the result in opposite ways, and may predominate, one or the other, irregularly; or may possibly tend to offset each other.

The average results for the three subjects first discussed follow, accompanied by the number of items entering into the average and by the mean variation. The average values

Table I

Measures of Fluctuation of the Result of a Single Series for Three Practiced Subjects

T—Threshold, on arbitrary scale, reduced to standard aperture (area $= 10\sqrt{2}$ sq. mm.).

$\sigma_{T/M}$—Standard deviation, relative to the mean for like conditions.

E_T—Probable error, in terms of arbitrary units, as T.

The last three columns give the times corresponding to T, and $T \pm E_T$ respectively.

Subject	T	$\sigma_{T/M}$	E	Time, Thousandths	
				Mean	Limits of $P.E.$
B.........	23.70	0.062	.99	25.7	23.6 to 27.9
C.........	26.85	0.056	1.01	19.5	17.9 to 21.3
L.........	23.48	0.049	.78	26.2	24.5 to 28.0

(M) are all in the neighborhood of unity, because, for the purpose of comparison, each item was, before averaging, divided by the mean of all results (T) for identical physical conditions. The figures given, both threshold and mean variation, are therefore expressed in terms of the general mean. Similar results, for a group of 33 unpracticed subjects (the first of a group of 96 subsequently to be discussed), appear in the fourth line. Here each item has been reduced to the average of the individual's four consecutive series as unity. (Table II.)

It will appear from the above that on the whole the variabilities of the successive results diminish somewhat, although in the case of the unpracticed subjects the change is in the opposite direction, but not to a significant extent. It appeared, however, that it would be out of the question, in using the method as a test, to expend the time necessary to run four or even three series as a routine matter. That consideration will therefore be dropped, and only the first and second series will be considered in what follows.

The thirty-three subjects just mentioned were the first of a series of 101, which will form the basis of the major part of this discussion. Visual acuity was taken, at the time

Table II

Comparative Values and Variabilities of the Results of Four Consecutive Series

B, *C*, and *L*, are three practiced subjects; and the last row gives the results for 33 unpracticed subjects.

N—Number of series concerned in the average.

M—Mean of all values, relative to the mean of all series for the same subject.

MV—Mean variation expressed in the same terms as *M*. These values, multiplied by 1.253, should be comparable with the values of σ stated in the third column of Table I. The four series consumed an hour's time, more or less, according to the practice and natural aptitude of the subject.

Subject	Series 1			Series 2			Series 3			Series 4		
	N	*M*	*MV*	*N*	*M*	*MV*	*N*	*M*	*MV*	*N*	*M*	*MV*
B.......	18	.996	.048	18	1.002	.069	18	1.003	.049	18	.993	.037
C.......	22	.985	.057	22	.998	.058	22	1.005	.042	22	1.012	.046
L.......	21	1.016	.058	21	1.020	.036	21	.981	.042	21	.964	.030
Various..	33	.993	.044	33	.992	.046	33	1.019	.054	33	.996	.050

of the retinal sensitivity measurement, in 100 of these with each eye separately, and in 91 of these, with both eyes together. Unfortunately, in 5 cases of the 91, retinal sensitivity proved to be too low to measure with the apparatus as constructed, and rather than deviate from the established technique, it was decided to record these simply as 'too low to measure.' Some idea of the probabilities in these five cases may be gained from the fact that the lowest measurement actually made in conformity with the established technique was $T = 12.7$, and the three next 14.3, 14.5, and 14.7, respectively.

The distribution of 192 measurements, two upon each of

96 subjects, is shown in Fig. 1. The small circles represent values of the 'sliding' averages of four of the original class-

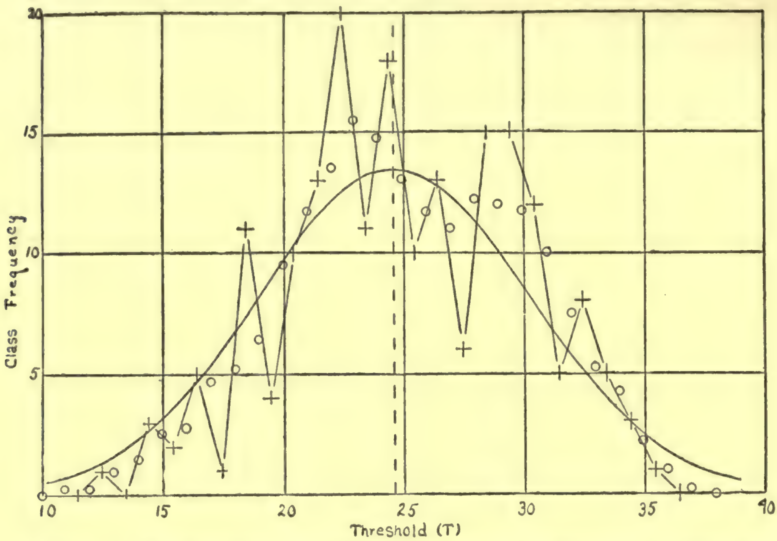


FIG. 1. Distribution of 192 measurements of retinal sensitivity; the values obtained from the first and second consecutive series of each of 96 different subjects. Abscissæ are (*T*). Class interval = 1. Class types are integral values + 0.4. Ordinates are class frequencies.

(*a*) Crosses and solid line—Frequency by classes.

(*b*) Circles—'Sliding' averages of four class frequencies.

(*c*) Smooth curve—Closest fitting normal distribution.

Ten of a possible 202 measurements fail to appear in (*a*) and (*b*) owing to the fact that they were too low to be measured with the apparatus used. These were taken into account in computing the constants of the normal distribution (*c*). See text.

frequencies and the smooth curve is that of a normal distribution based upon characteristics of distribution derived as follows: (1) A mean drawn from the whole of the 192 measurements less the 10 highest, omitted to offset the absence of the corresponding 10 lowest of 202 possible measurements, which are not included in the 192 because of their being too low to measure; and (2) a measure of dispersion derived from the sum of the deviations of that fraction of the distribution lying above the mean.

The first of the curves in the diagram (*a*) exhibits irregularities due to the method of interpolation by which a numerical threshold was derived from the results of one series. This method favored the even integers; that is, of the 192 interpolated values, 119 were even integers plus a fraction, and 73 only were odd integers plus a fraction. Inspection of the figure will show this irregularity. However, the circles representing the average frequency of four consecutive classes, plotted against the average of the corresponding four class-types, represent values from which this systematic irregularity is eliminated, and among which other irregularities are also, in part, evened out. Nevertheless, in comparing the line of the circles with the curve of normal distribution, it would appear that there are in the former two maxima: one at about the point where $T = 23$, ($t = 27\sigma$) another at $T = 29$, ($t = 16\sigma$); with a minimum between them at $T = 27$, ($t = 19\sigma$). In other words, the plot suggests very strongly that the distribution of results is a bimodal one.

The relation existing between the results of the first and second series on each of the 96 subjects is shown in Fig. 2,

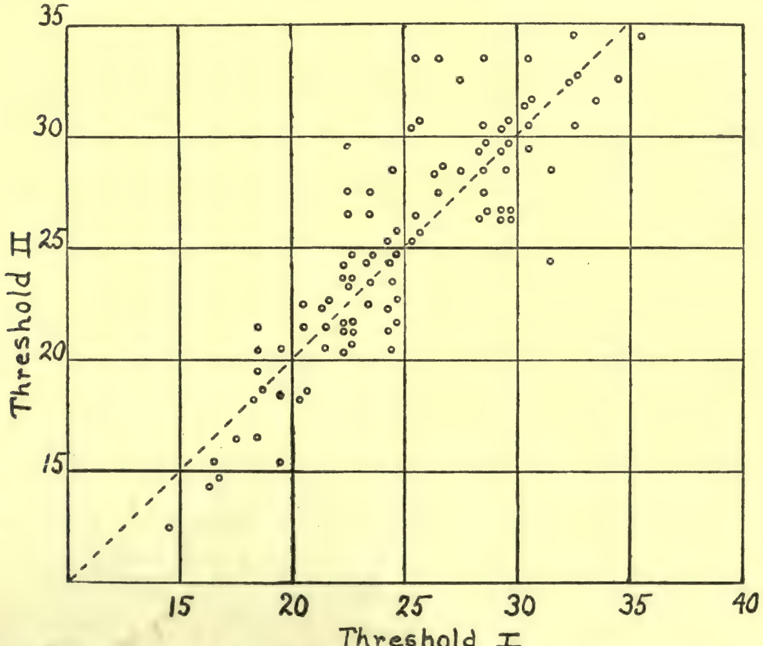


FIG. 2. The relation between the results of the first (I.) and second (II.) series, for each of 96 subjects. The oblique dotted line represents the condition of equality. Those plotted below this line gave a result on the second series lower than that of the first, and *vice versa.*

where the second of these is plotted directly against the first. The two means are I, 24.91 and II., 25.17; very close together. The oblique dotted line, drawn at 45° inclination, is the locus of the cases in which the two are exactly equal. The mean values just stated would indicate a tendency in the subjects to go higher on the second series, and a glance at the figure will show this tendency in general to be greater in those above the mean than in those below it and, in fact, to be reversed in a few cases at the other extreme. This shows itself also in the two standard deviations. That for the first series is 4.6, for the second 5.1. The correlation-ratio is 0.857 ± 0.018.

The characteristic equations derived from the above are:

$$x_1 = 0.77x_2, \qquad E_1 = 1.6; \text{ and}$$
$$x_2 = 0.96x_1, \qquad E_2 = 1.8;$$

in which the two probable errors (E) may be taken as an index of the variability of these inexperienced subjects, as a group, for the first and second consecutive trials respectively. This indicates a lower degree of reproducibility for the group of 96 subjects than for the three considered in Table I. There the values for E are of the order of 1.00 or less. The group of 33 various observers will be seen to compare very favorably with the three observers, *B*, *C*, and L (Table II., *M. V.*' s).[1] They therefore appear to show a better reproducibility than the group of 96 as a whole, of which they are a part.

This is incidental to the rather unusual way in which the method of correlation[2] has here been applied, for the purpose of estimating, from the results of a group, the expected standard deviation of repeated identical measurements upon an individual. In plotting the second set of results, *b*, against the first (*a*) as abscissæ, it will be seen that although the *a*'s are of identical class type for any column of *b*'s, yet in the case of any pair plotted in the column, the *a* is still in every case a variate from a possible mean, from which it deviates, on the whole, by an amount represented by its

[1] The probable error for all results of the group of 33, Table II., is 1.05, on the same basis as the values for E in Table I., for subjects *B*, *C*, and *L*.

[2] The correlations discussed here were performed by Pearson's method.

own standard deviation. Thus the standard deviation from their characteristic line of the b's ($\sigma_{2.1} = \sigma_2 \sqrt{1 - r_{12}^2}$) is not the standard deviation to be expected of repeated identical measurements, but is greater than the latter, as the standard deviation of a (sum or) difference exceeds that of either term; by the factor $\sqrt{2}$ when the two are of equal variability. Accordingly the two E's stated above should be divided by $\sqrt{2}$, and they then become $E_1 = 1.1$, $E_2 = 1.3$, which conform fairly well to the results stated in Tables I. and II.

Of especial interest in determining the value of this test is the question of the correlation existing between its results and those of the ophthalmologic tests. And of these, the relation with visual acuity, taken by the Snellen letters, was of especial interest. Visual acuity was taken for each eye separately, and for binocular vision, by hanging a test-letter card upon the screen of the apparatus, so that the conditions of the retinal sensitivity test were exactly reproduced for the letter test. The results of the correlations, performed by Pearson's method, are given in Table III. An incidental question, worked out by way of side issue, is the relation of binocular visual acuity (V_0) to the two values of monocular visual acuity obtained with the better eye, (V_1), and the poorer eye, (V_2), respectively. The equation for this value, derived from the correlation-ratios of the first order and the standard deviations of the second order, involving the variables 2, 3, and 4 only, is as follows:

$$V_0 = 0.75 V_1 + 0.16 V_2 + 0.18, \qquad (1)$$

the standard deviation of V_0 is $\sigma_{2.34} = 0.13$. It is to be remembered that the bulk of the results fell about the region of $V = 20/30$ to $V = 20/15$ (6/9 to 6/4.5 metric or 1.00 to 1.33 decimal) and the application of this equation is not safely to be extended outside these limits.

The interpretation of a characteristic equation such as the above is that V_0, calculated from the observed values of V_1 and V_2, is the expected value of binocular acuity, and that experimentally determined values may be expected to deviate from this as implied in the accompanying standard deviation, $\sigma_{2.34} = 0.13$.

Similarly, the characteristic relation between retinal sensitivity (T) and the three values of visual acuity above enumerated is:

$$T = 5.8V_0 - 2.3V_1 + 4.0V_2 + 16.2; \qquad (2)$$

and the standard deviation of T is $\sigma_{1.234} = 4.24$; the same re-

Table III

The Results of Correlation of Retinal Sensitivity (1) with Visual Acuity (2, 3, and 4) Measured with the Snellen Letters under the Conditions of the Experiment

The results from the letter chart have been reduced to decimal values (*e.g.*, 20/20 = 1.00; 20/15 = 1.33; 20/20 + 5/8 = 1.00 + 5/8 × 0.33 = 1.21) and the values of σ in division c of the table are expressed in such decimal units. Retinal sensitivity is expressed in arbitrary scale units, the same as those in which T is elsewhere expressed.[1]

ss—subscript indicating the variables involved and their interrelation.

N—number of cases involved.

r—correlation-ratio.

σ—standard deviation.

(*a*) *Numerical Designations of the Variables:*

1. Retinal sensitivity (mean of I. and II. series.)
2. Visual acuity (Snellen), binocular vision.
3. Same, better eye alone.
4. Same, poorer eye alone.

(*b*)

Correlation-ratios

ss	N	r, Zero Order	E_r	*ss*	r, First Order	*ss*	r, Second Order
12	86	0.375	0.062			12.34	0.177
13	95	0.321	0.062			13.24	−0.068
14	95	0.355	0.060			14.23	0.147
23	91	0.871	0.017	23.4	0.716	23.14	0.716
24	91	0.723	0.034	24.3	0.191	24.13	0.160
34	100	0.760	0.029	34.2	0.384	34.12	0.389

(*c*)

Standard Deviations

ss	N	σ, Zero Order	*ss*	σ, Second Order	*ss*	σ, Third Order
1	95	4.62			1.234	4.24
2	91	.271	2.34	.131	2.134	.129
3	100	.275	3.24	.125	3.124	.125
4	100	.247	4.23	.157	4.123	.155

[1] For a more complete explanation of the meaning of these symbols and of the methods of correlation see: Yule, An Introduction to the Theory of Statistics, Fifth Edition, London, 1919; Chap. IX ff. and Chap. XII.

servations as to the limits of applicability of this equation to apply, as in the case of equation (1).

These relations exhibit one or two points worthy of note. From equation (1), it will be noted that binocular visual acuity depends in a major degree upon the vision of the better eye. Whether it depends at all upon the vision of the inferior eye is an important question. V_2 being ordinarily somewhere about unity, the value of the V_2 term will be about 0.16, which is not greatly in excess of the standard deviation of V_0 from its value computed by the equation. However, while the plotted results show many cases where binocular vision was distinctly better than the vision of the better eye alone, there are but 4 cases in a total of 91 where the reverse is true, and this only to the extent of one class interval, equal to two single letters on the chart, or decimally expressed, equal to a difference of from 0.04 to 0.15 according to the portion of the chart involved. The mean value for V_0 is 1.22; for V_1, 1.16; and for V_2, 1.06; which also would indicate a definite extent to which, on the average, vision of the one eye is helped by the second, inferior, eye in binocular vision.

As to equation (2), the negative coefficient of V_1 would seem to indicate that the better eye has a negative contributory value to retinal sensitivity. By the same reasoning as the foregoing we see that in general the amount of the second term, V_1 being about unity, does not exceed the standard deviation of T (4.24). We must also bear in mind that the three V's bear high correlations, each to the other. The writer would interpret the equation as meaning that the better eye contributed to the result in the case of retinal sensitivity (that is, reinforced the poorer eye) only in so far as it does the same with reference to the letter chart; that is, only in so far as it coöperates with its inferior fellow in other respects. Retinal sensitivity is shown by the equations to be more highly dependent upon V_0 than upon either V_1 or V_2; which is perhaps to be expected, for retinal sensitivity was also taken binocularly.

This equation answers the question as to whether this test is to be looked upon (as indicated in an early paragraph of this paper) as supplementary to other tests in (*a*) a confirmatory sense or (*b*) as measuring something not already taken into account.

For the measures of retinal sensitivity in the 95 cases taken into account in the correlations, the standard deviation (σ_1) is 4.64. Equation (2) should give for T the closest value possible to compute from all the data on visual acuity. The variations of the actual measurements from this computed value are represented by $\sigma_{1.234}$, following equation (2), and equal to 4.24. Thus the attempt to express retinal sensitivity in terms of visual acuity is futile. It still deviates in much the same degree from the computed values, and we must conclude that retinal sensitivity, measured by the present method, involves a new characteristic of vision that is not to any significant extent taken into account in the measurement of visual acuity by means of the letter-chart.

In examining the relation of the retinal sensitivity results to the ophthalmologic findings as a whole, it was necessary to establish some reasonable means of numerical rating as to the latter. The ophthalmologic examination, of which the results were used, was part of the general medical examination of candidates for flying status, known at the Medical Research Laboratory as '609,' from the form number of the blank used to record the results. The criteria of ophthalmologic qualification, as practiced at the time this work was done, are here reproduced. The method of rating, as will be seen, was to score 1 against the candidate for a contingent disqualification, as implied in the following paragraphs, 3, 5, 8, and 9; to score 2 for disqualification in one particular; and to score 1 for each further point of disqualification. The rules following were interpreted literally. Following each paragraph, in parenthesis, are such supplementations and changes as were necessary to make the rating just and unequivocal, or remarks necessary to explain the nature of the test to one unfamiliar with the special practice.

1. *Visual Acuity.*—The minimal visional requirement for each eye is 20/20. If two or three letters are not read in the 20/20 line they may be offset by an equal number of letters read in the 20/15 line.

(The ratings taken in the experimental room were used, not those furnished by the ophthalmologist. In view of the qualification in the foregoing paragraph, a subject was passed if he read 20/20 less two letters, assuming in the absence of record of the fact, that he was able to offset these by two read in the 20/15 line. The decimal value 0.92 or over satisfied this condition.)

2. *Depth Perception at 6 Meters.*—An average depth difference of more than 30 mm. disqualifies the applicant.

(Two rods, 1 cm. in diameter, separated 6 cm. laterally, must not be more than 30 mm. apart in the direction of vision when set to apparent equidistance by the candidate. The rods were seen in silhouette against a bright background, and seen through an opening which did not permit vision of either top or bottom of the rods.)

3. *Maddox Rod Screen Test at 6 Meters.*—Esophoria of more than 4^{Δ} is a disqualifying factor if associated with less than 4^{Δ} of prism divergence, or if associated with diplopia in the lateral positions on the tangent curtain, or if associated with an amount of accommodation near the lower limits, or if associated with an amount of hyperopia near the disqualifying limit.

4. Esophoria of more than 10^{Δ} is a disqualifying factor, even if unassociated with any of the preceding conditions.

5. Exophoria of more than 2^{Δ} is a disqualifying factor if associated with an angle of convergence near the disqualifying limit, or if associated with diplopia in the lateral positions on the tangent curtain.

6. Exophoria of more than 5^{Δ} is a disqualifying factor even if unassociated with any of the preceding conditions.

7. Hyperphoria of more than $\frac{1}{2}^{\Delta}$ disqualifies the applicant without further supporting evidence.

8. *Maddox Rod Screen Test at 33 cm.*—Exophoria of 4^{Δ} may be considered the normal condition. Any considerable variation from this condition is to be interpreted in connection with the other associated tests. An exophoria of more than 12^{Δ} at 33 cm. disqualifies.

(1 to 7^{Δ} exophoria at 33 cm. were taken as the normal condition. Outside of these limits a question was scored.)

9. *Prism Divergence.*—Prism divergence of more than 9^{Δ} disqualifies the applicant if associated with an angle

of convergence near the disqualifying limit. If less than 4^Δ of prism divergence is found associated with more than 4^Δ of esophoria at 6 meters, the applicant is disqualified.

(Those grades of muscle abnormality which do not disqualify except contingently upon some other finding were marked with a question. One or more questions in the absence of final disqualification scored 1 only against the applicant.)

10. Prism divergence of more than 15^Δ disqualifies without further supporting evidence. Prism divergence of less than 2^Δ disqualifies without further evidence.

11. *Test of Associated Parallel Movements.*—The applicant is disqualified if the underaction or overaction of any of the extrinsic ocular muscles causes diplopia except in the extreme positions, where a small separation of the images may be disregarded. Nystagmus disqualifies if it is demonstrated except in extreme positions.

12. *Inspection of the Eyes.*—Any pathological condition which may become worse or interfere with the proper functioning of the eyes under the fatigue and exposure of flying, disqualifies the applicant.

13. *Accommodation.*—Accommodation is normal if it lies between limits 2 diopters above and below the mean for the applicant's age. Failure to read within these limits disqualifies.

TABLE OF MEAN VALUES OF ACCOMMODATION POWER

Age	Dptrs.	Age	Dptrs.	Age	Dptrs.	Age	Dptrs.
18	11.9	25	10.2	31	8.6	37	6.8
19	11.7	26	9.9	32	8.3	38	6.5
20	11.5	27	9.6	33	8.0	39	6.2
21	11.2	28	9.4	34	7.7	40	5.9
22	10.9	29	9.2	35	7.3	45	3.7
23	10.6	30	8.9	36	7.1	50	2.0
24	10.4						

14. *Angle of Convergence.*—An angle of convergence smaller than 40° disqualifies.

15. *Central Color Vision.*—If it is apparent that mistakes made by the applicant are due to color confusion and not to carelessness or failure to understand instructions, he is disqualified.

(Color-vision was left out of consideration in grading the subjects. There were but three cases not definitely found normal as to central color vision as follows: (1) no record as to color vision. Record otherwise incomplete, and subject therefore excluded from consideration; (2) no record as to color vision, disqualified on account of visual acuity and hyperphoria; (3) found to confuse reds and greens. No other disqualification or question.)

> 16. *Field of Vision for Form and Color.*—The normal visual field for form is largest; those for blue, red, and green are successively smaller in the order given. The color fields should be nearly concentric with the form field. Any marked contraction of the form field disqualifies the applicant for flying.
>
> 17. *Refraction.*—The applicant is disqualified if he cannot read 20/20 without more than one diopter of correction, either hyperopic, myopic or astigmatic.

Concerning paragraphs 11, 12, and 16, it was assumed that no disqualification existed unless specified. The records furnished the writer did not ordinarily state the findings on these points.

Of a total of 101 subjects upon whom retinal sensitivity measurements were made, there were 59 with records complete as indicated above, and only these are included in what follows.

It will be seen from the plot (Fig. 3*a*), that those showing the highest number of points as to ophthalmologic defects have, in general a low retinal sensitivity; curve *b* shows the average number of points of defect for each class as to retinal sensitivity; and curve *d* the average retinal sensitivity for each particular score. Both of these curves bear out the first statement made. The correlation ratio is -0.444 with a probable error of ± 0.070.

If, however, we exclude those having an ophthalmologic score of 4 or more against them, we get a correlation ratio of 0.009 ± 0.094 which is no correlation at all. The curve (*c*, Fig. 3) shows that there is no significant relation between retinal sensitivity and the ophthalmologic findings when those defective with a score of 4 or more are excluded.

We must conclude, therefore, that there is no relation between the defects elicited by the ordinary ophthalmologic examination on the one hand, and the present retinal sensitivity rating on the other; excepting those cases in which the clinical defects are comparatively numerous, in which these are signs of some rather general, perhaps distinctly pathological, derangement of the visual organs which comes to be generally reflected in their various functions, of which retinal sensitivity is one.

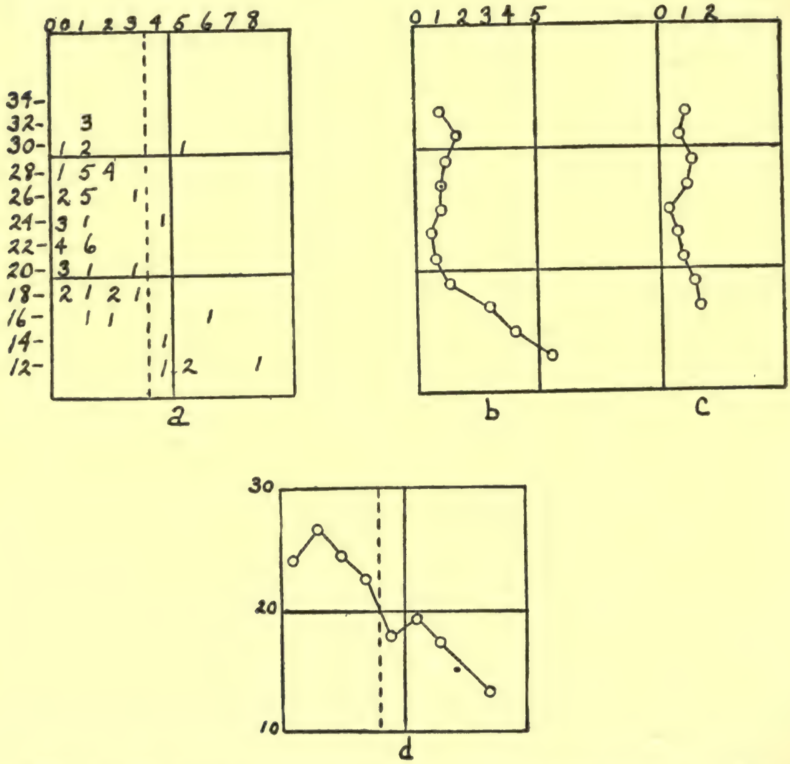


FIG. 3. The relation between retinal sensitivity measurements and the general degree of visual defectiveness found in the ophthalmologic examination.

(*a*) Classified results. Ordinates indicate class as to retinal sensitivity. The class-interval is 2 units on the scale, and the mean of the class is greater than the designation given by one unit. *E.g.*, in the 30-class, the lowest possibly included is 30.00, the highest 31.99, the mean (or type of the class) therefore is 31.00 minus. Abscissæ, points disqualification; see text.

(*b*) Average points of disqualification for each retinal sensitivity class.

(*c*) Same as (*b*), scores of 4 or over excluded.

(*d*) Average retinal sensitivity (ordinate) for each class as to points disqualification.

The cases falling to the right of the dotted line in (*a*) and (*d*) are those excluded from (*c*).

This conclusion will perhaps be borne out by inspection of the following synopsis of all results in the 59 cases considered. Where retinal sensitivity was too low to measure this is indicated by the word 'low' in place of the numerical result. Otherwise the synopsis should be self-explanatory:

Summary of Cases

Serial No.	Ret. Sens.	Clinical Description		No. of Cases
—	—	*Unquestioned, Score 0*		16
—	—	*Questioned, Score 1*		
		Esophoria	5	
		Exophoria	22	
		High prism divergence	7	
			34	
		Counted twice or more, less	9	
			25	25
		Just disqualified, Score 2		
101	29.8	Visual acuity	1	
62	17.0	Depth perception	1	
80	28.2	Exophoria at 6 meters	1	
19	19.9	Low convergence	1	
90	28.6	Hyperopia, 1 diopter, esophoria	1	
100	28.7	" 3 "	1	
16	18.6	Hyperopic astigmatism, corrected	1	
			7	7
		Disqualified, Score 3		
42	19.0	Visual acuity and hyperphoria	1	
83	21.7	Hyperopia, low accommodation	1	
91	27.8	Hyperopic astigmatism, low accommodation	1	
			3	3
		Disqualified, Score 4		
		Hyperopic astigmatism, poor depth perception, with:		
55	15.6	Convergent strabismus	1	
46	Low	Low visual acuity	1	
82	25.9	Deficient accommodation	1	
			3	3
		Disqualified, Score 5		
87	31.1	Hyperopia, exophoria, hyperphoria, low convergence and accommodation	1	
85	Low	Hyperopic astigmatism, low visual acuity and depth perception, hyperphoria	1	
77	Low	Hyperopic astigmatism, low visual acuity, depth perception and accommodation	1	
			3	3

Serial No.	Ret. Sens.	Clinical Description	No. of Cases
		Disqualified, Score 6	
39	17.5	Myopia, corrected; low visual acuity, depth perception and convergence; exophoria.......... 1......	1
		Disqualified, Score 8	
63	Low	Myopic astigmatism, strabismus, low visual acuity, accommodation and depth perception. Impossible to measure convergence................ 1......	1
		Total number of cases..........................	59

Summary of Conclusions

1. Measurements of retinal sensitivity made upon 101 subjects by a new method, previously described in detail, show individual differences far in excess of the accidental variations for the same individual and the same conditions.

2. The result of any one of as many as four consecutive series is of about equal reliability even with unpracticed subjects.

3. The distribution of the various results of retinal sensitivity measurement suggests a natural division of individuals into two groups, most numerous where the times of stimulation are approximately 27 and 16σ respectively.

4. Correlation of the results of the first series with those of the second, taken immediately after the first, in 96 subjects, shows almost identical characteristics for the two. The correlation ratio is 0.857 ± 0.018.

5. The characteristic equation derived from correlation of the retinal sensitivity results (average of first and second series) with visual acuity indicate that retinal sensitivity is somewhat dependent upon the visual acuity of the eye with poorer vision, but not upon that of the better eye unless the advantage of this latter is evident in enhanced binocular visual acuity.

6. Due allowance for the differences due to different visual acuity leaves individual differences in retinal sensitivity measurements unaccounted for, almost to the original amount. The standard deviation from 4.64, is reduced only to 4.24 when correction is made for the differences due to different visual acuity.

7. Correlation between retinal sensitivity and a rating based upon the ophthalmologic portion of the Air Service medical examination for flying status, indicates that in general no definite relation exists between retinal sensitivity and disqualification; except when the latter is based on findings which, in degree and character, indicate distinctly pathological conditions.

8. Incidentally, the method of correlation indicates that binocular visual acuity is dependent chiefly upon the eye which, used singly, shows the better vision, to the extent of nearly five times its dependence upon the eye with the poorer vision. This has exclusive reference to results obtained with the letter-chart.

Appendix

The apparatus used in this work differed from that used in the former work in the substitution of a perforated and painted metal slip for the punched card formerly used for the test-stimulus aperture. It was found that simply drilling and painting the slip would not suffice, since the paint tended to thin itself, by its surface tension, at the margins of the aperture, causing a ring-shaped shadow by which a photometric balance between screen and aperture was rendered impossible.

The following will give an idea of the technic found necessary in preparing replaceable apertures free from such defect:

Sheet zinc of 0.017 inch thickness was found suitable for the purpose. The hole at the center was drilled to admit the end of a reamer (known to the trade as a 'No. 1 taper pin reamer'), provided with a wooden sleeve adjusted so that the hole, reamed from the back until the sleeve just touches the metal, was of just the right size to fit a carefully sized gauge plug. This plug was made for the purpose and measured 0.1687 inch (4.285 mm.) in diameter. The margins of the aperture were then freed from burrs by passing a moderately fine file flat over the front and back surface of the metal.

Before painting the slips, the back of the hole was covered with surgeon's adhesive, and the hole filled with Alabastine (water) paint, mixed to a putty consistency. Beeswax was tried and worked almost as well. The filling was pared down flush with the face of the zinc by means of a clean-edged carpenter's chisel, and the face of the zinc cleaned free of excess filling material at the same time.

The face of the slip was then painted. Two coats of a good ordinary white oil paint were applied in the usual way, and as these nearly covered the metal, a coat of alabastine white was applied. The filling was then removed by first stripping the adhesive from the back and locating the center of the hole by pricking with a needle, and then carefully punching the filling out, from the face, piecemeal, without touching the parts which would form the margins of the hole. After this the reamer was run into the hole from the back, as in the first place, carefully and *only* until the collar on the reamer lightly touched the zinc.

The object of this procedure was that the edge of the opening should be perfectly sharp, and that the surface of the face of the slip should continue to be plane, accurately and exactly, until it met the cylindrical (slightly conical) surface cut by the reamer. If carefully and successfully done, it was readily possible to make the aperture disappear by adjusting the brightness of the back screen.

Considerable care had to be used in handling the finished slips, as they were rendered useless by the slightest soiling or chipping of the paint at the margins of the aperture.

ADULT TESTS OF THE STANFORD REVISION APPLIED TO UNIVERSITY FACULTY MEMBERS

BY HELEN HUBBERT CALDWELL

Dayton, Ohio, May 8, 1922

Object of Research.—The present investigation was the outgrowth of one previously made on the application of the Stanford Adult Tests to College Students,[1] and was carried on in the spring of 1919 while the writer was instructor in educational psychology at the University of Wisconsin. Its object was threefold. First, to determine the possible limits of performance for adults especially chosen because of reputed brilliancy or genius. Second, by recording the percentage of failures in each part of the test and relating them to the performance as a whole, to approach a better evaluation of the separate tests. Third, to test the validity of the assertion that there are significant differences in the results obtained from persons inclined to literary pursuits and persons inclined to scientific pursuits, the former being markedly superior in the so-called Verbal Tests, while the latter excel in what are called the Constructional Tests.

Securing Subjects.—The writer had not realized until she attempted to secure subjects for this investigation, how widespread and ofttimes violent among cultured and educated persons was the opposition to mental testing of any sort. Attempts to get at the reasons for such opposition were in most instances fruitless. Beyond the stock objections, "mind cannot be measured," and "mental tests take no account of individuality," not many would express themselves. They were content to condemn mental testing on general principles and refuse to become subjects 'even in the interests of science.'

[1] Caldwell, Helen Hubbert, 'Adult Tests of the Stanford Revision Applied to College Students,' *Jour. Ed. Psychol.*, Dec., 1919, Vol. X., No. 10, p. 477.

The first intention, which was to have only heads of departments and full, associate and assistant professors as subjects, had to be abandoned, and in four instances it was necessary to go below the rank of instructor. Records for three student assistants and one student fellow are included, but each of these students had an unusually brilliant record in his particular department, and might be classed as a 'genius' in the ordinary meaning of that term. Neither was it possible to secure as many subjects as desired. The original plan of the research called for thirty subjects in each group, sixty in all. Slightly over half of that number were actually obtained.

So far as was possible, faculty members chosen were those whose predilections for scientific or for literary pursuits were well established. The effectiveness of personal appeal as against telephonic communication was demonstrated in the securing of subjects. Attempts to arrange for the tests by telephone were without exception unsuccessful. Personal interviews with about seventy faculty members, in which the purpose of the investigation was explained, resulted in but thirty-six subjects.

Opinions Regarding the Tests.—Reasons—or shall we say excuses—for not coöperating were plentiful and some of them quite amusing. Most frequent was the plea, "I haven't time," and this was allowed to stand even after it was explained that but 45 or 50 minutes would be required. In nearly every case where coöperation was secured it was the declaration, "We aren't really testing *you*, you see; it's the *tests* we are testing; we want to see if they really do show differences between you literary folks and the scientists" (or *vice versa*, as the case might be), which finally won consent. Once the experimenter was soundly berated for even venturing to request the assistance of a very noted professor. She was told that his time was precious, that he was protected by University rulings from the annoyance of book agents, solicitors and peddlers and should be so protected from her intrusion. In another instance, refusal to take the tests was apparently based upon the pretense that it would be such a 'waste of time,' since 'no mental tests were worth anything anyhow." But the same gentleman added, "I may be a fool, but I don't intend to let anyone find it out if I can help it!" No doubt in many cases refusal was due to an unacknowledged fear of failure and the functioning of a 'defense mechanism.'

By no means all of those who finally agreed to submit to the experiment were friendly to mental testing. In at least four cases the tests were declared to be 'utter nonsense.' One subject considered all mental tests 'vicious,' while another characterized them as 'piffle.' Attempts to draw these reagents out and get reasons for their opinions failed. In four instances the attitude was one of complete boredom, while three reagents submitted with an air of amused toleration. Three subjects were apparently willing enough, but seemed unable to concentrate on the work in hand. Six were willing, but extremely self-conscious. The remaining subjects were alert, willing, and coöperated exceedingly well, with one exception where there was utter failure to coöperate —fortunately the only case of its sort which was encountered. This record is not included in the averages or results, since only the average adult tests were attempted. Tests 1, 2 and 3 were passed, 4 was not passed, 5 was a complete failure, 6 the subject positively refused to attempt, saying, "I am not interested, don't want to put my attention on it." In alternate 1 '*a*' was missed but '*b*' was correct. '*a*' of alternate 2 was missed, '*b*' was correct, '*c*' was not attempted. Instead the experimenter was told, "If you want my real reaction to these problems they bore me stiff." At this point the test was discontinued!

One subject declared himself as unalterably opposed to all attempts at mental testing "because they take no account of genius, and do not enable one to get at the *urge* behind the actual performance." When told that psychologists themselves were working on this very problem in an attempt to get at this element of effort or 'urge' by devising character tests, he replied, "It can't be done. I myself have spent hours trying to devise something of the sort." Though convinced that he could not 'get by' the immigration authorities at Ellis Island if required to take their tests, he emerged from the ordeal with the highest Intelligence Quotient it is possible for an adul to make, 121.8, and was the only one of the 36 subjects who did not fail on any part of the test. This instance and several others bear out Miss Downey's [1] statement that the reagent's score bore little relation to his own estimate of his ability.

The digit-span and other immediate memory tests called forth more objections than any others. In a number of cases, however, after the first refusal to try had been overcome, the tests were passed correctly. Several suggestions regarding the 'number tests' were made. Two subjects were confident they could pass them if they were presented visually instead of orally. A set of numbers so presented at the close of the test gave no better results with one of these subjects than the auditory presentation which is customary. The other subject, however, passed all three of the digit-span tests when the numbers were presented visually although he had failed on all three with the auditory presentation. A mathematics professor explained that mathematicians would not be apt to pass the digit-span tests because of their 'habit of thinking in wholes and generalities.' He himself passed them creditably. Only two of the subjects of the scientific group really objected to the 'number' tests, while seven of the literary group had to be coaxed to try them.

Results for all Subjects.—Two of the 36 records were discarded before making up averages, one in the literary group because of lack of coöperation, one in the scientific group because of six interruptions which occurred during the test. All tables, averages and conclusions drawn from the experiment are based upon the remaining 34 records. In most cases percentages and averages are given in whole numbers. The same notation is used as in the writer's previous paper on Stanford tests, the tests in the sixteen-year group being referred to as I—1, I—2, I—3, I—4, I—5, I—6, Al. 1 and

[1] Downey, June E., 'The Standard Adult Intelligence Tests,' *Jour. of Delinq.*, 1917, 2, 144–155.

Al. 2, and the eighteen-year-old tests being referred to as II—1, II—2, II—3, II—4, II—5, and II—6. Intelligence quotients are referred to as I.Q's.

The I.Q.'s for the entire list of faculty members tested range from 104.7 to 121.8. Only 23.5 per cent. fall below the Superior Adult level, and none below the Average Adult level. Nine per cent. rank as Very Superior. The average I.Q. is 114, 47 per cent. showing a higher I.Q. than this average while 50 per cent. fall below it. Seventy-six and five tenths per cent. reach or exceed the Superior Adult level. Time required for the tests varied from 22 to 60 minutes, the average time required being 40 minutes.

Table I

Showing, (*a*) the Number and Per Cent. of Failures on Each Test for the Entire Group; (*b*) the Number and Per Cent. of Failures for Those Having a Superior Adult Record; (*c*) the Number of Failures on Each Test in Miss Downey's Group of Wyoming University Faculty Members

(*a*) Entire Group Failures			(*b*) Superior Adult Failures		(*c*) Superior Adult Failures (Downey)
Test	No.	Per Cent.	No.	Per Cent.	No.
I—1....	0	0	0	0	0
I—2....	0	0	0	0	5
I—3....	2	6	1	3	1
I—4....	13	38	8	24	6
I—5....	17	50	8	24	4
I—6....	8	24	3	9	13
Al—1....	19	56	10	30	Not recorded
Al—2....	2	6	1	3	" "
II—1....	0	0	0	0	1
II—2....	13	38	7	20	10
II—3....	15	44	6	18	11
II—4....	5	16	3	9	10
II—5....	17	50	7	20	7
II—6....	3	9	1	3	8

Reference to Table I. (*a*) will show the percentage of failures on each test. Attention is called to the relatively large percentage of failures on tests involving immediate memory only, I—5, Al—1, II—3 and II—5. Even when the records of those below the Superior Adult level are eliminated this is still true. (See Table I. (*b*).) Note also the weakness on tests involving the use of the imagination, I—4 and II—2.

For a comparison of the rating of that part of the University of Wyoming faculty tested by Miss Downey with that part of the University of Wisconsin faculty tested, see Table I., (*b*) and (*c*), where the number of failures recorded in each test for those with an I.Q. indicating superior intelligence is tabulated.

Comparison of these results with those secured in the work with college students is interesting. (Table II.) In tests largely involving the verbal element the faculty ratings are distinctly higher—(I—1, I—2, I—3), in immediate memory tests [1] they are as low or lower—(I—5, Al—1, II—3, II—5), in the two tests which depend almost entirely upon use of the imagination (I—4, and II—2), faculty and students rate about alike. On the code test the faculty is slightly better, (I—6) for the logical memory and ingenuity tests, (II—4, II—6), the faculty record is considerably better, while the physical relations test, (Al—2), shows almost as

TABLE II

SHOWING THE PERCENTAGE OF FAILURES ON EACH TEST FOR THE SOPHOMORES AND JUNIORS OF RANDOLPH-MACON WOMAN'S COLLEGE AND THE FACULTY GROUP TESTED AT THE UNIVERSITY OF WISCONSIN

Test	Percentage of Failures Recorded		
	Sophomores	Juniors	Faculty
I—1	33⅓	12½	0
I—2	12½	12½	0
I—3	25	12½	6
I—4	37½	37½	38
I—5	33⅓	29⅙	50
I—6	29⅙	29⅙	24
Al—1	58½	62½	56
Al—2	83⅓	62½	6
II—1	70	50	0
II—2	25	42	38
II—3	42	12½	44
II—4	46	33⅓	16
II—5	54	70	50
II—6	50	62	9

[1] Miss Downey's 'digit-span tests' + Al.1.

great a contrast between faculty and students as the verbal tests. Comparing these results with Miss Downey's we also find the faculty to excel on vocabulary. We cannot agree that they are better on digits reversed, nor does it appear that our student group tested excelled on code, paper-cutting, digits forward or logical memory. In fact for the code and logical memory tests quite the reverse is true. It would not be safe to draw general conclusions regarding reactions of students as compared with faculty from either set of results.

Results for Separate Groups.—We turn next to a consideration of results obtained from the two separate groups in this problem. There were 17 subjects in each group. For convenience we shall call them the Literary Group and the Scientific Group. The former was composed of persons from the various departments of the University which might be termed *literary* in contradistinction to those definitely *scientific*, and included one librarian, two professors of History, one professor and two associate professors of English, one professor and one assistant professor of Latin, two professors, two assistant professors, one instructor and one assistant in Romance languages, one assistant professor of Semitic languages, one professor and one assistant professor in journalism.

The I.Q.'s for this group ranged from 104.7 to 121.8, the average being 114.3, with 47 per cent. above and 53 per cent. below the average. Average time required for the test was 41 minutes, with 47 per cent. completing it in less than the average time for the group and 51 per cent. requiring more than the average time. Reference to Table III. will show the number of failures on each test. Of the five Verbal Tests only one was missed and this by 3 persons; with the exception of II—2 results of the Constructional Tests are fair and results on the Immediate Memory Tests make the poorest showing.

Consider next the Scientific Group which was made up of one assistant professor of mechanical practice, one professor of steam and gas engineering, two associate professors, two instructors, one student assistant and one research fellow in physics, three assistant professors in chemistry, one assistant professor in agricultural chemistry, one assistant professor in agronomy, one instructor in manual arts, one associate professor of mathematics, one professor and one assistant professor of genetics. I.Q.'s here ranged from 107.3 to 121.8, the average was 113.7 and 53 per cent. were above while 47 per cent. were below this average. Time required varied from 27 to 60 minutes with 38½ minutes as the average and 59 per cent. requiring less than the average time for the group while 41 per cent. required more than the average time. Table III. shows the number of failures made in the Scientific Group on each test. Of the five Constructional Tests one is perfect and one was missed only once,

TABLE III

SHOWING THE NUMBER AND PERCENTAGE OF FAILURES ON EACH TEST BY EACH GROUP AND CLASSIFYING THE TESTS AS VERBAL, CONSTRUCTIONAL AND IMMEDIATE MEMORY TESTS. LIT. = LITERARY GROUP, SCI. = SCIENTIFIC GROUP, V = VERBAL, C = CONSTRUCTIONAL, I M = IMMEDIATE MEMORY

Test	No. Failures		Per Cent. Failures		Classification of Test
	Lit.	Sci.	Lit.	Sci.	
I—1	0	0	0	0	V
I—2	0	0	0	0	V
I—3	0	2	0	6	V
I—4	5	8	15	24	C
I—5	9	8	26	24	I M
I—6	3	5	9	15	C
Al—1	10	9	29	26	I M
Al—2	2	0	6	0	C
II—1	0	0	0	0	V
II—2	8	5	24	15	C
II—3	6	9	18	26	I M
II—4	3	2	9	6	V
II—5	9	8	26	24	I M
II—6	2	1	6	3	C

the showing on the other three being hardly so good as that on the Verbal Tests. As was the case with the Literary Group the Immediate Memory Tests are very poor.

Making now a somewhat more direct comparison of the results for the two groups, we find a number of facts of interest. First, as regards the attitude shown by the two groups, on the whole the Scientific Group made better subjects, were less opinionated, adopted a more impersonal attitude toward the tests and seemed more interested in the purposes and results of the experiment. There were fewer objections to tests in general and fewer criticisms of individual tests as 'inadequate,' 'tiresome,' 'unfair,' 'no test at all,' than in the Literary Group. There was also less annoyance evidenced when a test was not passed. Of course the experimenter followed the technique recommended by Terman and the results were not made known to the subjects, but in such a test as the vocabulary test and in the logical memory test and the digit-span tests the subject was pretty well aware of the character of his performance without being told.

With regard to I.Q.'s there is a difference of only 0.6 between the average I.Q.'s for the two groups, not a sufficient difference to bear any significance. As to the time required to complete the tests the Scientific Group were apparently somewhat speedier, their average being 38½ minutes, while that for the Literary Group was 41 minutes. Only five, or 30 per cent. of the Literary Group, completed the tests within the average time of the Scientific Group. One reason for this time difference may lie in the fact that more definitions were given in the vocabulary test, and more exhaustive explanations of the difference between abstract terms, on the part of the Literary Group than on the part of the Scientific Group. Another reason may lie in the fact already mentioned, that there were fewer comments made in the nature of objections or criticisms during the progress of the test by the Scientific Group than by the Literary Group.

If, now, we classify the tests according to Miss Downey's method and indicate the number of failures on each test for each group we have as a result Table IV.

We find, taking all the Verbal Tests into consideration, that there are 3 per cent. fewer failures on the part of the Literary Group than on the part of the Scientific Group, while with the Constructional Tests exactly the opposite is true and the Scientific Group makes 3 per cent. fewer failures

TABLE IV

CLASSIFYING THE TESTS AS VERBAL, CONSTRUCTIONAL AND IMMEDIATE MEMORY TESTS AND SHOWING THE NUMBER OF FAILURES IN EACH TEST BY EACH GROUP

Verbal Tests

	I—1		I—2		I—3		II—1		II—4	
	Lit.	Sci.	Lit.	Sci.	Lit.	Sci.	Lit.	Sci.	Lit.	Sci.
No........	0	0	0	0	0	2	0	0	3	2
Per Cent. .	0	0	0	0	0	6	0	0	9	6

Constructional Tests

	I—4		I—6		Al—2		II—2		II—6	
	Lit.	Sci.	Lit.	Sci.	Lit.	Sci.	Lit.	Sci.	Lit.	Sci.
No........	5	8	3	5	2	0	8	5	2	1
Per Cent. .	15	24	9	15	6	0	24	15	6	3

Immediate Memory Tests

	I—5		Al—1		II—3		II—5	
	Lit.	Sci.	Lit.	Sci.	Lit.	Sci.	Sci.	Lit.
No....................	9	8	10	9	6	9	9	8
Per Cent.	26	24	29	26	18	26	26	24

than the Literary Group. In the Immediate Memory Tests taken as a whole, there is a 1 per cent. difference in favor of the Scientific Group. So far, there seems very little evidence that the tests really enable differentiation between literary and scientific ability.

Let us, however, pursue the matter a little further and examine the results more closely by characterizing the performance of each subject on each test as *excellent*, *medium* or *poor*. Such a characterization of performance is necessarily based upon an arbitrary decision as to just what is meant by these standards, but since the same limits are set for both groups of subjects no unfairness can result from such a comparison. Arbitrarily, then, in the Verbal Tests we set the following standards, P standing for *poor*, M for *medium* and E for *excellent:*

Test I—1	Score 69 or below	= P
	Score 70 to 84	= M
	Score 85 to 100	= E
Test I—2	Score below 8	= P
	Score of 8	= M
	Score of 10	= E
Test I—3	Less than 3 contrasts	= P
	Three contrasts	= M
	Four contrasts	= E
Test II—1	Score 79 or below	= P
	Score 80 to 89	= M
	Score 90 to 100	= E
Test II—4	Neither passage	= P
	One passage	= M
	Both passages	= E

In the Constructional Tests our standards are as follows:

Test I—4	Less than 3 correct	= P
	Three correct	= E
	Four correct	= E
Test I—6	Two errors—More than 4 minutes	= P
	One error—3′ 1″ to 4 minutes	= M
	One error—3 minutes or less or Two errors—2 minutes or less	= E
Test II—2	No division made. Either E or P, meaning in this case failure.	
Test II—6	One problem	= P
	Two problems	= M
	Three problems	= E

The Immediate Memory Tests we divide as follows:

Test I—5	One correct	= P
	Two correct	= M
	Three correct	= E

Test Al—1	None correct	= P
	One correct	= M
	Two correct	= E
Test II—3	One correct	= P
	Two correct	= M
	Three correct	= E
Test II—5	One correct	= P
	Two correct	= M
	Three correct	= E

Analysis of results according to these standards of performance gives the figures shown in Table V. In the Verbal Tests we have the following averages: 12.2 of the responses of the Literary Group were excellent as against 9 of the

TABLE V.

SHOWING THE NUMBER OF *Excellents*, *Mediums* AND *Poors* FOR EACH GROUP ON EACH TEST, TESTS BEING CLASSIFIED AS VERBAL, CONSTRUCTIONAL, ETC.
E = EXCELLENT, M = MEDIUM, P = POOR

Verbal Tests	Lit.			Sci.		
	E	M	P	E	M	P
I—1	17	0	0	13	4	0
I—2	10	7	0	12	5	0
I—3	14	3	0	10	6	1
II—1	14	3	0	6	9	2
II—4	6	8	3	4	11	2
Averages	12.2	4.2	0.6	9	7	1
Percentages	71.8	24.7	3.5	53	41.2	5.8
Constructional Tests						
I—4	11	1	5	7	2	8
I—6	8	4	5	7	1	9
Al—2	5	10	2	12	5	0
II—2	5	—	12	12	—	5
II—6	7	8	2	7	9	1
Averages	7.2	4.75	5.2	9	3.4	4.75
Percentages	42.3	27	30.7	53	20	27
Immediate Memory Tests						
I—5	2	3	12	1	3	13
Al—1	2	6	9	2	7	8
II—3	1	5	11	0	3	4
II—5	2	2	13	0	4	13
Averages	1.75	4	11.25	0.75	4.25	9.5
Percentages	10.3	23.5	66.2	4.4	25	70.6

Scientific Group; there were 4.2 responses classed as medium in the Literary Group and 7 in the Scientific Group; 0.6 of the Literary Group's responses were poor as against 1 of the Scientific Group. Or, putting it another way we may say that in the Literary Group we find 61 responses out of a possible 85 graded *excellent*, or 71.8 per cent., 24.7 per cent. *medium* and 3.5 per cent. *poor*, while in the Scientific Group we find 53 per cent. excellent, 41.2 per cent. medium and 5.8 per cent. poor. Judged in this manner the Literary Group does show itself superior to the Scientific Group on the Verbal Tests.

For those tests termed *Constructional* we find an average of 7.2 excellent responses for the Literary Group *versus* 9 for the Scientific Group; the Literary Group averages 4.75 medium responses, the Scientific Group 3.4; there are 5.2 poor responses in the Literary Group compared to 4.75 in the Scientific Group. Stated in terms of per cent., the Literary Group furnished 42.3 per cent. excellent, 27 medium and 30.7 poor responses as against the 53 per cent. excellent, 20 per cent. medium and 27 per cent. poor responses of the Scientific Group. Here the advantage is slightly in favor of the Scientific Group, though it is not nearly so marked as is the better showing of the Literary Group in the Verbal Tests.

The Immediate Memory Tests show so slight a difference between groups as to make the working out of percentages scarcely worth while. However, we find the Literary Group showing 10.3 per cent. excellent, 23.5 per cent. medium and 66.2 per cent. poor responses, while the Scientific Group shows 4.4 per cent. excellent, 25 per cent. medium and 70.6 per cent. poor.

There is still another way of comparing these two groups, that of further analyzing some of the test results. The tests which lend themselves most readily to a more complete analysis than we have given are the Vocabulary, Code and Ingenuity Tests. If we look more closely at the results of the Vocabulary Test (I—1 and II—1) we find in the Literary Group a score ranging from 86 to 100 and time required from 2 to 8 minutes, averages being 95.3 for the score and 4′ 52″ for time required. It is in this test that we should expect to find the most marked superiority for the Literary Group and this seems to be the case. Scores for the Scientific Group range from 78 to 96, time required from 3 to 11½ minutes, with averages of 82 for the score and 6′ 12″ for the time. The Literary Group has an advantage, then, of 13.3 in score and 1′ 20″ in time in tests I—1 and II—1, both Verbal Tests.

Next, let us consider the results of the Code Test, I—6. Here we find errors for the Literary Group ranging from 0 to 8 with an average of 1.3. Time range is 1′ to 9′ with an average of 3′ 16″. The Scientific Group does not improve much upon these results, having an error range of 0 to 5, average 1.5, and a time range of 1½′ to 6′, average 3′ 30″. So we cannot grant any superiority in performance in this Constructive test to the Scientific Group.

Turning to another of the Constructional Group, the Ingenuity Test, II—6, and making a detailed analysis of its results we find '*a*' missed by one member of the Literary Group, but by none of the Scientific Group. Time range was from 20″ to 5′ for the Literary Group, average time 1′ 34″, while for the Scientific Group it was from 20″ to 4′ with an average time of 1′ 24″; '*b*' was missed by 6 of the Literary Group which required from 25″ to 5′ or an average of 2′ 9″ for this part of the test. Four of the Scientific Group failed on it, their time range being from 45″ to 5′, average time 2′. On '*c*' 7 of the Literary Group failed, time required was from 1 to 5 minutes, average time 2′ 42″, and 6 of the Scientific Group registered failure, time ranged from 45″ to 2½′ with an average of 1′ 26″. Combining the three parts of the test and taking the average for the whole we find the Literary Group requiring 6′ 25″ with 14 failures while the Scientific Group required 4′ 50″ and registered 10 failures, a difference in time of 1′ 35″ and in failures of 4. In this one of the Constructional Tests then, there does seem to be superiority on the part of the Scientific Group.

Conclusions.—Since none of the remaining tests are capable of so detailed an analysis, we shall not attempt further minute comparison, but will try to summarize briefly the results arrived at.

1. The first object of the research was "to determine the possible limits of performance for adults especially chosen because of their reputed brilliancy or genius." As was explained above, this purpose could not be completely carried out because of the difficulty of securing the desired subjects. Fifteen of those who took the tests may, however, be said to have the reputation in university circles of possessing exceptional ability. Tabulation of the I.Q.'s and the tests missed by these subjects appears below. (Table VI.) It

TABLE VI

SHOWING THE I.Q.'S AND TESTS MISSED FOR 15 SUBJECTS CLASSED AS EXCEPTIONALLY BRILLIANT

I.Q.	Tests Missed
121.8	———
121.8	Al—1
119.2	I—5
118.7	II—2
118.7	II—2
118.7	II—4
118.7	Al—1, II—2
116.6	I—4, I—5
116.1	I—6, II—2
114.0	I—4, I—5, I—6
113.0	I—5, II—2, II—5
113.0	I—4, Al—1, II—3, II—5
113.0	I—5, II—3, II—5
110.4	I—4, I—5, Al—1, II—3, II—5
109.8	I—4, Al—1, II—2, II—3, II—5

will be seen that in one instance the I.Q. falls slightly below that of the Superior Adult level, that in two cases it reaches the highest possible score for an adult, I.Q. 121.8. It was a matter of surprise to the writer that only one subject made an absolutely perfect record. True, failure on either Al—1 or Al—2 had no effect on the test score or the I.Q., but it was her expectation that more subjects would have an I.Q. of 121.8 and that more would have perfect records. The fact that one subject not classed as 'brilliant,' whose record therefore is not included with those of the fifteen mentioned, had an I.Q. of 121.8 and that several other reagents not included in the 'exceptional' list had higher I.Q.'s than some who do appear in it, leads again to a questioning of the real value of the I.Q. as now computed for adults. Certainly it is of no great help in close comparisons within a group such as we have been attempting here.

2. Recording the percentage of failures in each test as has been done in Tables I. and II. shows that, as was the case with college students, there are relatively too many failures on the Immediate Memory Tests to make them of real value for adults as they stand. It would be interesting to know whether visual presentation of these tests would result in fewer failures.

3. There is slight evidence that there is a difference in performance between the two groups, Literary and Scientific, on the two sorts of tests. This difference is more marked upon minute analysis of certain separate tests. It is not sufficiently great, however, to warrant the use of these tests as aids in assigning courses to college entrants, or in determining the 'bent' of any particular subject, unless supplemented by other tests.

4. One or two interesting problems growing out of this one might be mentioned as fields for future investigation. *a.*—Presentation of half of the Immediate Memory Tests visually and half auditorially, and comparison of the number of failures resulting under the two methods. *b.*—The large percentage of failures on I—4 and II—2, both of which depend so largely upon the imagination, suggest the giving of these tests to (1) a group of adults of college age who lack college training, (2) a group beyond college age who have no college training, (3) a college student group whose ages average about the same as those of group (1), (4) a faculty group corresponding in age and number to group (2). Comparison of results so obtained might show two things. First, the extent to which age affects the element of imagination; second, the extent to which training and education affect it. It might then be determined just how large a part tests of imagination *per se* should play in a test for General Intelligence.

CUMULATIVE CORRELATION

BY J. CROSBY CHAPMAN

Yale University

If we consider a variable x_1 to be dependent upon a number of other variables x_2, x_3, x_4, x_5, etc., we may weight, by the ordinary method of partial correlation, the values x_2, x_3, x_4, etc., so that the residual $x_{1.2.3 \ldots n}$ will be a minimum. Under these conditions the partial regression coefficient of x_3, for example, is dependent not only on x_2 but also on x_4, x_5, x_6, etc. While in many cases this is what the conditions of the problem demand, there are certain practical situations in which the following question is pertinent. Having decided to use a variable number of tests in a serial order and weighting the first test unity, what weight must be attached to the succeeding test in order that the two tests combined will exhibit the maximum correlation? Furthermore, supposing the first and second tests are weighted according to the above procedure, and regarding this combination as a single variable, what must be the weight attached to another variable so that the new cumulative correlation is a maximum. At each step, in addition to the weighting, the resulting correlation is desired.

To translate into symbols, what is the value of m, n, p, etc., in the following correlations so that each correlation in succession may be a maximum: $r_{1.2+m3}$; $r_{1.2+m3+n4}$; $r_{1.2+m3+n4+p5}$; etc., where m is the weight attached to the third variable; n, that to the fourth; p, that to the fifth, etc. This procedure may be called cumulative correlation as opposed to straight partial correlation. It must not be confused with multiple correlation. In order to avoid misunderstanding, it is well before proceeding further to make clear what meaning the subscripts carry. Here $r_{1.2+m3+n4}$ is the correlation between x_1 and $x_2 + mx_3 + nx_4$ where the value of n is derived on the assumption that $x_2 + mx_3$ is a single variable. It might, except for inconvenience, be written $r_{1.(2+m3)+n4}$. Similarly $\sigma_{2+m3+n4}$, which is used later is the equivalent of $\sigma_{(2+m3)+n4}$. The equation for five variables expressing the relation between x_1, x_2, x_3, x_4, x_5, is:

$$cx_1 = x_2 + mx_3 + nx_4 + px_5,$$

where c = constant defined later.

The formulæ for such operations are well known and comparatively simple. They are deduced from first principles below, though certain steps in algebra and calculus have been omitted. Let x_1, x_2, x_3, x_4, x_5, represent the variables, expressed as deviations from the means; the problem is what weight m must be attached to x_3 to make the correlation between x_1 and $x_{2+}mx_3$ a maximum. Writing this correlation as $r_{1.2+m3}$ we have

$$r_{1.2+m3} = \frac{\Sigma x_1(x_2 + mx_3)}{(\Sigma x_1^2)^{\frac{1}{2}}(\Sigma(x_2 + m_3)^2x)^{\frac{1}{2}}}\,.$$

$$= \frac{r_{12}\sigma_2 + r_{13}\sigma_3 m}{(\sigma_2^2 + 2r_{23}\sigma_2\sigma_3 m + \sigma_3^2m^2)^{\frac{1}{2}}}\,. \qquad (x)$$

By differentiating with regard to m, and equating to zero, it can readily be shown that the value of m which makes $r_{1.2+m3}$ a maximum is given by the equation

$$m = \left(\frac{\sigma_2}{\sigma_3}\right)\frac{r_{13} - r_{12}r_{23}}{r_{12} - r_{13}r_{23}}. \qquad (a)$$

As might be expected, post facto, $m = b_{13.2}/b_{12.3}$ where, as in the usual Yule notation,

$$x_1 = b_{12.3}x_2 + b_{13.2}x_3.$$

Equation x may for convenience be written

$$r_{1.2+m3} = \frac{r_{12}\sigma_2 + r_{13}\sigma_3 m}{\sigma_{2+m3}}, \qquad (b)$$

where σ_{2+m3}, the denominator of equation x, is given by the formula

$$\sigma_{2+m3} = (\sigma_2^2 + 2r_{23}\sigma_2\sigma_3 m + \sigma_3^2m^2)^{\frac{1}{2}}. \qquad (c)$$

We may now consider a and b as the basic formulæ; from them the value of m, the most favorable weighting, and the value of the new cumulative correlation is obtained.

As the next step it will be necessary to make the following substitution

$$r_{4.2+m3} = \frac{r_{42}\sigma_2 + r_{43}\sigma_3 m}{\sigma_{2+m3}}. \qquad (d)$$

To add a further variable x_4, we consider the sum $x_2 + mx_3$ as a single variable, and by employing the basic formulæ a and b, proceed at once to write down the value of the weight n, to be attached to x_4 and also the value of the new cumulative correlation.

$$n = \left(\frac{\sigma_{2+m3}}{\sigma_4}\right)\frac{r_{14} - r_{1.2+m3}r_{4.2+m3}}{r_{1.2+m3} - r_{14}r_{4.2+m3}}. \qquad (e)$$

With this value of n determined

$$r_{1.2+m3+n4} = \frac{r_{1.2+m3}\sigma_{2+m3} + r_{14}\sigma_4 n}{\sigma_{2+m3+n4}}, \qquad (f)$$

where

$$\sigma_{2+m3+n4} = (\sigma_{2+m3}^2 + 2r_{4.2+m3}\sigma_{2+m3}\sigma_4 n + n^2\sigma_4^2)^{\frac{1}{2}}. \qquad (g)$$

For the next step the two following substitutions must be made:

$$r_{5.2+m3} = \frac{r_{52}\sigma_2 + r_{53}\sigma_3 m}{\sigma_{2+m3}}, \qquad (h)$$

$$r_{5.2+m3+n4} = \frac{r_{5.2+m3}\sigma_{2+m3} + r_{54}\sigma_4 n}{\sigma_{2+m3+n4}}. \qquad (i)$$

Proceeding to introduce x_5 with its most favorable weighting p, we have

$$p = \left(\frac{\sigma_{2+m3+n4}}{\sigma_5}\right)\frac{r_{15} - r_{1.2+m3+n4}r_{5.2+m3+n4}}{r_{1.2+m3+n4} - r_{15}r_{5.2+m3+n4}}. \qquad (j)$$

Whence, knowing p,

$$r_{1.2+m3+n4+p5} = \frac{r_{1.2+m3+n4}\sigma_{2+m3+n4} + r_{15}\sigma_5 p}{\sigma_{2+m3+n4+p5}}, \qquad (k)$$

where

$$\sigma_{2+m3+n4+p5} = (\sigma_{2+m3+n4}^2 + 2r_{5.2+m3+n4}\sigma_{2+m3+n4}\sigma_5 p + p^2\sigma_5^2)^{\frac{1}{2}}. \qquad (l)$$

This process of combining each added variable with all the previous variables weighted as shown, and substituting in equations a and b may be carried on indefinitely. At each stage, a new weighting is derived, and the increase in correlation shows the value derived from the addition.

It is perhaps advisable to point out specifically what are the differences between the weightings as obtained by this method [1] and the ordinary regression coefficients. It has already been mentioned that $m = b_{13.2}/b_{12.3}$, so that m is so chosen that x_1 and $x_2 + mx_3$ yield a maximum correlation. No other factor entered into the determination of m; its value is absolutely *independent* of the variables x_4, x_5, x_6, etc. The value of the corresponding partial regression coefficient is *dependent* on all the other variables x_4, x_5, x_6, etc. For the accurate investigation of independent causation the partial regression coefficients are the correct instruments; but for the investigation of the effect of adding a single test to a battery already weighted as though complete in itself, the cumulative correlation process has a distinct field of usefulness because it provides a weighting for any variable introduced in a definite series which is independent of the inclusion or exclusion of any of the later variables in the series. More important is the fact that it may also be employed, with a small amount of error, for the practical purpose of determining the weights of a given series of tests. The values of the weightings, so obtained, will not differ very considerably from those found by the ordinary partial procedure.

For purposes of illustration, four problems[2] have been selected for which the usual partial regression coefficients and $R_{1.(234-n)}$ were available. For each of these problems the partly analogous values found by the cumulative method have been determined by continued substitution in equations

[1] I understand from correspondence with Dr. H. A. Toops that he is using a somewhat similar procedure. Not having had an opportunity to examine the equations he employs, I cannot tell the extent to which his assumptions are identical with those made in this article.

[2] Problem 1 is taken from Yule's 'Theory of Statistics,' p. 242. Fourth Edition. Problems 2 and 3 were kindly supplied me by Dr. J. E. Anderson who had already determined the partial regression coefficients. Problem 4 is taken from Kelley's 'Educational Guidance,' p. 94.

a to *g* for the four-variable problem, and by using equations *a* to *l* for the five-variable problem. The variables are included in the order shown by the subscript; as a general rule, this follows the magnitudes of the zero order correlation coefficients with the variable x_1. As will be readily seen, for each order of the variables chosen, there will be corresponding alterations in the values of the resulting weightings m, n, p, etc. The best plan in ordinary work is to introduce the independent variables in the order of the magnitudes of their zero order correlations with the dependent variable.

Problem I.

$$\begin{array}{lll} r_{12} = .52, & r_{23} = .49, & \sigma_2 = 41.7, \\ r_{13} = .41, & r_{24} = .23, & \sigma_3 = 5.5, \\ r_{14} = -.14, & r_{34} = .25, & \sigma_4 = 23.8. \end{array}$$

By ordinary partial correlation procedure

$$c'x_1 = 1 \cdot 0x_2 + 4 \cdot 3x_3 - 1 \cdot 2x_4,$$

where $c' = (b_{12.34})^{-1}$ (Yule notation).

$$R_{1.234} = .63.$$

By cumulative correlation procedure

$$cx_1 = 1 \cdot 0x_2 + 3 \cdot 7x_3 - 1 \cdot 1x_4,$$
$$r_{1.2+m3+n4} = .62.$$

Problem II.

$$\begin{array}{llll} r_{12} = .33, & r_{23} = .46, & r_{35} = .46, & \sigma_2 = 6.6, \\ r_{13} = .29, & r_{24} = .39, & r_{45} = .47, & \sigma_3 = 5.7, \\ r_{14} = .25, & r_{25} = .49, & & \sigma_4 = 5.8, \\ r_{15} = .23, & r_{34} = .29, & & \sigma_5 = 7.3. \end{array}$$

By ordinary partial correlation procedure

$$c'x_1 = 1.00x_2 + .86x_3 + .72x_4 + .00x_5,$$
$$R_{1.2345} = .39.$$

By cumulative correlation procedure

$$cx_1 = 1.00x_2 + .81x_3 + .64x_4 + .00x_5,$$
$$r_{1.2+m3+n4+p5} = .38.$$

Problem III.

$$\begin{array}{llll} r_{12} = .33, & r_{23} = .46, & r_{35} = .29, & \sigma_2 = 6.6, \\ r_{13} = .29, & r_{24} = .25, & r_{45} = .20, & \sigma_3 = 5.7, \\ r_{14} = .27, & r_{25} = .39, & & \sigma_4 = 5.4, \\ r_{15} = .25, & r_{34} = .31, & & \sigma_5 = 5.8. \end{array}$$

By ordinary partial correlation procedure

$$c'x_1 = 1.00x_2 + .70x_3 + 1.05x_4 + .63x_5,$$
$$R_{1.2345} = .41.$$

By cumulative correlation procedure

$$cx_1 = 1.00x_2 + .81x_3 + .98x_4 + .65x_5,$$
$$r_{1.2+m3+n4+p5} = .41.$$

Problem IV.
$$\left.\begin{array}{llll} r_{12} = .72, & r_{23}, = .61, & r_{35} = .55, & \sigma_2 = 5.19 \\ r_{13} = .62, & r_{24}, = .61, & r_{45} = .59, & \sigma_3 = 5.17 \\ r_{14} = .58, & r_{25}, = .82, & & \sigma_4 = 5.14 \\ r_{15} = .63, & r_{34}, = .66, & & \sigma_5 = 5.19 \end{array}\right\} \text{assumed equal.}$$

By ordinary partial correlation procedure

$$c'x_1 = 1.00x_2 + .50x_3 + .24x_4 + .12x_5,$$
$$R_{1.2345} = .76.$$

By cumulative correlation procedure

$$cx_1 = 1.00x_2 + .53x_3 + .24x_4 + .13x_5,$$
$$r_{1.2+m3+n4+p5} = .76.$$

If it is desired to pass from the relative weightings of the variables to prediction values and the absolute weightings in this procedure, it is necessary to determine the value of the constant c, for each problem. From the fact that $x_2 + mx_3 + nx_4 + px_5$ is considered as a single variable, it follows that in case of five variables

$$\frac{1}{c} = \frac{r_{1.2+m3+n4+p5}\sigma_1}{\sigma_{2+m3+n4+p5}}.$$

Finding the value of c for each problem and substituting we get for the cumulative equations

Problem I. $x_1 = .33x_2 + 1.23x_3 - .38x_4.$

II. $x_1 = .33x_2 + .26x_3 + .21x_4 + .00x_5.$

III. $x_1 = .29x_2 + .23x_3 + .28x_4 + .19x_5.$

IV. $x_1 = .37x_2 + .20x_3 + .088x_4 + .047x_5.$

These may be compared with the corresponding partial equations

Problem *I.* $x_1 = .33x_2 + 1.38x_3 - .38x_4.$
II. $x_1 = .32x_2 + .27x_3 + .23x_4 + .00x_5.$
III. $x_1 = .30x_2 + .21x_3 + .31x_4 + .19x_5.$
IV. $x_1 = .36x_2 + .18x_3 + .086x_4 + .047x_4.$

It is also interesting to trace the rise in value of the cumulative correlation coefficient with the addition of each successive variable. This is shown below:

PROBLEM

Cumulative Correlation Coefficient	I	II	III	IV
r_{12}	.52	.33	.33	.72
$r_{1.2+m3}$	.55	.37	.37	.75
$r_{1.2+m3+n4}$	.62	.38	.40	.76
$r_{1.2+m3+n4+p5}$	—	.38	.41	.76

From the statements that have been made and the results that have been shown, it can be seen that for the practical purpose of weighting the variables in accordance with their contribution to the total effect, there is no great discrepancy, up to five variables, between the ordinary partial correlation procedure and this much simpler cumulative method employing equations, the logic of which is fairly apparent, and which are relatively few in number compared with those necessary in the usual partial correlation work.

LEARNING WHEN FREQUENCY AND RECENCY FACTORS ARE NEGATIVE

BY JOSEPH PETERSON

George Peabody College for Teachers

Introduction

Of the more specific factors operating within the larger situation that compels learning, frequency effects seem to be principally, if not entirely, responsible for the backward elimination of errors in the maze. Such error elimination, and such operation of frequency factors, have been most clearly brought out in mental maze experiments, which eliminate the influence of various disturbing and irregular spatial factors. It has been shown that in mental-maze learning the probability of one's getting successfully past a blind alley without subsequent returns is greater near the goal-end than at the entrance-end, and that the learning coefficient constantly increases from the entrance to the goal.[1] In a ten-blind maze the learning coefficient at the first blind was found to be 1.031, whereas it was 1.333 at the tenth, for the first trial. This advantage at the latter part of the maze for the elimination of entrance to blinds seems to be based on frequency factors, and is traceable to the fact that in each trial the forward runs will necessarily exceed the backward runs by unity. This may be indicated by the equation

$$f - 1 = b,$$

in which unity is the only constant term. The absolute values of f and b constantly decrease toward the distal part of the maze, and therefore the ratio of f to b increases, being greatest at the last blind. This is what gives the frequency

[1] Peterson, Joseph, 'The Backward Elimination of Errors in Mental Maze Learning,' Jour. Exper. Psychol., 1920, *3*, 257–280. The 'learning coefficient' of a blind was defined in this report as 'the ratio of the expected number of runs past that blind in the forward direction *to* the expected number of runs into it (from either direction).' These expectations are based on the laws of probability.

advantage to the forward runs past blinds, over the backward runs and runs into blinds, an advantage that increases regularly toward the distal end of the maze.

If frequency factors could be exactly balanced against themselves so far as positive and negative effects are concerned, and if the same thing could be done with recency factors, we should have a rather unique situation for the study of learning, and could try out experimentally some of the current views regarding the nature of the learning process, at any rate in a particular if not in a crucial case. To attempt to do this was the object of the present investigation.

It will be necessary on the start to distinguish between frequency of stimulus and frequency of response. The one sort of frequency effects may be balanced without a balance of the others being effected. In the present study we have attempted to balance exactly the frequency-of-stimulus factors making, respectively, for and against learning at any part of the maze, and we have also attempted to do this for the recency factors. At the same time, when the effects of frequency and of recency of *response* are considered, we find that both in the case of frequency and in that of recency the odds are greatly against learning; that is to say, these factors are negative in their effects, tending to fix the erroneous responses rather than the correct ones. In a later study we hope exactly to equate both the stimulus and the response frequency and recency factors.

It may seem confusing to distinguish between the operation of frequency and of recency factors in the stimulus, on the one hand, and in the response on the other, but the distinction seems necessary if we clear our study of certain confusions which may pass unnoticed in mere theoretical discussions but which are met when exact experimental control is attempted. In the former case, frequency of stimulus, the stimuli making for changes in behavior, are probably entirely extero-ceptive, while they are both proprio-ceptive and extero-ceptive in the latter. However, an analysis on this basis is not yet practicable. When we speak of response in the present report, we must consider a choice as a unitary response, and must not for the present attempt to analyze the responses further. We are attempting to get a quantitative hold on our problem, and to do this it is necessary that we thus define our terms. This meaning of response is, however, not far different from that currently used in discussions of learning factors, except that our units are more nearly equal, being always simply a choice of one of a presented pair of letters. We do not deny that behind different choices lie different backgrounds, but these differences are nevertheless produced by the choices themselves which we have attempted to make equal.

Method of Procedure

The method of the present investigation is identical, so far as the experimental part is concerned, with that used in our earlier study cited above, with the important exception that on each wrong choice the subject is immediately brought back to the entrance and again given the alternative choices at the first blind, without any indication being made to him that anything irregular has happened, and without anything occurring to call his attention to the fact that a return has been made. Of the various maze situations that have been presented either to animals or to man, such a procedure is of course possible only with the mental maze. The instructions to the subject are these:

> A maze, you know, is a winding way to some goal, but it has many blind alleys which will lead you to error if you enter them. I have drawn a maze which I will show you when you have completed this experiment. Wherever there is a choice of two alleys, each pathway is designated by a letter. No two parts are lettered the same. I call out two letters at a time, and you are to choose one of them; then I call out two more; and so on, till you get to the goal. Whether the right letter comes first or second in the pair called out, is determined wholly by a chance-order schedule, arranged by flipping a coin; so you need not attend to which letter comes first. In fact, if you try to choose according to some pre-determined plan or order, or try to make the letters you choose spell words, you will never learn the way to the goal. The problem is to see with how few errors you can get to the goal, and how soon you can learn to make no errors at all. You are through the experiment when you get to the goal the third successive time without any error. You will be told each time when the goal is reached, and also the number of errors made in reaching it. Remember, accent and order of calling out the letters have no significance. Don't attend to them. Keep in mind also that where we go in the maze depends on your own choices.

Note.—Please do not inform any one else about the nature of this experiment. All subjects must begin ignorant of it.

Let us illustrate the process of carrying out the experiment by following an actual record. The experimenter having Figure 1 before him and sitting behind a screen so that unconscious facial expressions and movements will not influence the subject, hands the instruction sheet to the latter, asking him to read it carefully. While the subject is reading the instructions the experimenter records his name, the date, and any other conditions of the experiment to be noted; and just when the subject returns the sheet the experimenter also records the exact time to the nearest minute, and immediately calls out *N-V*. The subject chooses *N*, which the experimenter records and underlines, and he also puts a cross over it because the choice is an error. The subject, of course, does not see the record. *V-N* are then called out in the same tone of voice and accent as before, the right alternative now being called first as indicated in the chance-order schedule. This time *V* is chosen. It is quickly recorded and underlined, and *I-L* are presented. *I* is chosen, and is re-

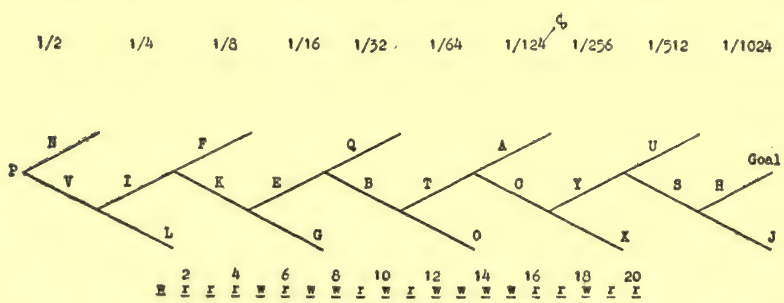


FIG. 1. The mental maze and the chance-order schedule. The fractions above the maze indicate the chance of passing in a single forward run the corresponding right alternatives, or of entering the wrong alternatives.

corded and underlined, after which *K-F* are called out. *F* is now chosen and is recorded by the experimenter with a cross over it. *N-V* are called, *N* coming first because *w* is the fifth letter in the schedule, which the experimenter has been following with his left index finger, and this is the fifth choice. Each *w* in the schedule means to call out the wrong letter first, while each *r* indicates that for the stimulus it represents, the right alternative must be presented first in the pair. *V* is now chosen by the subject and is recorded without any qualifying mark whatever, because it is neither the first of the alternatives called nor a wrong choice. Thus the process continues, probably for nearly an hour, when the subject finally reaches the goal. The experimenter now says "Goal!" and after recording the time, quickly counts and announces the number of errors. He then tells the subject that the second trial begins, and noting that the second letter in the schedule is an *r* (for right first), he calls out *V-N*. Each trial thus begins as indicated by the letter of its number in the chance-order schedule so that the order of presentation in successive trials is never the same, even if the subject makes no errors, or if he makes the same errors in two different trials. It will be noted that in each case in which the subject chooses the first of the two alternatives presented, the choice is recorded and underlined, and, in addition, erroneous choices are marked with a cross so that they can readily be counted up at the close of the trial. The subject when he reaches the goal is told the number of errors made but he must himself find out which of his many choices are wrong.

When the experiment is completed the experimenter has a full record of all the stimuli presented to and the responses made by the subject, even to showing at each alternative whether the first or the second of the letters presented at any time was selected. He can therefore study in detail any subjective tendency noted in the choices, whether it be to choose the first or the second of the alternatives in each pair, or to prefer now the one and now the other in any larger regular or irregular cycle. A thorough study of the subject's choices may in any case be helpful in throwing light on the subject's learning and on his peculiar difficulties. On the other hand, the subject himself has had a real problem to solve, one on which he has a good deal of room to exercise his own constructive ability and to try out various hypotheses, reactions which are often obvious enough to the experimenter.

This learning process is therefore not a matter of merely remembering a series of syllables, words, or objects in a fixed order, determined in advance by the conditions of the experiment, as has usually been the case in association experiments. And it is important to note that in the present form of learning, the subject, in the solution of his problem, by whatever methods he may adopt, leaves a complete, permanent record that can be analyzed in detail, even to determining the laws of association influential in each choice. From such a record, then, we are able to check up on some of the theories which flourish so well in the absence of objective facts.

A glance at the maze, Fig. 1, will show that the chances are even, that is, 1 to 1, that the subject will make the right choice at the first blind; that they are 1 to 4 that he will also successfully pass the second blind; 1 to 8 that he will get past the first three blinds; and so on. The chances of getting past all the blinds without being thrown back is, of course, the product of all the chances of passing the several blinds by themselves, or $\frac{1}{2}$ raised to the power of the total number of blinds. In the present maze the chances of reaching the goal without any error whatever is $(\frac{1}{2})^{10}$, or 1 to 1024. None of my subjects has yet succeeded in doing this, although some of them have had very good luck. Now if any one of the blinds is chosen and an immediate return to the beginning is made by the experimenter's calling out the *N-V* alternatives again, it is obvious that the mere frequency and recency of the wrong choice will operate toward the continuation of the error at this juncture, that is, *against* the learning of the maze. Indeed, the chances are that frequency and recency response effects will thus in 1023 times out of 1024 count negatively for the learning, while their stimulus effects are equally balanced for and against learning. That the effects of frequency and recency factors in response are very real is obvious to anyone who examines in detail the individual records as tabulated for intensive study. We have here then a case of constructive learning in which frequency and recency influences are decidedly against the formation of the proper associations for the solution of the problem.

Subjects

Twenty college students and graduate students were used as subjects in the present experiment with the modified mental maze, five men and fifteen women. None of them had had previous experience with either form of the mental maze. No introspections were called for during the experiment, as that would introduce variable factors, but some of the subjects made significant remarks, either spontaneously during the experiment or in reply to questions after it was completed. These remarks became part of the record, as did also various notes on the behavior of subjects during the experiment. Four of the subjects, all women, were unable to learn the maze in the time available, so their records are not included in this report. Three of these subjects worked 72, 98, and 110 minutes respectively, without any promise of success, while the fourth one, a woman who was having difficulty in carrying some of her courses, persisted in the invariable repetition of her erroneous choice of *N*, declaring that she was 'all in the dark anyway, and might as well go one way as another.' Our results are therefore based on the records of sixteen subjects, all of whom completed the problem. In every case the learning was completed at a single sitting.

Results

Table I. shows the general distribution of errors made by sixteen subjects to the point of complete learning of the maze. All but seven of the subjects made no errors after the first trial, but this trial was sometimes very long. The errors made in all trials after the first, by seven subjects, are shown separately in the table; and the total number of entrances to the goal, or goal choices, is also indicated, because these choices must be taken into consideration in the determination of the expected choices in the several parts of the maze. The table shows the distribution of errors of each subject separately as well as the total number of errors at each bifurcation by all the subjects, first for the first trial only, and then for all subsequent trials by seven subjects. It also gives the number of errors (based on the total number of errors made *plus* the goal choices) that would be expected at each part of the maze on the law of probability. On the basis of probability alone we should expect as many errors in the first blind

TABLE I

DISTRIBUTION OF ERRORS IN THE SEVERAL PARTS OF THE MAZE BY SIXTEEN SUBJECTS

Subject	Parts of Maze										Total Errors	Goal Choices	Total
	H	*L*	*F*	*G*	*Q*	*O*	*A*	*X*	*U*	*J*			
First trial													
A......	10	4	3	2	1		1	1	2	1	25	1	26
B......	85	49	11	7	3	3	1				159	1	160
C......	31	6	2	4							43	1	44
D......	41	28	15	1	36	1	1	1		1	125	1	126
E......	61	19	15	3	6		4	11		1	120	1	121
F......	112	21	17	8	4	3	2	1			168	1	169
G......	88	55	9	7	5	1		1			166	1	167
H.....	130	41	36	21	29	9	2	1	1	1	271	1	272
I......	2	1	2	4	2	2	1	1	1		16	1	17
J......	24	52	16	5	23	3	2		1		126	1	127
K.....	142	42	22	6	7	11	1	1	1		233	1	234
L......	12	10	5	9	3		1	2	2		44	1	45
M.....	6	1	2	1	1	1		1	1	1	15	1	16
N.....	22	16	5	6	3	5	4	1	1	1	64	1	65
O......	13	9	5	4	2	1	1	1	1		37	1	38
P......	141	69	17	11	4	3	4	1	1	2	253	1	254
Total....	920	423	182	99	129	43	25	24	12	8	1,865	16	1,881
Expected.	940.5	470.3	235.1	117.6	58.8	29.4	14.7	7.3	3.7	1.8	1,879.2	1.8	1,881
All subsequent trials													
A.....									1	1	2	1	3
B......			1	1	3		3	3			11	1	12
C......	3	32	30	6	1	4	2	2			80	5 (2–6th)	85
E......									1		1	1	2
G......										1	1	1	2
I......				1							1	1	2
L......	1										1	1	2
Total, all.	924	455	213	107	133	47	30	29	14	10	1,962	27	1,989
Expected.	994.5	497.3	248.6	124.3	62.2	31.1	15.5	7.8	3.9	1.9	1,987.1	1.9	1,989

as in all the following blinds *plus* the number of goal choices. This is because the chances are just even for choosing the first blind or going past it, and all the passes must end in some later blind or in the goal. The same thing holds for each subsequent blind; the choices made at the juncture where it appears should be divided evenly between it and all blinds beyond it, with the goal choices included in the latter. It will be seen that of the first-trial choices there were 920 errors made at the first blind, while 940.5 were to be expected; at the second blind there were 423 errors and 470.3 expected; and so on, a constantly decreasing series both in actual and in expected errors toward the distal end of the maze. Table II. shows these same totals in per cent., the expected errors, given in bold type, being the same in the first as in the subsequent trials or in all trials combined.

TABLE II

PERCENTAGE DISTRIBUTION OF ERRORS IN THE MENTAL MAZE (SIXTEEN SUBJECTS)

	Parts of Maze										Total Errors	Goal Choices	Total
	N	*L*	*F*	*G*	*Q*	*O*	*A*	*X*	*U*	*J*			
First trial, Obt'd..	48.9	22.5	9.7	5.2	6.8	2.3	1.3	1.3	.6	.4	99.0	.9	99.9
All trials, Obt'd ..	46.5	22.9	10.7	5.4	6.7	2.4	1.5	1.5	.7	.6	98.9	1.4	100.3
Expected........	**50.0**	**25.0**	**12.5**	**6.3**	**3.1**	**1.6**	**.8**	**.4**	**.2**	**.1**	**100.0**	**.1**	**100.1**

In general the correspondence of actual to expected errors in all parts of the maze is rather close, both in the first trials only and in all trials together of all the sixteen subjects who completed the maze learning. This shows that probability laws are indeed operative to a large extent even though the choices are voluntarily made by the subjects. But in both the first trials and in all trials combined we find a larger percentage of errors toward the distal end of the maze than expected on the laws of chance. This tendency of increased errors beyond expectations toward the goal is, of course, the effect of learning, whatever has brought about the learning, an effect that is usually very obvious to the experimenter during the first trial. Many subjects, in fact, make no errors after once reaching the goal, having learned to avoid every blind before this place is reached.

A very striking effect of the modification in this experiment effecting complete returns after erroneous choices is that which shows itself in the order of elimination of errors. *In* *the present form of the mental maze errors are eliminated in the forward direction*, whereas in the previous form not involving a return with each error, as well as in the mazes used for animals, the opposite tendency appears, as we have already seen. It was shown that backward elimination was mathematically attributable to the operation of frequency factors, which, as we mentioned above, are in the present case actually negative. In both cases, however, frequency and recency factors are not the real determining conditions of the learning process. This is the important result of our studies that will receive more explicit attention in a later part of this report.

The forward elimination of errors in the present form of the mental maze is revealed in a study of Tables III. and IV.

TABLE III

DISTRIBUTION OF THE FINAL ERRORS OF THE SEVERAL SIXTEEN SUBJECTS IN THE MODIFIED MENTAL MAZE

	Parts of the Maze										Total
	N	*L*	*F*	*G*	*Q*	*O*	*A*	*X*	*U*	*J*	
Number of final errors.	2	0	1	1	0	1	1	2	4	4	16

Table III. shows the number of *last errors* by our sixteen subjects in each blind of the maze. All but five of these subjects made their last errors at or beyond the seventh blind, and only two made their final errors in the first blind, one of these stating after the experiment that his error at this place was due to a choice purposely made to verify an idea that he had respecting it. Half of the sixteen subjects made their final errors in the last two blinds.

But the tendency to eliminate errors in the forward direction in the modified mental maze is even more obvious when we follow an individual in detail in the process of the learning. In Table IV. the progressive elimination forwards is shown by merely numbering all the errors of any one subject in the order of their occurrence, and by putting these several numbers in the columns corresponding to the blinds in which they occurred. The table shows separately the distribution and the order of occurrence of the errors made by three subjects, *E*, *G*, and *M*. Space does not permit the giving of the responses of some of the longest records, but those given are

Table IV

The Distribution and Order of All the Errors Made by Certain Subjects in the Modified Mental Maze

Parts of Maze										
N	*L*	*F*	*G*	*Q*	*O*	*A*	*X*	*U*	*J*	
Subject *E*										
1–9	10–13									
14–21	22									
23–26	27									
28–33	34									
35–40	44, 46	41, 42		43, 45						
48–50		51		47						
52–54	55									
56–59				60						
64	61–63									
68, 69	65, 67		66							
72, 73	70, 76	74, 75		71						
77–79	80, 81									
82, 84, 85 ...				83		86				
87, 88		90, 91	89							
92		94	93			95–97	99		98	
107, 108		100, 101					102–106			
		109, 110					111			
		112					113			
114, 115		116, 117					118, 120			Goal
Second trial .								121		Goal
Subject *G*										
2–58, 64	1, 59, 60	62	61	63						
66, 73–74	65, 70–72	67–69								
76, 78, 82	75, 77, 80						79			
84–86, 92	81, 83, 89	87, 88	90							
93, 99, 101 ..	91, 95, 96	94, 97		98, 100						
104, 106	102, 103									
110, 111	107–109									
113, 115	112, 114									
124	116–123									
143, 146	125–142	144								
148, 150	145, 147									
152, 153	149, 151		154, 156							
155, 158	157		159							
160, 161			163, 165	162, 164	166					Goal
Subject *M*										
2, 3, 5, 6	1	4	7							
8, 9		12		10	11		13	14	15	Goal

fairly representative, the first two of the medium or poor students, and the last, the responses of subject *M*, of the best subjects. The table is to be read as follows: Subject *E* chose *N* of the first alternative nine times in succession; the next four times he passed the first blind successfully, that is, chose *V*, and in every case entered the second blind. Then again he entered the first blind successively eight times. The twenty-second error was made at the second blind after having successfully passed the first. Errors 23 to 26, inclusive, are again in the first blind (at the first alternative), number 34 is in the second, 35 to 40 in the first, *N*, 41 and 42 are in the third, or the *F* blind, the first two having been correctly passed; and so on. Note that for each error in any blind beyond the first one, the subject must successfully choose the right alternative at all previous junctures or critical points; for example, error number 85 was made in the first blind, then the next choices at the first six blinds were all correct, and the 86th error was made by entering the seventh blind, *A*. Both subject *E* and subject *G* learned very slowly to avoid the early blinds, and they were never quite sure of passing them successfully till near the end of the experiment. On the contrary, subject *M* readily learned to avoid choices that were unfruitful, the result being a more regular forward progression of errors, with the last error occurring in the final blind, *J*.

In general, the method of learning may be characterized as follows: One usually begins by making a great number of errors in two or three of the first blinds, according to what would be expected on the laws of probability. These errors are repeated, sometimes a surprising number of times, especially by the passive, rather conventional type of person. Variations in response then occur, often without any purpose or forethought, and the way becomes cleared to the blinds beyond, which in turn give trouble just as the first ones did. In correcting the errors at these alternatives the subject often loses the effect of his earlier responses and again falls into error at the blinds first encountered. These are more readily eliminated now, however, and the subject gradually pushes his conquest over errors further and further back toward the distal end of the maze. In proportion as the subject keeps in mind the situation, or retains the effects of his earlier experiences, the series of repetitive responses each terminating in an error become more and more far-reaching until finally the goal is reached. In later trials few errors, if any, are made. That is, the subject as a rule clears his way as he goes, gradually enlarging his range of orientation, until the whole situation is either consciously reacted to as a unit, or at any rate is made effective somehow in the determination of his reactions. Some subjects seem rather suddenly to come to a recognition of the means of locating their errors and hence eliminate them, and it is usually obvious to the experimenter just when and at what point this recognition comes. It is really surprising, though, how slowly this idea, that erroneous choices throw one back to the *N-V* alternative, dawns on some of the subjects. In fact, the poorer subjects have constantly to recur to the first errors—the choice of *N*, *L*, or *F*—and they rather gradually and unconsciously wear themselves out, as it were, of these erroneous responses. Their responses differ from those of the better subjects in the fact that in the enlargement of the range of reaction already described, they frequently lose what they have gained and lapse back into earlier erroneous responses terminating the shorter ranged reactions that are characteristic of the earlier stage of the learning. But when once the principle, that errors throw one back, is clearly perceived, these relapses do not occur. The difference between the good and the poorer subjects is principally one of the explicitness with which the situation and the principle of error elimination are grasped. But from our experiments with the mental maze, as well as from those described as rational learning,[1] it seems that one may come to follow a useful principle in learning even though one is not aware of the fact or of the principle. Presumably many persons in practical life are unconsciously forced by the general consistency of circumstances thus to follow principles, and it would be a great error to ascribe the uniformity of such behavior to imitation and suggestion as is commonly done by certain writers on social psychology.

[1] Peterson, Joseph, 'Experiments in Rational Learning,' *Psychol. Rev.*, *25*, 1918.

There seems, then, in the case of all subjects, good or poor, to be some gradually accumulative effect of the successive repetition of errors in a monotonous situation, an overlapping of stimulus effects of some sort that amounts to an increase in the range of receptivity to the stimuli of a complex situation, some effect the neural and general physiological basis of which we are yet very ignorant, but which reveals itself very well in the reactions and the verbal statements of the subjects of the present experiment. For instance, one bright, mentally alert boy in Peabody Demonstration School, on whom the experiment was tried, soon noted the return to the *N-V* situation in connection with certain of his choices, and he laughed heartily each time when he was thrown back. Using that return as a criterion of the degree of success of his choices and explorations in the maze, he soon solved the problem. Several of our subjects noted unpleasant affective states associated with the recurrence of the first often repeated letters, *N-V*, *I-L*, etc. One adult woman, who had been thrown back a great number of times, finally showed signs of attention to the recurrence of the *N-V* situation, and, after the completion of the problem, said that the letters at these early blinds got to be very repugnant to her. "I just got to hate these letters," was her way of expressing it. Some subjects after long repetitions without any sign of recognition of the fact that errors threw them back, would in time perceptibly sigh when the *N-V* situation was presented, following an error; and yet after such evidence of discomfort they frequently continued to make the error, ignorant of the principle which might help them reach the goal. It is thus clear that something more than cerebral activities became invloved in connection with these repeated failures, physiological processes which are undoubtedly in large part visceral. The means by which these accumulative organic effects arise and inhibit the erroneous responses in learning is unknown.

We have already indicated that a number of our subjects perceived clearly, after a few experiences of being thrown back, the significance of these set-backs; and that they checked their errors explicitly and voluntarily by them. Such subjects made much more rapid progress in the elimination of errors at successively encountered blinds, and they seldom made any lapses into errors once overcome, as did subjects who could not at the end of the experiment explain how to overcome errors. But we must not overlook the fact here that this more explicit perception is itself only the conscious aspect of an intra-organic process now possibly more completely cerebral in its location than is true when strong affective disturbances, not accompanied by clear perception of the meaning of the situation, arise. To deny 'stamping in' effects of right responses or 'stamping out' effects of wrong responses to feeling states *per se* and to ascribe such effects to perception or to the recognition of meanings, is of course inconsistent. In both cases there are physiological bases of which we are yet ignorant, the ignorance being in part due to the fact that we have been content to stop with the mental states themselves in our tracing of causal relations. This error is not unlike that of attributing certain acts to will, to attention, or to any other of the spontaneous faculties still in good repute in much of our current psychological literature. The perception of the meaning of the set-backs is itself but a response, a kind of behavior revealed in the reactions of the subject, even though expressed verbally by the subject.

An Analysis of the Influence of Frequency and Recency Factors

Having considered the method of learning in the modified mental maze and the order of elimination of errors, we may now turn our attention to a consideration of the rôle of frequency and recency factors in the process. It will be recalled that during the learning by each subject a complete record was made of all the successive alternatives presented to him as well as one of his individual responses in the exact order of their occurrence. From this record we are, fortunately, able to tabulate each subject's responses in detail, both as to order and location in the maze (or in the series of stimulus situations) and as to their agreement or disagreement with expectations based on frequency and recency factors as determined by all past responses up to the point in question. The first step in this analysis is to tabulate in a form like that shown in Table V., all the responses of each subject. A separate line is allowed for each forward movement from the initial alternatives at *N-V*, and on each of these lines are placed pencil checks in the appropriate columns for the several choices, each check indicating a choice of the letter (or part of the maze) at the head of the column. Each forward movement continues, of course, until some error is made, or until the goal is reached. When all of the subject's responses to the point of no further errors are tabulated in this manner, the record is ready for an analysis of the frequency and recency laws of response. For this purpose we use the following abbreviations: *B* indicates a response that is in accordance with the expectations based on both recency and frequency, while *b* is for a response of the opposite nature, contrary to frequency and recency expectations; *Rf* represents a response in agreement with recency expectations and contrary to frequency expectations, and *Fr* is just the reverse of *Rf; R* is for a response in agreement with expectations on recency when frequency factors just balance, while *r* is just the reverse, contrary to recency; finally 1 designates the first response to any one of the alternatives at a blind, a response which cannot therefore be analyzed according to recency and frequency factors. With these symbols in mind we go over the individual record and write in ink over the erasable checks the appropriate symbol. When the ink dries the checks are erased, leaving only the symbols. We get then a record such as that shown in Table V. of subject *L*. A study of this record will make the process clear. *L* first made three erroneous choices of *N* at the first alternative. The first one is marked 1, and the other two, being in agreement with both recency and frequency expectations, are marked *B*. It is obvious that now if these laws determine the learning all subsequent responses should be choices of *N;* but the fourth response is a choice of *V*, contrary to both these laws, hence this response is marked *b*. The subject now having passed

Table V

An Analysis of All the Responses of Subject *L* to the Point of Complete Learning of the Mental Maze. *B* means that Response is in Agreement with Expectations based on Frequency and Recency; *b*, the Reverse, Contrary to these Laws; *Fr*, in Agreement with Expectations on Frequency, Contrary to Recency; *Rf*, the Reverse of *Fr*; *R*, in Accord with Recency when Frequency Factors Balance; *r*, Contrary to Recency; 1, first Response to a Pair of New Alternatives

Trial	Parts of the Maze																			
	N	*V*	*L*	*I*	*F*	*K*	*G*	*E*	*Q*	*B*	*O*	*T*	*A*	*C*	*X*	*Y*	*U*	*S*	*J*	*H*
(1) . . .	1																			
	B																			
	B																			
		b	1																	
		Rf	*B*																	
		Rf	*B*																	
		R	*B*																	
		B		*b*	1															
	b																			
		Fr		*Rf*	*B*															
	b																			
		Fr		*Rf*	*B*															
		B		*Rf*	*B*															
		B	*r*																	
		B	*B*																	
		B	*B*																	
		B	*B*																	
		B	*B*																	
	b																			
	Rf																			
	Rf																			
	Rf																			
		Fr		*b*		*b*		1	1											
	b																			
		Fr		*Rf*		*Rf*	*b*													
		B	*Fr*																	
	b																			
		Fr		*b*		*Rf*		*r*	*B*											
		B		*Rf*		*Rf*	*b*													
		B		*Rf*		*R*	*R*													
	b																			
		Fr		*Rf*		*B*	*B*													
		B		*R*		*B*	*B*													
		B		*B*		*B*	*B*													
		B		*B*		*B*	*B*													
		B		*B*		*B*	*B*													
		B		*B*		*B*		*b*	*B*											
		B		*B*		*B*		*Rf*	*B*											
		B		*B*		*B*		*Rf*		*b*		1		1						
		B		*B*		*B*		*Rf*		*Rf*		*B*		*b*		1		1		
		B		*B*		*B*		*Rf*		*Rf*		*B*		*R*		*B*		*B*		
		B		*B*		*B*		*Rf*		*Rf*		*B*		*B*	*b*					
		B		*B*		*B*		*R*		*R*		*B*		*B*		*Rf*				
		B		*B*		*B*		*B*		*B*		*B*		*B*		*r*		*b*		1
(2) . . .	*b*																			
		Fr		*B*		*B*		*B*		*B*		*B*		*B*		*B*		*Rf*		*B*

Summary of L's Reactions

Trial	Law	*N*	*V*	*L*	*I*	*F*	*K*	*G*	*E*	*Q*	*B*	*O*	*T*	*A*	*C*	*X*	*Y*	*U*	*S*	*J*	*H*	Total	Wrong	Right
1st.	*B*	2	22	7	11	3	13	5	1	3	1		5		3		1	1				78	21	57
	R		1		1		1	1	1		1				1							7	1	6
	Rf	3	2		7		3		5		3					1						24	4	20
	Fr		6	1																		7	1	6
	b	6	1		3		1	2	1		1				1	1			1			18	9	9
	r			1					1								1					3	1	2
2d..	*B*				1		1		1		1		1		1		1				1	8	0	8
	R																					0	0	0
	Rf																		1			1	0	1
	Fr		1																			1	0	1
	b	1																				1	1	0
	r																					0	0	0

the first alternative (the first blind) chooses *L* of the *L-I* alternative. This being the first choice here, is marked 1. Again *L* chooses *V*, which is in agreement with recency of response but contrary to frequency, and is so marked. So we go on through the entire record till the goal is reached. *L* made one error in the second trial, the choice of *N*, so all the second-trial reactions are also recorded. A summary of *L*'s reactions as here classified is given at the base of the table, but in this summary the uncritical reactions marked 1 are omitted. In this summary of the first trial we find, for example, a total of 78 *B*-responses, 21 of which were wrong and 57 of which were right choices; a total of 7 *R*-responses, with 1 wrong and 6 right. There are 24 *Rf*-responses with 4 wrong and 20 right; 7 *Fr*-responses, with 1 wrong and 6 right; 18 *b*-responses of which 9 are right and 9 wrong; and 3 *r*-responses, 1 wrong and 2 right.

In this way each of the sixteen records has been treated, and the several summaries have been condensed into Tables VI., VII., and VIII. The method is seen to be similar to that which we applied to the learning of white rats in the animal maze.[1] In Table VI. are summarized the various critical or significant responses of the sixteen subjects, sepa-

[1] Peterson, Joseph, 'Frequency and Recency Factors in Maze Learning by White Rats,' *Jour. Animal Behav.*, 1917, 7, 338–364.

Table VI

Classification of All the Significant Responses Made by the Sixteen Subjects in the Modified Mental Maze

	Kind of Response						
	B	*R*	*Rf*	*Fr*	*b*	*r*	Total
First Trial							
Number right	1,104	78	508	203	445	29	2,367
Number wrong	839	43	232	279	288	27	1,708
Total	1,943	121	740	482	733	56	4,075
2d to 6th trials							
Number right	213	10	67	22	27	7	346
Number wrong	20	6	14	17	36	4	97
Total	233	16	81	39	63	11	443
All trials (summary)							
Number right	1,317	88	575	225	472	36	2,713
Number wrong	859	49	246	296	324	31	1,805
Total	2,176	137	821	521	796	67	4,518
All trials in per cent.							
Per cent. right	**29.1**	**1.9**	**12.7**	**5.0**	**10.4**	**.8**	**60**
Per cent. wrong	**19.0**	**1.1**	**5.4**	**6.6**	**7.1**	**.7**	**40**
Total	**48.1**	**3.0**	**18.1**	**11.6**	**17.5**	**1.5**	**100**

Table VII

Classification of All the Significant Responses of the Sixteen Subjects in the Modified Mental Maze to the Point in Each Part of the Maze where the Learning Was Complete

	Kind of Response						
	B	*R*	*Rf*	*Fr*	*b*	*r*	Total
Number right	702	38	437	225	472	36	1,910
Number wrong	859	49	246	296	324	31	1,805
Total number	1,561	87	683	521	796	67	3,715
Per cent. right	**18.9**	**1.0**	**11.8**	**6.1**	**12.4**	**1.0**	**51.4**
Per cent. wrong	**23.1**	**1.3**	**6.6**	**8.0**	**8.7**	**.8**	**48.6**
Total per cent.	**42.0**	**2.3**	**18.4**	**14.1**	**21.1**	**1.8**	**100.0**

Table VIII

Percentage of Responses of Sixteen Subjects in Agreement with Expectations Based on Frequency and on Recency Factors, and also Contrary to these Expectations

Responses are classified according to whether they are right or wrong.

	Kind of Response						Total		Difference
	B	*R*	*Rf*	*Fr*	*b*	*r*			
Right responses									
With frequency	18.9			6.1			25.0		
Against frequency			11.8		12.4			24.2	
Wrong responses									
With frequency	23.1			8.0			31.1		
Against frequency			6.6		8.7			15.3	
Totals	42.0		18.4	14.1	21.1		56.1	39.5	16.6
Right responses									
With recency	18.9	1.0	11.8				31.7		
Against recency				6.1	12.4	1.0		19.5	
Wrong responses									
With recency	23.1	1.3	6.6				31.0		
Against recency				8.0	8.7	.8		17.5	
Totals	42.0	2.3	18.4	14.1	21.1	1.8	62.7	37.0	25.7

rated first on the basis of first and later trials, and then of all trials together up to the point of complete learning. The summary of all trials together is also given in percents, this part of the table being in bold face type. In this table, are summarized the classifications of 4518 responses, exclusive of insignificant responses. It is obvious that the analysis of the first trial responses is more important for the understanding of learning than is that of later trials, after the goal has been reached, because in later trials the effect of the learning has already become a considerable amount at different parts of the maze. In the first trial there were 1104 *B*-responses (in harmony with recency and frequency) in the right choices and 839 in the wrong ones; but there were also 955 correct choices *contrary* to frequency expectations (*i.e.*, 508 *Rf* + 445 *b*) and only 520 (or 232 *Rf* + 288 *b*) wrong responses contrary to frequency. Moreover, in the right responses there are

78 + 508, or 586, in accordance with recency, and 445 + 29, or 474, against recency expectations. Of the wrong responses there are 275 (or 43 *R* and 232 *Rf*) in accordance with recency expectations and 315 (or 288 *b* and 27 *r*) against such expectations. Surely nothing very decided for frequency and recency influences can be made out from these data. A detailed study of individual records in the form of Table V. shows, indeed, the influences of frequency and recency, but these influences are of such a nature as to fix the erroneous responses as much as they fix the correct ones. In fact, as we have already said, learning in the present form of the maze problem must go on *in spite of* the effects of both frequency and recency. But we have previously shown, in the last reference cited, that even in the usual type of maze learning by rats, learning effects must in many cases go directly against frequency and recency influences.

But the data in Table VI. are somewhat deceiving in as much as certain parts of the maze are learned earlier than other parts, and hence in these places of earlier learning the responses will continue to be one hundred per cent. in agreement with both frequency and recency expectations until the other parts are also learned. Thus in Table V. we note that column I. has 12 *B*, 1 *R* and 3 *Rf* correct responses, and column *K* has 14 *B*, 1 *R*, and 3 *Rf* correct responses beyond the last correct choice. These are cases of over-learning in which the responses must of necessity be one hundred per cent. in accordance with either frequency or recency, or both. It is possible to eliminate all such over-learning responses from our tables and find the influence of frequency and recency factors only up to the point of learning, say one response correct after which no errors occur in each part of the maze. This has, therefore, been done in the present study by omitting from our summaries all the responses occurring in each individual's record at places where learning is complete, counting only the first of the final series of correct responses at each place. For example, in case of subject *L* whose record is shown in Table V. all the responses below the horizontal lines in the several columns under right choices have been omitted. The right choices are *V*, *I*, *K*, *E*, *B*, *T*, *C*, *Y*, *S*, and *H*. The results of this method are summarized in Table VII., in which we find, therefore, the total number of each kind of response up to the point of complete learning in the several parts of the maze brought together. These results are shown both in absolute numbers and in per cents. in the same table. Here we do not need to take note of the several trials, but can consider all trials together. The summary of per cents. affords easy comparison with the results given in Table VI. obtained by the more unsatisfactory method. In Table VII. it is seen that expectations based on frequency and recency of responses are satisfied oftener in the wrong responses than in the right ones. This is probably due to the persistence of wrong responses when once started in any part of the maze, a persistence that had to be overcome by factors other than those here studied.

Table VIII. is an analysis of the percentage data given in Table VII. While this analytical table shows a preponderance of responses in accordance with frequency and recency on the whole, it also shows that in the case of frequency there are more wrong responses in accordance with the law than there are right responses, and more right responses against it than wrong responses against it. In the case of recency there is almost an exact equality of *R*- and *r*-responses in the right choices on the one hand, and in the wrong ones on the other; that is to say, *agreement with recency has no relation at all to the rightness of the response, and in the case of frequency, agreement with the law is somewhat more likely to occur in the wrong than in the right responses while disagreement is more likely to be found in the right than in the wrong responses.* This result agrees with our theoretical determinations in the early part of this paper, that learning in the present form of the maze is not effected by recency or frequency factors, or by both combined, and it gives additional support to our conclusion of the year 1917, based on the learning of white rats in the ordinary maze problem, that such learning must go on in spite of—rather than because of—frequency and recency influences. This statement applies only to the *modification of the order* of successive responses and the inhibition of irrelevant acts, not, of course, to the mere fixing of any act, whether right or wrong, in a determined series. So far as the writer is aware, his is the first case of experimental work that has put to actual quantitative test the current and orthodox doctrine, coming down to us from the extreme association psychology, that learning is brought about wholly or in large part by the effects of recency and frequency factors. We have now on hand extensive results of experiments on about six hundred white and negro children in multiple choice learning, which agree substantially with our present results in affording evidence against this doctrine derived from the study of 'the association of ideas.'

Comparison with the Results of Other Experiments and Theoretical Considerations

The data of the earlier experiment on maze learning were obtained from an analysis of the responses of seventeen white rats in their first three trials in mazes. Two mazes of different form were used, but there is close correspondence between the results in the two cases. We shall therefore give only the summary table here for easy comparison with our present results.

Table IX

A Summary of the Distribution of 1,232 Responses in the First Three Trails of Seventeen White Rats in Maze

Figures give the per cent. of total responses and symbols have same meaning as in Table V.

Trial Number	*B*	*R*	*Rf*	*Fr*	*b*	*r*	Totals
1	33.7	8.6	2.9	3.3	34.9	16.5	100
2	41.4	6.9	5.4	9.1	27.2	9.9	100
3	48.3	6.1	6.3	9.2	23.4	6.6	100
1–3 together	41.7	7.1	5.1	7.7	27.9	10.5	100

In this table the symbols have the same significance as in the present study. Unfortunately, in the case of animal learning, no account was taken of whether the reaction in each case was right or wrong, but the complete study, to which the reader is referred, shows no evidence that frequency and recency factors bear any real causal relation to the learning. Indeed, though it was not possible in the animal maze to effect complete returns with each error, as in the present experiment, evidence showed that frequency and recency tended more strongly toward the fixing of wrong acts than toward the correction of errors, and that responses in accordance with the expectations based on these laws increased in successive trials (as a *result* of the learning), reaching one hundred per cent. with perfect learning.

While no other quantitative study of frequency and recency effects in constructive learning seems to have been made, there has recently appeared from the Chicago laboratory an interesting and ingenious study of a complex form of learning based on ideation and controlled by quantitative methods.[1] Unfortunately the author, although paying some degree of attention to frequency factors, or repetition,—without always being clear as to what it is that is repeated—accepts the traditional view of the effects of frequency, and makes no quantitative analysis with a view to evaluating such effects. It is, of course, possible that his method and data do not permit of such an analysis. Numerous statements as to the effects of repetition are made, however, statements that are not supported by adequate evidence, if by evidence at all; and they seem to be contradicted by other statements. The methods and results are too complex to explain here; we can take note only of some of the statements regarding the method of learning. We are told that "Perhaps the most important factor in the abstraction and selection of the essential elements of our problems is frequency of repetition of responses to those elements . . . those elements which served most frequently as stimuli were the first to acquire definite meanings and special value for purposes of control" (p. 104). Again: "The most obvious factors in the selection and accentuation of essential elements were frequency of repetition of elements and their relative

[1] Peterson, John C., 'The Higher Mental Processes in Learning,' *Psychol. Monog.*, 920, *18*, No. 7 (Whole No. 129).

nearness to a goal, or end of action" (p. 120). Yet explanations in subjective terms are used to supplement these effects of frequency and nearness to the goal, and certain mental functions seem to be elevated almost to the position of faculties with spontaneous powers, a practice that is still current in much of our psychology and that is, indeed, hard to avoid wholly, in appearance at least, because of the complexity of mental processes and of their stimulation by various indirect means. Thus we find in the present case that the all-important frequency factors are supplemented by 'attention,' by 'foresight of consequences,' by 'perception of relations,' by 'the setting up of hypotheses,' etc. In the early stages of learning, "owing to a lack of acquaintance with the elements of the problem, the span of attention to the elements is very narrowly limited. Subjects who have passed beyond the early stages of learning also revert to this form under the stress of emotion, self-consciousness, or fatigue, which tend seriously to narrow the span of attention to the elements of the problem. There are some individuals who, through native limitations of the span of attention seem doomed to a marked dependence upon this primitive form of arousal of hypotheses. They drew rapidly and, disregarding recently observed and often well-known facts, frequently jumped to unwarranted conclusions from which they showed little ability to extricate themselves by crucial tests" (p. 112). "Sometimes hypotheses originated in explicit analysis and comparison of the elements of one or more series. . . . At its best this process involves a clear insight into the causal relations of the phenomena in question and carries its own confirmation" (p. 113). The nearness of any response to a goal is also emphasized, as in such cases, as well as in less complicated circumstances, the consequences of tentative or hypothetical draws can be foreseen (47 ff.).

The *facts* of behavior here described agree well with those that have been observed and mentioned in our own experiments on the higher forms of learning; but the explanations not only imply spontaneous mental forces, but also controls within which frequency must operate. This implication contradicts the author's expressed view as to the importance of this factor. Occasionally these mental powers seem to be built up as a direct result of repetition, so that we seem to move in a circle. Indeed the author unwittingly reveals the fact that frequency factors operate only within limits imposed by certain other factors which really give direction to the learning that is either strengthened or counteracted by frequency and by recency. These quotations remind one of Thorndike's well-known explanation of selection in learning in terms of frequency, based on the very curious assumption that the various elements in the situation are 'responded to each once,' [1] an assumption of some controlling factor that prevents the mere repetition of the first erroneous act, expected on the basis of his all-sufficient principle of frequency as well as on that of recency. The sailing is, of course, clear when this assumption is granted.

Attention, foresight of ends and of consequences, hypotheses, etc., are but the subjective aspects, as it seems to me, of a selective process that is itself to be explained, and the explanation of it—though it certainly lies within the relationships of the life processes and the total stimulating conditions—is undoubtedly one of the most important problems before psychology today. The study in question, though it overlooks this problem, concealing it inadvertently behind subjective terms or mental forces, nevertheless gives many observations which seem to me to point out the direction of a solution. The author not only notes, as our own researches have shown, that subjects frequently get to respond selectively on the basis of some principle before they become explicitly aware of the principle,[2] but he also recognizes 'the combined associative pull of a number of disconnected elements observed in rapid succession' (pp. 111, 112). The explicit perception of relations and of principles seems to be but the mature manifestations, in subjective terms, of selective responses based upon the cumulative results of successive and simultaneous stimuli, not the causes of them. The

[1] Thorndike, E. L., 'The Psychology of Learning,' 1913, p. 41.

[2] *Op. cit.*, pp. 53, 109.

several quotations given above, as well as the objective results of the experiments, are entirely compatible with such a view. It is therefore probable that a detailed study of the responses in this type of complex analytical learning would also have revealed, as our present study has done, the inadequacy of frequency and recency factors as directors of the learning process.

Our notes as well as our quantitative data show that in the modified mental maze learning, the effects of any stimulus and of its resulting response do not fade away immediately. The range of the series of responses, as we have seen, gradually enlarges itself by breaking away from such repetition of errors as the effects of frequency and recency would bring about. On the subjective side of the responses we noted, in correspondence with these changes, that certain choices come to be avoided either with a dislike for the stimuli immediately following them—the *N-V* situation—or with explicit recognition of the fact that they are errors and throw one back. A cumulative effect of the successive stimulus-reaction processes seemed to build itself up and to inhibit errors. The neural changes corresponding to this modification in behavior are not yet understood, but the general fact itself that some change goes on bringing about an overlapping of stimulus effects is indisputable.

When an individual is stimulated in any manner the nerve impulses flash through the organism in but a small fraction of a second. But there is considerable evidence to show that the effects of this stimulus do not so immediately fade away. Probably the responses of muscles and glands set up other afferent nerve impulses, which in turn, being reflected up through cortical and subcortical coördinating centers, or synapses, out through efferent nerves, bring about further responses; and so on. Thus streams of impulses must pour in and out in various directions for some time, gradually decreasing in intensity—or in extent according to the 'all-or-none' law—until the disturbance is entirely abated except for changes of permeability of synapses that it has effected. These streams of impulses therefore will exist contemporaneously with subsequent stimuli and exert important directive influences on the nerve impulses these stimuli set up. Thus certain cumulative dispositions to responses in a measure consistent with all the stimulating circumstances will gradually take shape in the organism. All stimuli, moreover, have an effect on the organism in proportion to the general organic needs at the time of their occurrence. Thus when one is in a state of hunger the food stimulus is heightened. Every response tends to be complete, or consummatory, to the extent that it brings about a satisfaction—in the biological and not necessarily the psychic sense of the term—of these organic needs. It may bring recognition of self by others, or food, or it may promote self assertion, or arouse sex activity, or stimulate play activity when there is readiness for it in the way of 'surplus energy' or what not, etc. The overlapping of stimulus effects may thus be conceived as due, in part at least, to the rebounding of nerve impulses which involve the action of visceral muscles and glands as well as of skeletal muscles, and to the interactions of various organs upon one another.

It is this conception which the writer outlined in 1916, doubtless very imperfectly, as an hypothesis susceptible of experimental test.[1] A theory that does not lend itself to some sort of verification by experiment is of little value. It has been found, in conformity with the expectations based on the theory in question, that rats learn more readily to avoid errors in a maze with short blind alleys than in one with long blind alleys but in every other respect exactly identical with it.[2] It is not, of course, contended that no other explanation of these facts is possible; none has yet been given, however, and it seems clear that frequency and

[1] Peterson, Joseph, 'The Completeness of Response as an Explanation Principle in Learning,' *Psychol. Rev.*, 1916, *23*, 153–162. This first statement of the theory seems to have been badly misunderstood by one reviewer, who, by an unfortunate omission, misquotes an important passage. See Wiltbank, R. T., 'The Principles of Serial and Complete Response as Applied to Learning,' *ibid.*, 1919, *26*, 277–286. Mr. Wiltbank, by some means that I cannot understand, concludes that my theory neglects present external stimuli, and he consequently sets up to explain certain data, a theory that seems to be identical with my own theory except in name.

[2] Peterson, Joseph, 'The Effect of Length of Blind Alleys on Maze Learning: An Experiment on Twenty-four White Rats,' *Behav. Mon.*, Ser. No. 15, 1917.

recency factors fall short here. The present study shows in another form of learning, with human subjects, the inadequacy of these factors to account for the modification in behavior effected.

Indeed, many writers have shown some degree of recognition of this inadequacy of frequency and recency factors as complete explanation principles in learning. Thorndike has explicitly argued against the assumption that learning can be based on frequency of exercise,[1] and has added his 'law of effect,'[2] which is, of course, not an explanation at all but only a very rough statement of fact. Watson, though generally taking a rather extreme view with reference to the importance of frequency factors in learning, has recognized the probable operation of such other 'physiological principles' as 'reënforcement, inhibition, and summation of stimuli.'[3]

In the general psychology texts we still find attention, will, the perception of meaning, etc., playing the rôle of real selective agents in higher forms of learning; hence the inadequacy of the traditional associationism as applied to learning has not been so evident to the uncritical reader. It is in animal psychology, where these 'faculties' have not so obviously functioned, or where the assumption of such *forces* has been more obviously questionable, that the real problem of learning has been laid bare. It is here that our ignorance of the process has become evident. In this field the first tendency—as is seen in the writings of Lloyd Morgan and Thorndike, among the experimentalists—was to invoke pleasure and pain as the selective agencies. On the basis of frequency and recency alone it has thus far been impossible to explain how an animal reacts to situations, as we know that we ourselves do, rather than to a mere succession of stimuli.

But even if we grant the simultaneous operation of many stimuli on an individual by the means suggested above—the overlapping of stimulus effects—we are still not free from the obligation to explain how the more consistent acts survive

[1] Thorndike, E. L., 'Animal Intelligence,' 269 ff.

[2] Thorndike, E. L., 'Psychology of Learning,' p. 4.

[3] Watson, J. B., 'Behavior,' 275 f.

over the less consistent ones, or otherwise stated, how erroneous acts are eliminated. While in its technical aspects this is fundamentally a problem for neuro-physiology, the suggestion here of an hypothesis subject to experimental test may be *a propos*. If we grant that the basic life processes, controlled by stimuli (mainly internal) through the autonomic nervous system, are not only continuous but in a very real sense consummatory, it may be agreed that other acts are subservient and contributory to them. The synapses involved in the neural impulses directly affecting these autonomic processes may be thought of as permeable to a high degree. When these processes become impeded in any manner nerve impulses flow into various channels bringing about random, uncoördinated acts involving the skeletal muscles. How are these impulses again drained off as it were into the channels of the consummatory acts? Let us assume that each of these uncoördinated acts, causing conflict in the animal's behavior, brings about a shower of afferent impulses from the skeletal muscles, and that the consummatory acts—including not only the autonomic acts but also the acquired acts or habits making for their realization—bring fewer of these rebounding, afferent impulses because of less conflict. This assumption is certainly supported by introspective as well as by objective evidence; and the consummatory acts seem to spread their effects less extensively throughout the organism than do the random, uncoördinated acts. Now, it can be shown mathematically, on exactly the same principles that we have applied to the backward elimination of errors in the maze,[1] that these random impulses will be eliminated in the order of their proximity, in neural relationship, to the consummatory acts. The learning coëfficient, based on frequency, will be greatest at the synapses most closely associated with the autonomic processes basic to the present consummatory acts, because of the greater scarcity of rebounding impulses from these activities. Or if there is no greater scarcity of afferent impulses from these activities, the impulses may not be reflected back so directly to the specific synapses involved in the random conflicting acts, but may tend only to heighten the consum-

[1] See the first reference, above.

matory processes themselves and thus add to the general motivation. It is well known that certain processes, such as play and sex activities, tend toward their own continuation. In either event it can be shown mathematically that the spreading impulses will gradually drain off into the channels of those promoting the consummatory acts, first at the synapses next to those of these end-acts, and then, when these synapses have been sufficiently changed in resistance, at those next to them.

We shall have to develop this hypothesis more fully in another paper. The important point here is that such an application of our theory of maze learning is also applicable to nerve impulses in learning; and that it can be tested out experimentally by a study of the order of elimination of random acts in relation to their proximity to the consummatory acts.

Inhibition and facilitation must also be reckoned with in any complete theory of learning, but we are yet a long way from understanding their respective rôles in the modification of behavior.

Our experimental results, by a quantitative method applied to the laws of frequency and recency in learning, seem conclusive against the adequacy of these factors as explanation principles in learning; and these results must stand or fall independently of any relation to the theoretical considerations in this paper. It is plain that our knowledge of the real neural basis of learning is yet very limited. Whatever the neural processes in detail prove to be, it would seem that intelligence in general, and as manifest in learning in particular, consists in an arrangement of neuro-muscular mechanisms of such a nature that the organism is not subject to the mere contingency of the order of the occurrence of stimuli, but reacts to situations on the basis of general consistency. The degree of complexity of any situation to which one can learn to react efficiently is dependent largely on one's innate ability. On the basis of such a view it is obvious that one may profit by errors as well as by successes, a fact that is not easy to explain on the frequency-recency view.

Journal of
Experimental Psychology

Vol. V., No. 5 October, 1922

INFLUENCE OF VISION IN ACQUIRING SKILL

BY H. A. CARR AND ELLA B. OSBOURN

University of Chicago

In an article in a recent number of this Journal, one of the writers[1] reported the results of an experiment on the value of vision in maze learning. A certain amount of vision was found to be extremely effective in reducing the number of trials and errors required to master the problem so that the maze could finally be traversed without the aid of vision. Under the conditions of this experiment, the subjects could see not only the movements of the arm, hand and stylus, but also the pattern of the maze. Vision, however, did not enable the subjects to distinguish the true pathway from the cul de sacs. The motion of the hand and stylus was thus perceived in relation to the fixed pattern of the maze, and presumably this condition materially aided the subjects in identifying and remembering the correct pathway in visual terms. The conditions were somewhat analogous to those involved in learning to write, for in this case the motions of the hand and pen are perceived in relation to the permanent record of those movements.

The experimental procedure, however, failed to duplicate the conditions which obtain during the acquisition of most types of skill, such as walking, dancing, swimming, ball tossing, and typewriting by the touch method, for obviously in these cases the movements are not perceived in relation to some objective system of coördinates.

It was the purpose of the present experiment to create a situation more similar to those under which most acts of

[1] Harvey Carr, 'Influence of Visual Guidance in Maze Learning,' Jour. Exper. Psychol., IV.

skill are acquired. As a consequence our experimental procedure was arranged with a view to excluding the sight of the maze pattern while permitting the subjects to see the movements of the arm and stylus. An ordinary form of stylus maze was employed, differing in size and pattern from that previously used. The particular device used to exclude the sight of the maze pattern was a circular disc of thin aluminum plate, firmly attached to the stylus so as just to clear the surface of the maze when the stylus was in position. This disc was sufficiently large to cover completely the maze pattern for all possible positions of the stylus in the maze. Thus when vision was permitted, the subjects could see only the motion of the hand and the stylus with its attached disc, while any perception of this motion in relation to the objective pattern of the maze was prevented.

Five groups of subjects were tested. All subjects were without previous maze experience. All members of group I. learned the maze entirely without the aid of vision. This group is termed the 'standard.' Vision was excluded in the customary manner by the use of a cloth screen. To make the conditions comparable, the members of this group were also required to use the stylus with the attached disc. Groups II., III., and IV. were permitted the use of vision during the initial three, five and seven trials respectively and were then required to complete the mastery of the maze with vision excluded by means of the screen. The members of group V. were permitted the use of vision until each individual was able to run three perfect trials in succession, and then were required to complete the mastery of the problem with vision excluded. The average number of trials in which vision was permitted (not counting the three perfect runs) was 14.9.

The comparative data are given in Tables I. and II. In Table I. are given the visual conditions of learning, the number of individuals in each group, the average number of trials and errors required to learn, and the percentages of saving in trials and errors due to vision. In Table II. the error data are further analyzed into the two classes of

TABLE I

GROSS RECORDS AND PERCENTAGES OF SAVING DUE TO VISION

Groups	Number of Subjects	Trials	Errors	Percentages of Saving	
				Trials	Errors
I. Standard	15	22.1	181.4	—	—
II. 3 trials	15	16.7	87.4	24.4	51.8
III. 5 trials	15	18.9	95.4	14.4	47.4
IV. 7 trials	15	17.2	79.7	22.1	56.0
V. 14.9 trials	17	17.7	61.4	19.9	66.1

TABLE II

FORWARD AND BACKWARD ERRORS

Groups	Number of Subjects	Forward Errors	Backward Errors	Percentages of Saving	
				For.	Back.
I. Standard	15	58.6	122.8	—	—
II. 3 trials	15	30.7	56.7	47.5	53.8
III. 5 trials	15	32.6	62.8	44.2	48.9
IV. 7 trials	15	33.6	46.1	42.6	62.4
V. 14.9 trials	17	28.5	32.9	51.3	73.2

forward and backward errors. The forward errors consist of all entrances into cul de sacs while the subject is moving forward in the direction of the goal. All sections of the true path retraced and all cul de sacs entered while the subject is moving back toward the entrance of the maze are counted as backward errors. These data justify the following conclusions:

1. Vision, when permitted, is always effective in reducing the number of trials and errors involved in acquiring the ability to run the maze without the aid of vision.

2. Vision is much more effective in reducing the number of errors than of trials.

3. As the amount of vision is increased, its influence upon errors at first decreases and then increases. This statement applies to both types of error.

4. No definite relationship apparently obtains between the amount of vision and its effect upon the number of trials necessary for learning.

5. Vision is more effective in the elimination of the backward errors than in the elimination of the forward ones.

Vision is effective in reducing the number of errors not only during the time when vision is permitted but also during the subsequent period of mastery in which the use of vision is excluded. These two stages in the learning process may be termed the visual and the post-visual periods respectively. Obviously the way in which vision operates will differ in the two cases, and hence any explanation of the effect of vision upon learning must consider these two periods separately.

The error data for the two periods of learning are given in Table III. The first column states the length of the visual

TABLE III

ERROR RECORDS FOR VISUAL AND POST-VISUAL PERIODS

Groups	Visual Period				Post-Visual Period			
	Length	Total	Ave.	Per Cent.	Length	Total	Ave.	Per Cent.
II.....	3	25.8	8.6	39.1	13.7	68.2	4.9	59.0
III.....	5	40.2	8.04	50.5	13.9	45.8	3.3	43.9
IV.....	7	39.6	5.65	42.2	10.2	62.1	6.1	72.2
V.....	14.9	87.5	5.9	60.0	2.6	32.5	12.5	92.6

period. The second column gives the total number of errors prevented by vision during this period. The third column states the average saving of errors per trial, while the fourth column states the total saving in terms of its percentage relation to the number of errors made by the standard group for the corresponding period of learning. The remaining four columns give the same values for the post-visual period. The data of this table justify the following conclusions:

1. Vision is invariably effective in reducing the error scores for both the visual and the post-visual periods of learning.

2. Measured in terms of the total number of errors pre-

vented and the percentage of saving, vision exerted its greatest effect during the post-visual period in three of the four cases. Measured in terms of the number of errors prevented per trial, vision was most effective during the visual period for its shorter lengths, but when the visual period is long the subsequent effect of vision is the greater.

3. With the exception of the records for group IV., the total and the percentage values for the visual period increase with the amount of vision permitted. With the same exception, the opposite relation obtains for the average values.

4. Measured in terms of the average and percentage values, the influence of vision during the post-visual period at first decreases and then increases with the amount of vision permitted. No very definite statement can be made concerning the relation between the total values and the amount of vision, unless the values for group IV. be regarded as an exception.

These values for the post-visual period are complicated by the fact that the length of this period is highly variable from group to group.

5. Vision was not very effective during the initial trial of the visual period. The initial error score of the standard group was 23.7, while the corresponding value for the 62 individuals who were permitted the use of vision was 20.

6. The removal of vision operated to increase the error scores to a slight extent during the first two trials of the post-visual period. The comparative data are not given in the tables.

7. The subsequent effect of a given amount of vision is almost as great as the effect that would have been obtained if the use of vision had been continued until the maze was learned. This statement is based upon the data of Table IV. In the middle columns are given the error records of groups II., III., and IV. for the post-visual period. The error records for the standard group for the corresponding periods of learning are given in the first column, and similar data for the group that was permitted the use of vision for the entire period of learning are given in the final column.

Presumably the records of groups II., III., and IV. would have approximated those of group V. if the use of vision had not been discarded. In other words, three trials with

TABLE IV

ERROR RECORDS FOR POST-VISUAL PERIOD

Group I.	Group II.	Group III.	Group IV.	Group V.
115.6	47.3			28.2
100.4		54.6		23.6
88.0			25.8	21.0

vision (group II.) operated to reduce the error record during the post-visual period of learning from 115.6 to 47.3, while the continuance of vision throughout this period would have further reduced this score by 19.1 errors. Likewise a continuance of vision would have reduced the error scores of groups III. and IV. by 31.0 and 4.8 respectively.

8. The mastery of the maze in visual terms does not necessarily enable the subject to traverse it successfully with vision excluded. Group V. that was permitted the use of vision until each individual was able to run three perfect trials in succession required on the average 2.6 more trials to learn to run it when vision was excluded. Individuals differ in their ability in this respect. Of the seventeen subjects, nine ran the maze perfectly when the possibility of vision was eliminated, while five required from 8 to 15 trials with an average of about one error per trial to relearn it under the new conditions.

There are several ways in which vision may manifest its effectiveness during the visual period. It may give the subject greater confidence and assurance and thus eliminate much of the futile and aimless exploration characteristic of the periods in which the subject becomes confused, lost or discouraged. It may permit of a more accurate and definite localization of the goal and enable the subject to attain a better conception of the course of the true pathway. It may also aid in the definition of the cul de sacs.

On the basis of the above conceptions, one might expect that vision would be effective as long as its use is permitted, and hence that the total number of errors prevented would increase with the length of the visual period. With the exception of the records of group IV., this assumption is justified by the data of Table III. However, one cannot assume that the number of errors prevented will be proportional to the number of trials, for naturally more errors will be prevented in the initial trials in which many errors are normally made. Rather one should expect that the number of errors prevented per trial will decrease as the amount of vision is increased, and with one exception this assumption is also justified by the records.

The exceptional records of group IV. for the visual period may find their explanation in the following conception: During the course of the experiment, it was observed that the subjects invariably watched their movements very closely during the first few trials, then attempted to dispense with its use for several trials, and finally employed it again in the later trials. On the basis of this observation, it is suggested that the relatively poor records of group IV. may be due to the fact that they attempted to dispense with the use of vision from the fifth to the seventh trials, while the excellent records of group V. are due to the fact that vision was again employed after the period in which its use was discarded. Of course it is also possible that the records of group IV. may be due to some chance factor.

The effectiveness of vision during the post-visual period is due to one or both of two factors: (1) A memory of the visual data concerning the maze pattern acquired during the preceding period, and (2) to a more perfect tactual motor coördination resulting from the previous visual guidance. There are no data which indicate the relative effectiveness of these two factors. So far as the first factor is concerned, we would assume that the more adequate and accurate the conception, the greater will be its effectiveness, and that its accuracy, within limits, will vary directly with the length of the visual period. The effectiveness of the tactual motor coördination will vary with the degree of its perfection, and this latter may be expected to vary with the number of trials in which vision is permitted. In other words, it may be assumed that the effectiveness of the two factors during the post-visual period will increase in a general way with the length of the preceding visual period, and hence that, all other variables being neglected, the total number of errors prevented during the post-visual period will increase with the number of trials in which the use of vision was permitted.

The influence of these factors is, however, to a considerable extent counterbalanced by several other variables: One is the length of the post-visual period. It is obvious that, other things being equal, the shorter the visual period, the longer will be the post-visual period, and the greater will be the number of errors prevented. A second is the position at which the post-visual period begins. With a short visual period, the after effects of vision will be exerted at a time when more errors are normally made than is the case with a longer visual period. Third, the factor of visual memory is likely to disintegrate with time. So far as the first two variables are concerned, it is evident that the total number of errors prevented during the post-visual period will vary inversely with the length of the previous visual period. As to the third variable, it is evident that its effectiveness will tend to decrease with the length of the post-visual period and hence to vary inversely with the length of the visual period.

The total amount of error saved during the post-visual period is thus related to the length of the preceding visual period in two opposing ways,—one in which the number of errors saved will vary with its length, and one in which the number will vary inversely with the length. These two antagonistic tendencies may well account for the fact (Table III.) that no definite relation obtains between the amount saved during the post-visual period and the number of trials in which vision was permitted. The lack of any correlation between the two sets of data may also be due, in part, to the peculiar records of group IV. for the visual period. If vision was not effectively utilized throughout the visual period, this fact would obviously complicate its post-visual effect.

As a subsidiary experiment, we investigated the influence of the sudden introduction of vision upon the performance of a thoroughly mastered act, and the relation between the amount of such disturbance and the extent to which vision was employed in its acquisition.

Group I. had been required to master the maze without any visual aid. After the maze was learned in this manner, the screen was removed and the subjects were required to run the maze with vision permitted. Likewise all those groups that were permitted the use of vision during the initial trials and had acquired the ability to run the maze with vision excluded were now tested as to their ability to perform this act while visually observing it. In all cases the subjects were tested until all disturbance due to the introduction of vision was eliminated and three perfect successive trials were obtained.

The comparative data are given in Table V. In the

TABLE V

INFLUENCE OF VISION AFTER LEARNING

Groups	Number of Subjects	Per Cent. of Subjects Affected	Ave. Trials	Ave. Errors
I. Standard	15	87	6.9	7.7
II. 3 trials	15	27	2.0	2.2
III. 5 trials	15	53	3.2	3.7
IV. 7 trials	15	93	5.5	7.8
V. 14.9 trials	17	65	4.1	4.0

third column are given the percentages of the individuals affected for the various groups. The fourth column gives the average number of trials per affected individual required to eliminate the disturbance, while in the last column are found the average number of errors per affected individual involved in learning to adjust to the novel situation. Obviously these values are a measure of the degree of the resulting disturbance. These data justify the following conclusions:

1. The majority of the individuals (65 per cent.) were

disturbed by the introduction of vision, and this disturbance persisted for from one to fifteen trials. Usually not more than one or two errors were made per trial. The disruptive influence of vision was slight but quite difficult to eradicate.

2. A very considerable number of the subjects were immune to this influence and were able to traverse the maze without error.

3. With the exception of group IV., those subjects who had learned the act with visual aid were less disturbed by the introduction of vision than were those individuals (group I.) who had been denied that aid. The standard group (I.) and that permitted seven trials with vision (group IV.) were about equally affected.

4. With the exception of group IV., the amount of the resulting disturbance, measured in terms of the number affected and the number of trials and errors required to readjust to the situation, varies directly with the amount of visual guidance employed in acquiring the act.

The fact that the standard group that had been denied the use of vision in learning the maze was the most disturbed by the introduction of vision is to be explained largely in terms of the illusory conception of the magnitude of the movements which subjects normally develop under these circumstances. It is a well-known fact that subjects who learn a maze with vision excluded build up some sort of a conception of its visual appearance, and that almost invariably they grossly overestimate its dimensions. Naturally all subjects who are permitted the use of vision in learning the maze will develop a more correct conception of its magnitude. As a consequence the members of the standard group, when vision is introduced, will be distracted by the disparity between their conceptual and perceptual impressions, and will make a number of errors in so far as they attempt to guide their movements in terms of vision.

One would naturally expect that the greater the amount of vision involved in the acquisition of an act, the less will be the disturbance resulting from its later introduction. As a matter of fact the opposite relation obtains and the fact is not readily explained. Two suggestions may be offered: (1) Overconfidence and carelessness. The more nearly the act was learned while vision was permitted, the greater is the subject's confidence in his ability to run the maze correctly when vision is again introduced, and this confidence begets carelessness and resulting error. Especially does this explanation apply to the members of group V. They were permitted the use of vision until they ran three perfect trials in succession, and were then requested to make three more perfect runs with vision excluded. By this time, they were overconfident, somewhat bored and quite careless, and errors were made as a consequence. (2) Inaccuracy of visual conception. While the visual groups developed a more accurate conception of the magnitude of the movements than did the standard group, yet their conception of the details of the maze pattern may be more inaccurate. The greater the amount of vision permitted, the more detailed will be the conceptual schema developed, and it is assumed that the greater the wealth of detail, the more numerous are the possibilities of error. Mistakes will naturally be made in so far as these subjects attempt to guide their movements by means of these erroneous notions.

The fact that such a large number of individuals in all groups were not disturbed by the introduction of vision indicates that individuals differ very materially in the extent to which they normally rely upon a visual control in acquiring and performing such an act of skill.

DIFFERENCES IN THE ORAL RESPONSES TO WORDS OF GENERAL AND OF LOCAL SIGNIFICANCE

BY VIVIENNE R. McCLATCHY

University of Texas

This investigation in free association is concerned with an experimental study of differences—both quantitative and qualitative—in the oral responses to words of general and words of local significance. Interesting and significant facts relating to temperament and to mental mechanisms in general have been found by various writers in the qualitative and quantitative data furnished by word association experiments. Jung (3) analyzed and classified the kinds of responses given by various types of individuals, but he also noted in a general way the presence of certain signs, such as long reaction time, that he believed to be indicators of emotional complexes. Hull and Lugoff (2) made a study of various complex signs in free associations, chief among which were repetition of stimulus words, extremely long and extremely short reaction time, and complete inability to make any response to certain stimulus words. Dooley (1) is interested in using these various signs as means of studying common complexes rather than individual complexes. She is dissatisfied as to the use and value of reaction time length in word association studies. In the following investigation, conducted with 90 university students of undergraduate rank, 42 of whom were boys and 48 girls, I have attempted to determine the associational tendencies as indicated (*a*) by the reaction time, and (*b*) by the type of oral responses given by normal individuals to oral stimuli, when stimulus words of different classifications were used. The classifications, so far as stimulus words are concerned, are of two kinds, *i.e.*, general and local. So far as responses are concerned, I have assumed, *a priori*, that I should find both personal and impersonal answers, as well as responses of general and of local import. In other words, the responses to each of these groups could be classified in two ways—local or general, and personal or impersonal.

For stimulus words, two groups were used—words which had one of two connotative values. Either they were words of general significance, which might have been used with any group of normal people, or they were the names of local places which the students frequent, or the names of local organizations, or words to which the students themselves had assigned peculiar meanings. Forty-four of each kind, promiscuously sorted, were used. For convenience of the reader, I shall list these separately:

General Words	*Local Words*
long	McFadden's (drug store)
Dodge automobile	chili (favorite article of food)
coca cola	X———— (name of dean)
slow	Theta fraternity
ideal	Sigma Chi fraternity
cigaret	Woman's Building (dormitory)
over	Crescent Theatre
poker	Lake Austin
justice	Beta fraternity
sheep	Kappa fraternity
stem	Kappa Alpha fraternity
bank check	Students' Council
whistle	German (weekly dance)
automobile	Pi Phi fraternity
beautiful	problems (mathematics and physics)
fraternity pin	Cactus Tea Room
river	Newman Hall (girls' dormitory)
white	Deep Eddy (swimming resort)
rush	Curtain Club (dramatic club)
swift	"Pig" (campus dog)
wicked	Driskill Hotel
toddling	S. A. E. fraternity
fraternity house	buzzard (student who idles on campus)
heavy	San Marcos (pleasure resort)

hammer	Maverick Cafe
lion	Post Road (popular drive)
Ford automobile	Co-op (university book store)
prejudice	"Cosine" (nickname for mathematics professor)
cheese	date (social engagement)
fish	library
afraid	grave digger (gossip)
drink	Blunderbuss (April 1st journal)
comfort	Griffith's Drug Store
Buick automobile	Phi Mu fraternity
fruit	"B" Hall (boys' dormitory)
dance favors	campus
smooth	Majestic Theatre
mathematics	Clark Field (athletic field)
butterfly	Zeta fraternity
Bible	quiz
sweet	psychology
Stutz automobile	Kappa Sigma fraternity
false	Arrow-head (social club)
crap shooting	bust (colloquial for failure)

This list of words, as oral stimuli, was given individually to each of the ninety students. The oral responses and the reaction time, measured by a stop-watch graduated to one one-hundredth of a minute, were recorded on a blank arranged for the purpose.

The responses were sorted as to whether they were general and local, as defined above, or personal and impersonal. By personal responses, I mean those which have an obvious personal significance, such as 'sister—Mary,' 'cash—none'; while impersonal responses are those which seem to have no peculiar personal meaning and which might have been given by any normal person under any circumstances.

Table A contains the list of general stimulus words with the number of responses classified in two ways—local and general, or personal and impersonal. Columns II. and III. represent one classification of the ninety responses given to each of the stimulus words, and Columns IV. and V. represent another classification of the same responses. For ex-

Table A

Frequency Table for General Stimulus Words

Stimulus	Number of Responses						
I	II	III	IV	V	VI	VII	VIII
	Local	General	Personal	Impersonal	Total Time	Av.T.	Av.D
Dodge automobile	34	56	43	47	.47	.054	.028
long	5	85	7	83	.27	.03	.011
coke	56	34	17	73	.40	.044	.014
rush	48	42	30	60	.35	.039	.012
swift	5	85	8	82	.36	.04	.014
wicked	1	89	4	86	.34	.037	.013
toddling	57	33	18	72	.33	.036	.007
fraternity house	75	15	15	75	.46	.051	.016
heavy	9	81	12	78	.39	.043	.021
hammer	2	88	2	88	.30	.033	.009
lion	3	87	4	86	.33	.036	.007
Ford automobile	11	79	36	54	.48	.053	.021
prejudice	14	76	24	66	.47	.052	.021
cheese	2	88	6	84	.28	.031	.006
fish	11	79	8	82	.36	.04	.014
afraid	0	90	7	83	.38	.042	.015
drink	5	85	7	83	.36	.04	.013
comfort	0	90	40	50	.37	.041	.025
Buick automobile	19	71	45	45	.47	.052	.020
fruit	1	89	16	74	.31	.034	.008
dance favors	26	64	42	48	.49	.054	.015
smooth	1	89	7	83	.31	.034	.005
mathematics	76	14	55	35	.40	.044	.017
butterfly	31	59	11	79	.37	.041	.011
Bible	16	74	10	80	.33	.037	.013
Stutz automobile	29	61	34	56	.39	.043	.010
sweet	0	90	6	84	.23	.025	.006
false	4	86	4	86	.39	.043	.016
crap shooting	3	87	12	78	.33	.036	.008
slow	2	88	3	87	.28	.031	.014
ideal	6	84	17	73	.77	.085	.007
cigaret	6	84	9	81	.34	.037	.015
over	2	88	1	89	.32	.035	.010
poker	3	87	10	80	.39	.043	.014
justice	2	88	4	86	.41	.045	.022
sheep	0	90	2	88	.35	.039	.017
stem	1	89	3	87	.34	.037	.010
bank check	20	70	37	53	.42	.046	.016
whistle	23	67	24	66	.40	.044	.018
automobile	2	88	6	84	.33	.036	.012
beautiful	8	82	14	76	.32	.035	.012
fraternity pin	37	53	36	54	.49	.054	.020
river	19	71	1	89	.29	.032	.008
white	0	90	1	89	.23	.025	.006

ample, where Dodge Automobile was the stimulus word, the numbers 34 and 56 represent the ninety responses given to the word; the numbers 43 and 47 represent the same responses in a different classification. Column VI. represents the total time for the responses to each stimulus word; Column VII. is the average time; and Column VIII. is the average deviation.

Table B contains the same sort of information in regard to local stimulus words.

For quantitative investigation we may tabulate thus:

To General Stimulus Words from a Total of	3,960
Responses of Local Significance	675
Responses of General Significance	3,285
or	
Responses of a Personal Nature	698
Responses of an Impersonal Nature	3,262
To Local Stimulus Words from a Total of	3,960
Responses of Local Significance	2,709
Responses of General Significance	1,251
or	
Responses of Personal Nature	1,051
Responses of Impersonal Nature	2,919

These results permit the following interpretations: (*A*) for stimulus words of a general nature, out of 3,960 responses, 17 per cent. were of local significance while the larger per cent. (83) were of general significance; (*B*) these same responses classified as personal and impersonal showed 17.6 per cent. personal and 82.4 per cent. impersonal; (*C*) where words of peculiar local significance were used as stimulus words, out of 3,960 responses 68 per cent. were local, as opposed to 32 per cent. general; (*D*) these same responses in the personal-impersonal classification were 26 per cent. personal and 74 per cent. impersonal.

Comparing the two groups, I find that general stimulus words get one fourth as many local responses and more than two and one half times as many general responses as do local stimulus words. When the responses are classified as personal and impersonal, the local stimulus words get about one and one half times as many responses of a personal nature and not quite an equal number of impersonal responses as do the general stimulus words.

The time is interesting in showing a difference between the two groups of stimulus words. The responses to the general

TABLE B

FREQUENCY TABLE FOR LOCAL STIMULUS WORDS

Stimulus			Number of Responses				
I	II	III	IV	V	VI	VII	VIII
	Local	General	Personal	Impersonal	Total Time	Av.T.	Av.D.
McFadden's	72	18	6	84	.35	.039	.011
chili	17	73	15	75	.28	.031	.008
Dean's name	88	2	27	63	.47	.052	.022
Curtain Club	18	72	23	67	.42	.046	.026
"Pig"	7	83	10	80	.35	.039	.014
Driskill Hotel	49	41	32	58	.48	.053	.024
S. A. E. Fraternity	78	12	48	42	.53	.058	.017
buzzard	14	76	2	88	.37	.041	.015
San Marcos	55	35	9	81	.37	.041	.012
Maverick Cafe	34	56	24	66	.42	.046	.011
Post Road	24	66	25	65	.47	.052	.017
Co-op	34	56	8	82	.38	.042	.016
cosine	12	78	12	78	.38	.042	.009
date	51	39	63	27	.48	.053	.029
library	32	58	15	75	.32	.035	.009
grave digger	8	82	14	76	.42	.046	.012
Blunderbuss	81	9	21	69	.40	.044	.015
Griffith's	59	31	7	83	.45	.050	.019
Phi Mu Fraternity	84	6	54	36	.65	.072	.025
B. Hall	90	0	10	80	.40	.044	.016
campus	53	37	10	80	.38	.042	.010
Majestic Theatre	54	36	21	69	.40	.044	.019
Clark Field	87	3	11	79	.32	.035	.008
Zeta Fraternity	70	20	49	41	.45	.050	.026
quiz	90	0	30	60	.52	.058	.034
psychology	73	17	51	39	.41	.045	.018
Kappa Sigma Fraternity	87	3	37	53	.55	.061	.035
Arrowhead Club	84	6	15	75	.41	.045	.011
bust	69	21	19	71	.39	.043	.012
Theta Fraternity	85	5	42	48	.55	.061	.026
Sigma Chi Fraternity	83	7	34	56	.48	.053	.028
Woman's Building	74	16	34	56	.41	.045	.019
Crescent Theatre	71	19	15	75	.41	.045	.012
Lake Austin	67	23	11	79	.35	.039	.011
Beta Fraternity	78	12	40	50	.55	.061	.024
Kappa Fraternity	80	10	47	43	.52	.058	.021
K. A. Fraternity	79	11	27	63	.54	.060	.016
Students' Council	81	9	30	60	.50	.055	.020
German	63	27	8	82	.33	.036	.011
Pi Phi Fraternity	79	11	29	61	.46	.051	.014
problems	84	6	12	78	.31	.035	.030
Cactus Tea Room	69	21	16	74	.41	.045	.012
Newman Hall	52	38	22	68	.43	.047	.016
Deep Eddy	90	0	6	84	.29	.032	.008

stimulus words ranged from .085 minute for *ideal* to .025 minute for *sweet* and *white*. The responses to the local stimulus words ranged from .072 for *Phi Mu Fraternity* to .031 for *chili*. The average for the general stimulus words was .041, while the average for the local stimulus words was .047, a slight increase. In general, we may say that the responses to the local stimulus words were somewhat slower than to the general stimulus words.

From the data in the tables as well as from a study of the actual responses, the following results are suggested:

1. As has been indicated elsewhere, the average time for responses to local stimuli was somewhat longer than that for responses to general stimulus words. If interpreted, as is usually done in the case of long reaction time records, this result would seem to indicate that on the whole there was more of a tendency to inhibitions for local stimulus words than for words of a general nature. The next problem was to find where these inhibitions existed if at all. Choosing those stimulus words which showed an average reaction time of .05 or above, I find that nine out of seventeen of these were the names of Greek letter fraternities. The individual responses were indicative of a very prevalent anti-fraternity spirit known to be present in university circles among those who are not members of fraternities. One individual who had been no little chagrined at not having been invited to join a fraternity, and who had heard much adverse criticism at home in regard to fraternities gave very long reaction times and, on being questioned, seemed utterly incapable of expressing herself sufficiently to make any explanations. Several other non-fraternity individuals showed rather quick reaction time except in the cases where the names of fraternities were the stimuli. In these cases the reaction times were long and the responses obviously 'cover reactions.' Another of those stimulus words showing long reaction time was the name of the dean before whom the students have to appear for infringement of rules either of conduct or of scholarship. The name of a local hotel, where many of the student social affairs are held, gave an average reaction time

of .052 minute; 'Post Road,' the name of the best automobile drive in this part of the country, suggested various escapades by its average time of .052; 'date,' used in the sense of social engagement, brought out many possibilities of sex inhibitions; 'Blunderbuss,' the name of an April First yellow journal which has caused the expulsion of several students from time to time, came in this group with an average time of .050; the word 'quiz' was peculiarly significant at the time the records were made, due perhaps to the fact that mid-term examinations were in progress; the one remaining stimulus word of this particular group of seventeen was 'students' council,' the name of the student government organization which investigates violations of the honor system, etc.

2. Scrutinizing the average deviations of the reaction time for each group, it was found that the average of the average deviations for the local stimulus words was higher than the average of the average deviations for the general stimulus words. This fact would seem to indicate that there was a greater tendency to individual variation in regard to time where stimulus words of local import were used than where general stimuli were employed.

3. With the result from item 2 in mind I then compiled the frequency tables, examples of which are found at the end of the article. From this compilation I found that the total number of different responses given to local stimulus words was 1,645, whereas the total number of responses given to general stimulus words was 1,531. This would seem to suggest that there was greater variety in the responses to local stimuli than to general stimuli, or less of a tendency to stereotypy.

4. In the case of general stimulus words the relation between the number of local responses and the number of general responses was very similar to the relation existing between the number of personal responses and the number of impersonal responses; this condition did not exist in the case of local stimuli where the responses were local by a large majority though singularly impersonal. General responses and im-

personal responses may both be assumed to carry a rather large degree of superficiality with them as opposed to local responses and personal responses which might seem to involve considerable feeling tone. If these last mentioned assumptions be granted, then I feel justified in saying that local stimulus words are accompanied by more feeling tone than are general stimulus words.

5. It is possible that association to general stimulus words come more readily because they are older, more frequently used, and therefore more firmly established than are associations to the local stimulus words. Such suggestion is not improbable when we recall that in certain cases of amnesia those usable associations which remain longest are those of early formation. The best evidence, however, which I have been able to gather does not support this possibility. In order to determine if such condition existed, I chose without regard to their time, four general stimulus words at random, *i.e.*, *long*, *swift*, *heavy*, and *fruit*. I then found that the four words showed an average reaction time of .0365 minute. I then chose in the same way four local stimulus words which I thought would probably not have any emotional tone connected with them. These were *chili*, *campus*, *library*, and *Co-op*, and their average reaction time was .0375. Then I chose four local stimulus words—*Post Road*, *Students' Council*, *Dean's Name*, and *S. A. E. Fraternity*—which I suspected of having an emotional tone. The average reaction time for these words was .0538. These comparisons show a greater difference of reaction time between the two groups of local stimulus words than between the general stimulus words and either of the groups of local words. This result leaves open the possibility of general stimulus associations coming more readily because they are older than associations to local words, but as far as this investigation is concerned, the possibility is unsupported.

6. Again it is possible that the slowness of reactions to local stimuli can be accounted for in the length of time involved in the mechanism of choice. As has been previously indicated, the variety of associations to local stimuli was

greater than that to general stimuli. It is possible that the increase in reaction time was due simply to the necessity of making a choice from a larger rather than a smaller number of associations.

I conclude, therefore, that so far as this particular investigation is concerned, we find (*a*) that the time required for response to general stimulus words or local stimulus words indicates more possibilities of inhibitions than does the response to general stimulus words; (*b*) there is a tendency to greater individual variation as far as time is concerned in responses to local stimuli than to general stimuli; (*c*) there is a greater variety of responses to local words than to general words; (*d*) there is more feeling tone attached to local stimuli than to general stimuli; or (*e*) associations to general stimuli are more firmly established and therefore appear more readily; (*f*) the mechanism of choice for a greater variety of associations increases the length of the reaction time.

EXAMPLE OF FREQUENCY TABLE

Stimulus—Deep Eddy

19	swim	2	suit
37	swimming	3	pool
1	river	1	picnic
3	bathing	2	beach
2	Sunday	1	Barton
1	dancing	1	last fall
4	water	1	good time
1	girl	1	boy
1	deep	1	date
1	lake	1	blue
1	bill board	1	spring board
5	boats	1	run
2	shute		

Stimulus—Newman Hall

10	girl	1	group
1	Spaniard	2	dance

19 girls	1 barbs
4 club	1 sign
1 news	1 B. Hall
1 building	1 mutton
3 brick	1 tough
21 Catholic	1 itch
6 name	1 Guadalupe
1 home	1 Edith
1 21st St.	1 Polly
2 nun	1 Cardinal
1 living room	1 sisters
1 grass	1 tall
1 cousin	1 rooms
1 position	

Stimulus—Problems

46 math (mathematics)	1 solid
1 slow	1 class
7 arithmetic	1 fish
2 box	1 physics
6 work	1 ethics
1 maze	1 books
5 trig (trigonometry)	1 tired
5 hard	1 plays
2 solve	1 pencil
1 example	1 tedious
1 algebra	1 sheets
2 analytics	

References

1. Dooley, S. Correlation of Normal Complexes. *Amer. J. of Psychol.*, 1916, Vol. 27, No. 2.
2. Hull-Lugoff. Complex Signs in Diagnostic Free Association. J. of Exp. Psychol., 1921, Vol. IV, No. 2.
3. Jung, C. Studies in Word Association. London: Heinemann, 1918.

THE CARDIO-PNEUMO-PSYCHOGRAM AND ITS USE IN THE STUDY OF THE EMOTIONS, WITH PRACTICAL APPLICATION

BY J. A. LARSON

Research Bureau, Berkeley School for Police

A Preliminary Communication

Over a year ago we started to use a deception test based upon the correlation between physiological and emotional activities.[1] In brief, the technique consists in securing a continuous blood-pressure curve taken synchronously with a respiratory and a timing curve. A chronograph, preferably of the Jacquet type, records the time in fifths of seconds, or greater intervals, while the reaction time is further obtained by the use of a signal magnet. Readings are taken on one arm by a Tycos sphygmomanometer (auscultatory method). A normal record is obtained without any stimulus word or action. This is followed by a short, prefatory, word preamble, while tracings are being recorded, after which a series of questions are given. The questions are very simple and direct and deal with indifferent matters as well as the subject under investigation. The questions are followed by having the subject or suspect reply to a set of Kent-Rosanoff stimulus words where the reaction time and the reply is recorded, tracings being obtained throughout. Every third word has to do with the subject of the investigation and the others act as controls. In cases where the subject is not a suspect, but thought to be suffering from a complex, the Woodworth questionnaire is used in place of questions concerning a crime. Marked variations at certain questions sometimes bring a complex to light. In addition to securing checks on a single suspect, it is often possible to obtain from forty to a hundred controls in some investigations. Thus, in certain cases a

[1] J. A. Larson, *Journal of the American Institute of Criminal Law and Criminology*, Feb., 1921.

crime is committed in a house in which there are many people. Here everyone is considered a suspect and is subjected to a test run, and it is finally ascertained that only one person is guilty, with the result that all of the other persons ran clear records. In every case the questions were the same and given so as not to arouse any emotion except that underlying the deception syndrome.

After subjecting several hundred individuals to the deception test the following results seem significant:

1. The association words with the time reaction do not give as satisfactory results as the cardio-respiratory changes. We can say this definitely in cases where the suspect has subsequently confessed where, although there were marked and striking changes in the tracings, the findings by the association method were not significant. Therefore, any test based upon the reaction time alone is not as satisfactory as the variations in the tracings.

2. Quantitative blood-pressure determinations, although highly satisfactory in Marston's work [1] are not infallible. Thus in many instances, suspects who subsequently confessed did not show a rise of blood pressure great enough to be classified as a rise due to deception according to Marston. In many cases the cardio-respiratory curves show marked changes but no marked rise. This is to be expected, since the reaction of the suspect is going to vary according to his makeup. In some instances there is a sudden depression or a marked irregularity in rhythm. In others again there may be a marked increase in pressure which may be sudden and only last a few seconds. In some cases there is a sudden depression which is followed by a sort of compensatory increase.

The ordinary galvanometric methods are unsatisfactory for the same reason that quantitative blood-pressure determinations alone are. Instead of measuring the effects of the emotions in terms of blood pressure, they are studied in terms of changes in electrical resistance. These are un-

[1] Subsequently corroborated by H. E. Burtt, JOURNAL OF EXPERIMENTAL PSYCHOLOGY, Vol. IV., Nos. 1 and 2.

satisfactory as long as they are purely quantitative. Even when studied by means of the string galvanometer, the changes are not as easy to interpret as those in the pneumo-cardiograph.

3. In every case of deception, as checked by the cardio-pneumographic tracings and confession, there are marked changes in the records. The effect of the suppression or repression varies according to the temperament and physical character of the subject. There may be an increase or decrease in frequency, a marked depression or excitation or a more or less summative effect. H. E. Burtt has had definite results from the study of the respiratory curve alone.[1] He has studied the inspiration-expiration ratio by a special device and concluded that systolic blood pressure has a greater diagnostic value than the breathing. He obtained satisfactory results with the blood pressure even though the determinations were quantitative. As to the respiratory effects, he confirmed the findings of Benussi.[2] Burtt did his work with Marston. In addition to changes to the respiratory as noted by Burtt, other striking changes may be noted. There may be sudden inhibitions, summative and excitive effects.

4. In all cases of deception yet encountered the curve differs from that of the controls or the person who does not repress when questioned. In one or two cases where there were more than forty individuals involved, a preliminary test was held and from this two or perhaps three persons were picked for further investigation because of certain irregularities in their records. It subsequently developed that the effect was due to some other guilty association brought out by some stimulus word. After this association complex had been cleared up, the suspect was not affected by the question pertinent to the matter under investigation. In most cases there is a certain increase in tension in the controls, but this is of a general and not specific nature and is never marked and is easy to control. It appears, if at

[1] *Loc. cit.*

[2] 'Die Atmungsymptome der Luge,' *Archiv f. die Gesamte Psychologie*, 31: 1914, 244–273.

all, before the subject matter being investigated is touched upon. In framing and giving the stimulus words and questions proper precautions must be observed so that anger is not brought into play. It was found from the introspection of all of the subjects that there was never any anger aroused by any of the questions. The innocent, although they did not like the idea of being under suspicion, were always glad to be given an opportunity to vindicate themselves. In most cases they were much interested in the procedure. It might be mentioned here that the criminal record of the suspect in no way vitiates the emotional effect. In fact, it has been our experience that the recidivists are as much, if not more, affected as the first offender. In all of our cases of deception as checked by confession, there are very definite changes which are usually very pronounced. If a mistake is made in picking the culprit, it is due to misinterpretation.

5. The marked irregularities due to the effects of repression involved in the deception process disappear with the confession. If, however, a subject maintains a repression in successive tests, as a rule the effects continue, although he may know the stimulus word or question and when it is coming. In all cases up to the present time, whenever a subject is given the same test after confession, the record was clear. This same thing occurs when the suspect confesses when first questioned. The only difference, if any, in this curve is that the curve may show a slight effect before the confession is made.

6. Physiological or pathological factors do not appear to interfere with the test, provided that the subject is able to understand the questions and is not unfit mentally, as in some of the imbeciles and psychotic individuals. Thus if a subject has an irregular heart action, this is ascertained in his control and these effects are considered in the interpretation of the remainder of the record. If a subject is temporarily unstable because of some worry or physiological strain such as fatigue, menstruation, etc., this in no way interferes with the effect of emotional disturbances.

7. The cardio-respiratory curve is very useful in the study

of the emotions. Thus Sidis utilized the tracing obtained by the pneumograph to differentiate between the various elements in the case of a woman with a dual personality. Pictorially, the individual is represented in two ways, first by his present physical condition as shown in his heart and respiratory rhythm, and second by his reaction under stress, during questioning which may involve him in some crime. Thus a phlegmatic individual or a person with a hypo-thyroid insufficiency does not have the same type of curve or react in the same manner as the nervous, dynamic type or the individual with a condition of hyper-thyroidism. In the cardio-pneumo-psychograms the persons resolve themselves into groups which at first glance would seem to depend upon the temperaments or dispositions of the individuals. The cause, however, seems to be deeper, for the emotional reactions of an individual may depend entirely upon his physiological or pathological picture, as Dr. Berman emphasizes in his interesting work. Records may be grouped physiologically according to age, sex and other factors. In short any factor, normal or abnormal, which affects the heart and respiratory activity to any extent will show upon the record. This effect may be only transitory or of momentary duration. In some cases, as in certain girls during menstruation, there may be some changes from their records at other times. The pathological factors such as arterio-sclerosis, improper cardiac functioning due to disease, abnormal conditions induced by puberty, the menopause, pregnancy, etc., may give the records a typical appearance. In addition to depending upon physiological and pathological factors, the appearance of the record may vary with the mental condition of the subject, which in turn, however, depends upon underlying conditions. Thus a patient of the manic-depressive type may run a rather high or a sub-normal blood pressure, depending upon his condition. (See the work of Dr. J. D. Ball.) Persons who might be grouped together physiologically may be separated by their emotional reactions to various stimuli.

Interesting records have been obtained with drug addicts. The transition from the very sick, moaning, miserable individual to the cheerful person may be shown graphically by comparing the record of the same man before and after an injection of morphine or other drug. This transition may occur within a few seconds.

With the aim in view of ascertaining how far abnormal individuals may be grouped according to type, we are making a survey of several thousand individuals in the penal and insane institutions of the state. These cases are treated first by securing controls and then probing for complexes. In this work we will have the assistance of Dr. J. D. Ball, as well as that of other psychiatrists.

The writer wishes to acknowledge his appreciation of the helpful coöperation and suggestions of Mr. Vollmer, of this department, throughout the entire investigation.

A STUDY IN GRADES AND GRADING UNDER A MILITARY SYSTEM

BY ROBERT L. BATES

Lieut. Col., and Associate Professor, Virginia Military Institute, Lexington, Va.

The data included in this study are those of the junior class of the academic school at the Virginia Military Institute. This report was, of necessity, limited to the one class. The group selected represents a good sample of the work that characterizes the institution. In common with most ranking military schools, a rigid system of grading obtains here. In this school is found, from one point of view, a peculiarly homogeneous group. However, as is true of any considerable number of individuals, there is a range of intellectual ability. The discipline is of the most democratic character, in that the same regimen holds alike for all. The time of the student is, in truth, wholly occupied with necessary military and scholastic duties. An honor system prevails in such manner that no suspicion attaches to a student's work. The system is inflexible to the point that each individual student knows his exact limitations as well as his rights. 'Shaving the regulations' is considered as reprehensible as an overt act. These facts might be deemed irrelevant did they not imply equal scholastic chance for all. Under this educational system definite abilities, especially in leadership, have a possibility to assert themselves and mediocrity is not concealed.

A daily rating for each student is required of the teacher, based on a quiz or the development of a topic on the blackboard. In the matter of rating students, each teacher grades independently with but incidental knowledge of the marks given in departments other than his own. At the end of each bi-weekly interval the teacher submits a written report of all the daily grades for the period. These are so posted that each student not only has knowledge of his own marks but is aware of his relative standing with respect to other members of his own or of different departments. At the end of each four-week period a monthly report is compiled in like manner, though not for publication. These latter supply the office data, and are the basis of individual reports that are sent to the student's parent or guardian. At the end of each semester examinations are conducted and these are rigorously graded. A period of three weeks is given to examinations at the end of each semester. In that time three examination periods are allotted to each subject pursued. The subject is, prior to examinations, divided into three parts and each test is based upon the respective part in consecutive order. This necessitates the student's review of but a limited portion of the semester's work for each part of the examination. 'Cramming' can be resorted to under the system, but the method of rating employed discourages it and minimizes its results. An examination period is of three hours' duration.

Promotion, success or failure in a course, is determined by effecting a resultant of class and examination marks, class marks being weighted as according to the formula,

$$\frac{2\ \textit{Class mark} + 1\ \textit{Examination mark}}{3}$$

The final class mark is the average of all recitations in the particular subject for the semester. For a passing mark no final grade must fall below 7.0 whether it be examination or class grade. This procedure, as previously noted, helps to offset the effects of 'cramming' in the attempt to elevate a deficient final class mark and also to prevent the cessation of work where the student already has a high mark. The standard in every instance is 10.0 regarded as perfect, 7.5 as proficient and 0 as complete failure. To receive credit for a course of study a grade of 7.5 must be obtained as a resultant of the above formula. Classes range in size from 16 to 25 individuals. The group reported here consisted of 34 individuals in two sections, each of which recited the same lessons but on alternate days. As duties are not relaxed on Saturdays, three recitations a week in each study is the rule.

The recitations are for one-hour periods barring, of course, the few minutes necessary to shift classes.

To the casual observer the system of rating students seems cumbersome, but after acquaintance with it and its exactions it becomes a matter of routine and is burdensome to neither student nor to teacher. Each student may know at any time how he is classified with respect to his fellow classmen.

As much has been written regarding the validity of marking systems, the chief question that arises is: Where an exacting system of rating students prevails, how do the ratings of the respective teachers compare? Growing out of this another question that arises is: To what extent do the findings corroborate or refute the experience of teachers in other institutions of learning? That these questions might be answered more fully measurements were made by the Thurstone Psychological Test (IV.).

The final grades for five studies that closely resembled each other in the technique of presentation were examined. No two studies were conducted by the same teacher. It may be a significant fact that the teachers of American Literature and of History are graduates of this institution where, as students, they were familiar with the system. The instructor's task is to get the relative standing of the students as judged by their class work. It is almost impossible to obtain a numerical expression of quantitative ability. The merits or defects of a system of strict grading are best illustrated in the analysis of the data.

In making comparisons the Product-Moment method of Pearson was used to determine the coefficients of correlation. The methods employed to obtain averages and deviations were those of common practice.

The highest general average for any subject was for Psychology. The lowest average was in Political Science. The difference between these two extremes was .243. The highest and lowest medians were respectively for Psychology and History, these ranging from 8.23 to 7.90, with a difference of .330. The median and average for each subject taught were nearly identical, the average in every instance being in excess of the median to a very small extent. This narrow variation for a series of averages indicates that grading was

Data Sheet Representing Grades Given in the Various Subjects and their Averages

Individ.	Avs. of Grades	Psychology	Political Sci.	American Lit.	English Lit.	History	Thurstone Test
1......	9.22	9.19	9.44	9.16	9.12	9.21	107
2......	9.15	9.06	9.33	9.06	9.11	9.18	116
3......	9.12	8.74	9.13	9.24	9.39	9.09	132
4......	9.06	9.33	9.20	8.67	9.00	9.11	63
5......	8.78	9.05	8.83	8.73	8.79	8.50	111
6......	8.68	8.82	8.83	8.61	8.71	8.42	120
7......	8.65	8.55	8.74	8.52	8.68	8.78	106
8......	8.48	8.22	8.95	8.48	8.46	8.28	86
9......	8.33	8.19	8.16	8.62	8.35	8.32	80
10......	8.29	8.56	8.03	8.38	8.14	8.32	76
11......	8.22	8.60	8.12	8.13	8.31	7.94	98
12......	8.18	8.15	7.93	8.26	8.36	8.21	89
13......	8.17	9.00	8.06	8.08	8.09	7.63	87
14......	8.14	8.00	7.90	8.23	8.36	8.23	80
15......	8.06	8.24	7.66	8.18	8.20	8.00	87
16......	8.06	8.02	8.15	8.07	7.93	8.11	61
17......	8.04	8.20	7.94	7.91	7.94	8.23	91
18......	8.04	8.24	7.99	7.97	8.23	7.76	112
19......	8.02	8.25	7.81	8.05	8.19	7.78	99
20......	8.01	8.09	7.76	8.14	8.31	7.74	98
21......	8.00	8.00	7.91	8.34	7.97	7.78	130
22......	8.00	7.88	7.81	7.99	7.96	8.36	94
23......	7.97	8.35	7.91	8.08	7.76	7.75	104
24......	7.95	8.05	7.78	8.14	7.91	7.86	94
25......	7.87	8.09	7.83	7.92	7.89	7.61	79
26......	7.87	8.24	7.77	8.00	7.63	7.72	89
27......	7.82	8.25	7.83	7.82	7.43	7.77	69
28......	7.74	7.65	7.85	7.81	7.79	7.62	62
29......	7.68	7.59	7.72	7.87	7.48	7.72	78
30......	7.68	7.81	7.89	7.48	7.54	7.66	59
31......	7.63	7.64	7.58	7.72	7.66	7.56	72
32......	7.56	7.52	7.62	7.63	7.50	7.55	89
33......	7.49	7.82	7.44	7.59	7.16	7.46	58
34......	7.45	7.66	7.22	7.53	7.42	7.43	97

Averages and Deviations

	Gen. Ave.	Psychol.	Pol. Sci.	Amer. Lit.	Eng. Lit.	Hist.	Thurstone
Average........	8.159	8.276	8.033	8.188	8.140	8.079	90.40
Median.........	8.040	8.230	7.910	8.080	8.055	7.900	89.00
Av. Dev........	.368	.375	.414	.344	.434	.421	15.45
S. D...........	.478	.474	.592	.437	.535	.507	19.25
Interquartile range........	8.31 7.84	8.58 8.00	8.45 7.79	8.50 7.91	8.41 7.78	8.34 7.72	105.00 78.00

COMPARISONS ON THE BASIS OF GENERAL AVERAGE OF GRADES

General Average	Average Grade in Quartile	Thurston Scores for Quartiles	Inter-quartile Score Average
9.22 8.31	8.859	103.64	
8.31 8.04	8.156	83.41	91.97
8.04 7.84	7.976	100.53	
7.84 7.45	7.645	73.94	

done on approximately the same scale by all five teachers. This fact is corroborated by a study of the interquartile range. The deviations, both average and standard, merely indicate the degree of dispersion of grades as posted by each teacher. Here we find no wide discrepancies because of individual variation in grading. The greater these deviations the greater is tendency of instructor to reward the better student with high grades while the grades of those doing less efficient work are reduced. The higher deviation from the mean also implies a greater tendency, on the part of the teacher, to approximate the quantitative scale. The lesser deviation from the mean proves a tendency to give neither extremely high grades nor to give extremely low ones. The dispersion as shown in the grades for the different subjects has a narrow range as compared with the Thurstone Psychological Test. But one inference can be made and that is,—there is a reluctance on the part of the teacher to mark the student of low grades as to his quantitative ability. In giving daily recitation marks the tendency is manifest, from a study of class marks, to give a grade of 10.0 only in exceptional instances. A recitation is deemed of high order when a mark between 9.0 and 9.5 is given. This is proved by the magnitude of the grade in the upper elevation of the first quartile. For the average individual reciting an average lesson a grade ranging from 8.0 to 8.5 is given. This is disclosed in the study both of the averages and of the inter-quartile range for the respective studies. It is also disclosed, on the basis of the general average, that 65 per cent. of the class make a grade of 8.0

or in excess. This is the median grade. It is also noted that the median is a close variant of the average. A high degree of uniformity is manifest for all subjects presented. A low class mark ranges from 7.0 to 7.5. This becomes evident upon a study of the lowermost quartile. Individual No. 34 has the lowest average, 7.45. The lowest single grade given in one subject was for Individual No. 33 who received a final grade of 7.16 in English Literature. It is common practice for a teacher to give less than a grade of 7.0 for a daily recitation, but compensatory grades in excess of this number maintain an average, for the less efficient student, at approximately the 7.5 position. A study of grade averages, however, manifests a reluctance on the part of the teacher to give less than 7.0 for a daily recitation. All grading ultimately reduces itself to the subjective process of judging, but the question of criterion for judgment is answered. The average recitation is estimated in terms of group performance.

The question may be raised that the grades as averaged include examination marks. It may be repeated that examination marks constitute the equivalent of but one third of the final grade. A collateral study of these discloses the fact that both examination and class grades constitute a close variant of each other.

Subjects Taught, Coefficients of Correlation and Probable Error

Subjects Correlated		Coef. of Correlation	Probable Error
History	– Amer. Lit.	.890	.024
History	– English Lit.	.880	.026
History	– Political Sci.	.868	.029
History	– Psychology	.734	.053
English Lit.	– American Lit.	.914	.019
English Lit.	– Political Sci.	.839	.034
English Lit.	– Psychology	.781	.045
Psychology	– Political Sci.	.715	.057
Psychology	– American Lit.	.802	.041
American Lit.	– Political Sci.	.848	.033
Thurstone Test	American Lit.	.560	.079
Thurstone Test	English Lit.	.547	.081
Thurstone Test	History	.367	.100
Thurstone Test	Political Sci.	.372	.099
Thurstone Test	Psychology	.396	.098
Thurstone Test	General Average	.467	.090

The coefficients of correlation existing between the various subjects are high, ranging from .71 to .91, inclusive. This indicates that ability in class work is located by the different teachers and that the students are arranged in the order of their class proficiency. The coefficients exceed the probable error by numbers ranging from 12 to 45. Six times this relationship implies adequate correlation.

A study of the Thurstone Psychological rating in the light of individual averages for the various subjects taught discloses a wider dispersion than do class marks. When the coefficients of correlation between respective subjects and Thurstone Test, and between the various subjects themselves are compared, approximately the same correlative dispersion is found though, on the whole, the coefficients of the former are but half the latter. A range of .20 was found existing between the highest and lowest of each series of coefficients. The Thurstone coefficients range from 3.7 to 7 times the probable error. The data show that concomitant higher abilities lie in the first quartile for both test and study series. The inverse relationship prevails for the lowermost quartiles of test and final grade series. This is indicated under the 'comparison on basis of the general average.' The data also show that the English studies have a higher correlation with intellectual ability, as estimated in terms of the Thurstone Test, than have the other subjects. The highest coefficient of correlation is for the two Literature courses. On the basis of class proficiency, as measured by final grades, the first 25 percentile of grades had a Thurstone score average of 103.64, and the fourth 25 percentile had a similar average of 73.94. At first glance the inter-quartiles present an anomolous distribution. The third and second appear to be interchanged. This was an important factor in lowering the coefficients between the averages and the Psychological test. This distribution accentuates the reason why intelligence tests practically never correlate to a high degree with class marks. The reason is: the inter-quartiles contain the great mass of average ability. Concentration, application, diligence and interest distort the significance of test scores. In other words, these factors determine whether or not one is to over-rank or under-rank (where the reverse factors hold) his intelligence rating. Where one has strong compensatory qualities—as exemplified for Individual No. 4—and in addition further complicates his endeavor by sacrificing speed for precision, a low score may obtain in spite of definite ability. Individual No. 34 admits a handicap of expression both oral and written. In the light of our limited data it would seem that both the upper elevation and the lower limit of scholastic attainment can be guaged by the intelligence rating, but it proves no criterion for the average, ordinary, normal ability.

Summary

I. Under a definite and strict system of grading, students can be rated successfully in the order of intellectual rank—qualitatively rather than quantitatively. What constitutes deficiency is a matter of judgment on the part of the teacher. In this group whose data are presented there is accord among instructors as to those who are deemed deficient. Psychological tests disclose a wider dispersion than do grades. The most deficient individual has a final grade not far below the passin mark. This is the evidence of a qualitative rating.

II. A criterion for giving daily marks may be effected that gives the qualitative relationship between students where they are under constant observation in class. Where teachers are brought into the open with their daily grades, the ranking of intellectual ability becomes a matter of high correlative significance.

III. There is a close conformity between teachers' ratings and test scores for the first and fourth quartiles. The average grades are taken as the basis of comparison. An examination by Thurstone's Psychological Test is not wholly comparable to a system of marking, in that, the one necessarily assumes the existence of quantitative relationships and the other but approximates such relationships. In effect, the class grade is qualitative in its significance. This fact does not impair correlation of two such series.

Intelligence rating with approved tests becomes a positive, though not absolute, means of determining the intellectual ability of the uppermost and the lowermost 25 percentiles of class proficiency. The intermediate quartiles are indicative of average ability. An individual of the average ability—in terms of the inter-quartile range on the Thurstone scale—may easily, even unintentionally, distort his class rating in a positive direction through concentration, application and interest. The 'personal equation,' in military parlance, plays a conspicuous part in securing high or low grades for the person of average ability.

STUDIES IN DISSOCIATION

I. Changes in the Auditory Threshold Induced by 'Crystal Gazing' [1]

BY LEE E. TRAVIS

The State University of Iowa

Introductory

1. *Statement of the Problem*

Recently there has been awakened a keen interest in the subconscious, abstraction, dissociation and related topics. These terms have often been used in a loose manner for the simple reason that so little is known about the phenomena which they purport to represent. We know very clearly what we mean by dissociation when the spinal cord is severed. Here the higher centers are actually separated from the lower spinal centers and the cortico-spinal neurones can no longer influence the motor responses determined by the impulse from the anterior horn cells. Since the reflexes are exaggerated under such conditions we can assume that the cerebral motor centers prevent the free play of simple reflexes.

The question arises as to whether voluntary abstraction or dissociation is at all related to actual operative dissociation. Can an individual voluntarily remove in any degree the inhibitory effect of cortical control and become a reflex machine? If all individuals cannot make this dissociation, are there some abnormal individuals who can do so? It is believed that these are fair questions to ask and the group of experiments that have been started, the first one of which is here reported, are being undertaken in an attempt to answer them.

This first report is a study of the changes in the auditory threshold for a relatively pure tone while 'crystal gazing.'

[1] This paper reports an experimental study cairied on in the laboratory of the Psychological Clinic at the State University of Iowa under the direction of Professor John J. B. Morgan, Director of the Clinic.

2. *Previous Work*

There is very little to be found in the literature regarding the threshold of hearing during states of abstraction and mild dissociation. Dunlap[1] made some experiments on 'the effect of physical interruptions in subliminal phases,' which, although conducted from an altogether different point of view, hinted at the possibility of the threshold changing under certain conditions. His findings were the results of the summation effect. He used a subliminal sound of constant pitch, and the observers were to respond at the instant of the appearance and disappearance of the sound by predetermined clicks. In his experiments no attempts were made to alter the state of the observer after the normal threshold had been obtained.

Various writers have undoubtedly supposed that changes in the thresholds of the different senses occur under certain conditions and especially in states of hypnotism and somnambulism. Most of the evidence for such suppositions has been gathered from pathological cases, and suggests the need of carefully controlled experiments on normal subjects.

3. *Method and Apparatus*

The individuals who acted as observers for this investigation were either juniors, seniors or graduate students in the State University of Iowa.

The main piece of apparatus used in the investigation was the fourth model of the Iowa Pitch-Range Audiometer,[2] about which only a few facts will be presented in order to make the experiments clear. The tone generated is comparatively pure and can be varied in pitch from 30 v. d. to over 7,000 v. d. A tachometer is so calibrated that the pitch being generated can be read at any instant. The vibration frequency can be changed according to the desire of the experimenter by means of a rheostat which is connected in

[1] Studies from the California Psychological Laboratory, *Psychol. Rev.*, 1904, XI., 308.

[2] For a detailed description of this apparatus see paper by C. E. Seashore, *Lancet*, Oct. 15, 1919.

series with the motor. This adjustment was not used in this experiment since the pitch was kept constant throughout each experiment. The tone can be produced in the observer's receiver by connecting through a double-pole knife switch.

The resistance and hence the intensity of the tone as heard in the receiver is controlled by a device which consists of a series of twenty-one non-inductive resistances, installed in series with the generator. The first contact represents minimal intensity; according to experimental findings, it is not absolute zero resistance. The second contact has a resistance of .000025 ohm, and the other steps increase throughout the series by increments of 250%, totalling 2,264.3 ohms resistance for the maximal intensity. The loudness of the tone can be varied and designated in terms of these resistances, as each of these contacts becomes a connection with the watch case receiver terminals when the switch is shifted from one to the other.

The remainder of the apparatus consists of a clear crystal for gazing about 10 cm. in diameter, mounted over a 40-watt blue, mazda light bulb; a silent signal key; and a watch case receiver fastened to a head-clamp. The apparatus just described is in the light-proof, sound-proof, and jar-proof room of the psychological laboratory, while the audiometer is in an adjacent room.

The observer was seated comfortably at a table of average writing height with the watch-case receiver held firmly against one ear by the head-clamp. On the table was the silent signal key and the crystal, both of which could be moved to convenient places by the observer. The mercury key was quite sensitive, but still some effort was necessary to make contact. When contact was made, a buzzer sounded at the elbow of the experimenter. Each subject was instructed to try several positions with regard to key and crystal positions in order to find the easiest and most comfortable one.

The threshold of the observer was obtained under conditions identical with those under which the experiment was conducted except that there was no crystal gazing. The subject was to attend to the tones coming in the receiver and respond every time he heard one. It might be well to state here that the threshold was ascertained both before and after the experiment.

The threshold having been found, the light was then switched on beneath the crystal into which the observer was asked to gaze. The crystal gave the appearance of the heavens on a clear, star-lit night. Each individual was told to begin the abstraction by recalling some pleasant experience or imagine himself in a wished-for or cherished situation. If these failed to cause him to forget his present surroundings, he was advised to attempt placing himself into a hypnagogic state. Several of the observers did approach this latter condition. When the subject fully understood what was expected of him, he was again left alone, to gaze into the crystal.

After from three to five minutes signals were given to the observer in series of three, the series at least a half minute apart. The intensity of the beginning tones was generally two contacts below the threshold. If the subject responded to tones of this intensity, the loudness was further decreased by one contact at a time until no responses were received. In case no responses were given on two contacts below normal threshold, the intensity was increased until the observer did respond.

The experiments lasted from one half hour to an hour, depending partly upon results obtained and partly upon the time which could be spent by the individuals acting as observers.

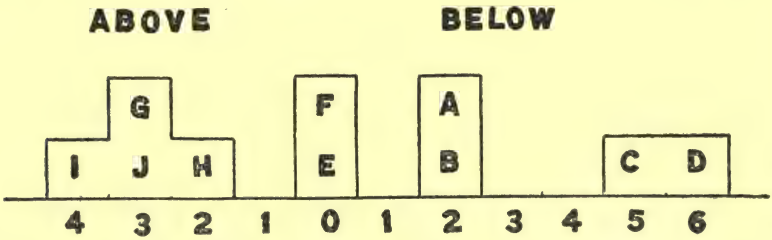
In the table each observer is represented by a capital letter. The thresholds, the tones of certain intensities and the responses to tones of certain intensities are represented by figures; each figure indicating the contact, a larger number being a louder tone. Column five is derived from columns three and four. To make this latter fact clear, let us take observer *D.* His normal threshold before and after the experiment was 14. Tones of intensity 8 were responded to accurately but no responses were made to tone of intensity 7. Hence his threshold during the experiment in a state of

4. *Experimental Data.*

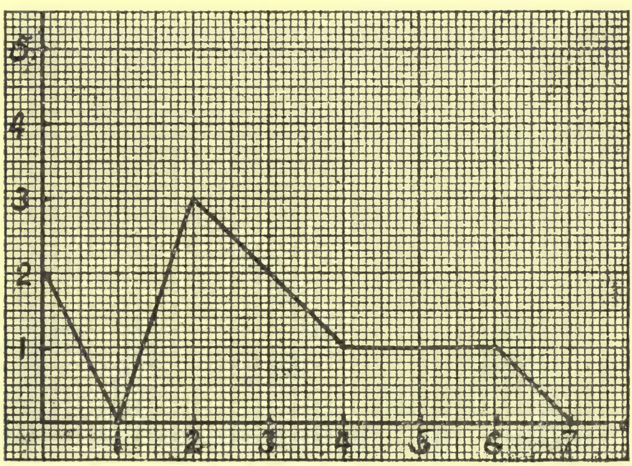
TABLE

Observers	Normal Threshold		Number of Tones of Certain Intensities													Number of Responses to Tones of Certain Intensities													Threshold During Experiment	Pitch	Length of Exp.	Time of Day	Errors
	Before	After	4	5	6	7	8	9	10	11	12	13	14	15	16	4	5	6	7	8	9	10	11	12	13	14	15	16					
A...	12	12					6	6	15											0	0	12							10	400 v. d.	60′	5–6 P.M.	3
B...	13	13								9	12	9											5	12	9				11	450 v. d.	40′	3:20–4 P.M.	6
C...	11	10	3	9	9	3	8	6								0	9	9	3	8	6								5	450 v. d.	60′	7–8 P.M.	6
D...	14	14				3	6	9	3	9	12								0	5	6	3	6	4					8	400 v. d.	30′	1:30–2 P.M.	2
E...	14	13									24	15	18											0	6	11			14–13	400 v. d.	40′	10:20–11 P.M.	1
F...	13	13								48	15	9	.										0	0	9				13	400 v. d.	30′	3:30–4 P.M.	0
G...	12	12							18	12	9	9	9	6								0	0	0	0	1	6		15	500 v. d.	60′	7:30–8:30 P.M.	0
H...	13	13								42	15	12	3	9									0	0	3	0	9		15	400 v. d.	45′	7–7:45 P.M.	0
I....	12	12							42	3	6	3	3	3	3							0	0	3	0	0	0	1	16	400 v. d.	30′	1–1:30 P.M.	0
J...	13	13								33	3	6	3	6	6								0	0	1	0	0	6	16	400 v./d.	40′	2–2:40 P.M.	2

abstraction was 8. In the cases of observers who required a tone above the normal threshold to provoke a response' that tone to which they would react was taken as their threshold during the experiment.



GRAPH I. Distribution of various subjects above and below their normal threshold.



GRAPH II. Distribution of changes in threshold during abstraction regardless of the direction of the change.

Upon investigation of the table, we find that the normal thresholds of eight out of the ten observers were changed during the experiment. The two whose thresholds were not changed probably were unable to abstract or go into a hypnagogic state. Referring to Graph I., we note that the normal thresholds of four of the observers were lowered, four were raised and two remained unaltered. In other words, four of the observers responded to subliminal stimulation, four required a stimulus above the threshold before they would respond and two of the observers reacted to stimuli on the normal threshold. Concerning the two observers whose thresholds remained the same throughout the experiments, nothing can be said except that they undoubtedly did not enter into the state experienced by the others.

One observer's normal threshold was lowered six contacts, each contact being an increase in ohms resistance of 250 per cent. over the previous one. Another's threshold was raised four contacts during the experiment. The other observers fell between these two extremes. Observer *D* remembered making only two responses when as a matter of fact he made twenty-four. During the experiment he made two false responses.

The case of observer *C* was slightly different. He did not know whether he responded or not, but knew the tones 'were coming across.' *A* and *B* were aroused by the muscular activity involved in pressing the key, when they made responses. *A* stated that he 'caught himself pushing the key.'

Responding seemed to bring these two observers, *A* and *B*, back to a normal state and they would not react again to subliminal stimulation until from three to five minutes had elapsed.

All of the observers whose normal thresholds were raised during the experiment stated that when they did hear the tone, it seemed very loud and abrupt. Immediately after a response to a sound which was far above the normal threshold, each of these individuals would respond to tones on the normal threshold and some time had to elapse before he would not react to such stimulation.

Interpretation of Results

To give an exact reason why the auditory threshold changes during states of abstraction is not an easy task. The work done in this investigation is too elementary to draw any definite and sharp-cut conclusions. At this time, it will have to suffice merely to give opinions and conjectures. We can only speculate as to why some observers responded to subliminal stimulation and others required sounds above the normal threshold to arouse them.

The preliminary practice meager as it was gave the subjects a certain attitude or set. This set carried over into the period of abstraction with four of the subjects so that the responses were made without the awareness of the observer and to stimuli that were below the normal threshold. We cannot be sure whether the difference in response was due to increased acuity of the receiving apparatus or to a facilitation in the nervous conduction brought about by the abstraction. If it were due to increased acuity, it seems that all observers should have reacted alike. If it were due to nervous set, then it is easy to conceive why some acted in one way and some in another. Questioning elicited the confession that those who required a stronger stimulus than their normal threshold tried not to listen for the stimulus; which means that they were set against it and so they required a more intense stimulus to break through this negative set. Those who obeyed the instruction and maintained the set to respond while they abstracted were enabled to react to a stimulus below their normal threshold.

This experiment indicates that the so-called 'automatic' activities that occur in dissociation need not be well practiced habitual responses. The important requirement seems to be that the subject accept the directions of the experimenter. The difference in response may be due to more or less readiness to accept the directions as given, or may be due to the fact that different individuals interpret the same directions differently and so maintain a different set.

Future Research

The sense of audition presents only one field to study changes in thresholds in states of abstraction. The visual and cutaneous senses are especially susceptible to a study of this sort and should yield as fruitful results as have been obtained in this investigation.

It would be interesting to compare the thresholds of the various senses during 'crystal gazing' with those in a hypnotic state. This could be quite easily studied with pathological cases.

Another phase of the problem which would undoubtedly lend itself readily to experimental investigation would be the ascertaining of the reaction time in crystal gazing and compare it with that under ordinary conditions.

After more detailed study has been made, it ought to be possible to standardize a test whereby an individual's ability to voluntarily abstract could be measured. If such a test could be devised, it should prove useful in detecting cases of hysteria.

Summary

This paper reports an experimental study of the effect of abstraction on the auditory threshold.

1. It was found that the auditory thresholds of eight of the ten observers changed during 'crystal gazing.'

2. The thresholds of four of the individuals were lowered while the thresholds of the other four were raised.

3. The responses made by those observers whose thresholds were lowered were without awareness.

4. When the individuals whose thresholds were raised during the experiment did hear the stimulus, it seemed loud and very abrupt.

5. It has not been sufficiently determined whether the responses to subliminal stimuli were due to increased acuity of the receiving organ or to the neural set involved, although the facts seem to point to the latter interpretation.

THE SELECTIVENESS OF THE EYE'S RESPONSE TO WAVE-LENGTH AND ITS CHANGE WITH CHANGE OF INTENSITY

BY I. A. HAUPT

Longwood Circle, Roland Park, Maryland

Introduction

The present study marks an attempt to deal quantitatively with the selectiveness of the response of the eye.[1] The eye shows selectiveness of response in three ways: (1) to wave-length; (2) to intensity; (3) to time of exposure of light. That is, both in its achromatic and its chromatic response the eye is unequally sensitive to wave-lengths of light; its responses, particularly its chromatic responses, are a complicated function of intensity of light, and they change irregularly with the time of exposure to light. In the latter regard the eye is very exceptional indeed as compared with the physical instruments. With increase of time of exposure, its response increases rapidly and irregularly, varying with the wave-length and the intensity of light, until a maximum is reached and then decreases, at first rapidly and then more slowly approaching zero as a limit. In short, the responses of the eye are a highly complicated function of three variables, the complication being much greater for the chromatic than for the achromatic response.

These peculiarities of behavior are supposedly evolutionary adaptations in the interest of the eye as an organ of seeing. They complicate and render difficult, however, many technical uses of the eye. From both points of view it is important to know them quantitatively and minutely, in order, first, that lighting and other important factors in our working environment may be most effectively adapted to the service of the eye and, second, that the most effective use may be made of the eye itself and in connection with instruments

[1] While the experimental part of the work was done in the Psychological Laboratory of Bryn Mawr College under the direction of Professor C. E. Ferree and Dr. Gertrude Rand, the author alone is responsible for the conclusions.

designed to extend and amplify its powers in the work of the arts and sciences. In the general subject of optics the emphasis given to the purely physical side has been entirely disproportionate. A great deal of time, money and energy have been expended on the study of light and physical optics, and little on physiological optics; yet both of these subjects are ultimately useful over a wide range of their application only in so far as there is knowledge of physiological optics. Thousands of dollars have been spent in the production of lights and lighting with never a question raised as to what the eye needs with regard to the factors of composition, intensity and distribution. This is but one example selected from a number. Along many lines of applied optics further advancement must come first from the physiological side.

Little is known in a quantitative way of the chromatic sensitivity of the eye or its variation with the factors wave-length, intensity and time of exposure. In prior studies in this laboratory all of the above aspects of the selectiveness of the achromatic response of the eye have been subjected to quantitative determination. No quantitative work has been done, here or elsewhere, however, so far as we know, on the selectiveness of the chromatic response. While the chromatic response has not in general the utility and practical importance of the achromatic response, still in proportion to what is known there is perhaps a greater practical need at this time of a detailed knowledge of the chromatic than of the achromatic powers and peculiarities of the response of the eye. It has been the purpose of this study to inaugurate a series of systematic investigations of this neglected phase of physiological optics.

It was considered desirable to conduct the investigation with foveal stimulation, since foveal vision is perhaps the greater and more important part of our seeing. Since the stimulus field had to be divided into standard and comparison halves, it was desirable to work with as large color fields as possible. In attempting to satisfy both of these conditions one is confronted with some difficulty because of the conflicting data and beliefs as to the exact size of the fovea.

According to Kollicker's[1] measurements, for example, the foveal depression is from .18–.225 mm. in breadth, the visual angle equivalent of which is .42–.92 min. These measurements are quoted and accepted by Helmholtz.[2] J. G. McKendrick says:[3] "The yellow spot has a horizontal diameter of .2 mm. and a vertical diameter of .8 mm. and it corresponds in the visual field to an angle of from 2–4 degrees. The fossa (*fovea centralis*) has a diameter of only .2 mm. which makes the angle ten times smaller." Koster[4] measured four cases and gives the dimensions of the rod-free area as .44–.90 mm., with .5 mm. as the most probable average diameter. Schaeffer[5] says "most observers give the foveal depression a diameter of 0.2–0.4 mm. or of 0.75–1.50 degrees." Tscherning[6] accepts and quotes these values. König[7] assigns 55–70 min. as the angular measurement of the apparent fovea of his right eye.

Howell [8] offers the following explanation of the discrepancies of the different observers: "In some cases the measurements are made from the bottom of the depression, the fundus, in others from edge to edge. . . . In the center of the fovea lies a smaller, very shallow depression spoken of as the foveola. While measurements on preserved specimens give the diameter as .2–.4 mm., ophthalmoscopic examination seems to indicate that in the fresh state it may be larger." The latter statement is in agreement also with Quain.[9] In the 'International System of Ophthalmic Practice' [10] one finds the statement that the fovea and the macula are identical. Foveola is the term applied to the depression. This is situated slightly outward and downward from the center of the papilla. The *Fundus Oculi* defines the limit of exact

[1] Kollicker, A., 'Verh. der Wurzb. med. Ges.,' III., p. 316, 1852.
[2] Helmholtz, 'Physiologische Optic,' I. ed., p. 21, 1867.
[3] J. G. McKendrick, 'A Textbook of Physiology,' Vol. II., p. 631, 1889.
[4] Koster, *Arch. f. Ophth.*, 1895, XLI., part 4, p. 1.
[5] Schaeffer, 'A Textbook of Physiology,' II., 1900, p. 1103.
[6] Tscherning, 'Physiological Optics,' 1904, p. 36.
[7] König, 'Gesammelte Abhandlung zur Physiologischen Optik,' 1903, p. 355.
[8] Howell, 'Textbook of Physiology,' 1907, pp. 314–315.
[9] Quain, 'Textbook of Physiology.'

vision by the macular ring 1.75–2 mm. in diameter or 4.5–5 degrees visual angle. Parsons describes the foveal region as elliptical, horizontal axis 0.3 mm., vertical axis 0.2 mm. The angular measurements of the fovea centralis are given as 55–70 minutes. The rod-free area subtends an angle of 3° 3′ and the macula 4° 12′. In recent experimental work Ives chooses a visual angle of 2° as falling well within the fovea, but gives no authority or citation in justification of his choice. The following table will serve as a summary of the more important estimates.

TABLE I

ESTIMATES OF THE SIZE OF THE FOVEA

Name	Angle	Diameter
Kollicker	42′– 52′	.18–.225 mm.
Koster	1° 40′–3° 26′	.44–.90 mm.
McKendrick	42′	.2 mm.
Schaeffer	42′–1° 30′	.2 –.4 mm.
Tscherning	42′–1° 30′	.2 –.4 mm.
König	55′– 70′	
Howell	1° –1° 30′	.3 –.4 mm.
Ives	2°	

The color field used by us in this work was a circular opening in the screen between the focusing lens and the eye, at a distance of 270 mm. from the eye. The visual angle subtended was 1° 55.8′.

II. EXPERIMENTAL

A. Purpose of the Investigation

In general the response of the eye to colored light takes the following course: at low intensities it is sensed as colorless; at a certain higher intensity, varying with the wavelength and the individual eye, the chromatic response is aroused. With further increase of intensity there is a more or less regular increase in the chromatic component of the sensation until a maximum is reached, the value of this component and the point at which the maximum is reached again varying with the wave-length and, within certain limits, with the individual eye. Still further increase of intensity results in a decrease in the chromatic component until a point is finally reached with very high intensity at which the sensation is again colorless or approximately so. The change of intensity results also, varying with the wave-length of the stimulus, in changes in the hue of the color sensation.

The present study includes the determination of the following points:

1. The amount of energy needed to give the threshold of color at seven points in the spectrum. These results enable one to compare the chromatic sensitivity to wave-length at the threshold of sensation and to plot a chromatic visibility curve at these minimum intensities.

2. The sensation is laid off in just noticeably different steps from the threshold to the maximum of chromatic response and as far beyond this point as produced any further changes in saturation or as the limits of the apparatus would permit. A comparison of the increments of light energy required to produce these just noticeable changes for any one of the stimuli reveals the selectiveness of the eye's chromatic response to intensity for that range of wave-lengths. A cross comparison between the series for the different stimuli at points representing an equal number of just noticeable differences from the threshold of sensation represents the selectiveness of response to wave-length and its change with change of intensity.

3. The point of maximum response has been determined for each of the stimuli used and for several observers and the amount of light required to give this response has been measured both radiometrically and photometrically. The wave-length and the radiometric specification fixes the point in absolute terms, the wave-length and the photometric specification in terms more in conformity with those used in lighting and other subjects in which practical applications are made. Another important feature of these determinations is the knowledge they give of the relative power of response of the eye to the different wave-lengths. That is, the response is laid off in just noticeably different steps until a maximum is reached. The total power of response may in each case be rated directly as the number of steps from

zero to maximum. To those who believe that the difference threshold may be taken as a unit in terms of which to rate sensation intensities this estimate should be considered correct within the limits of experimental errors: to those who do not believe that the just noticeable difference represents equal steps or amounts of response the rating can be considered as only approximate, but as nearly correct as possible in the psychophysical procedure. Total power of response is a new feature in the comparative study of the characteristics of response of the sensory mechanism. A correlation between the total power of response and sensitivity is possible and even probable, but by no means *a priori* necessary.

B. Apparatus

It is obvious that if the variation of the chromatic component is to be traced in just noticeable difference steps from minimum to maximum and back again to minimum we must be able to produce two stimuli independently variable in intensity over a wide range of intensity, identical in composition, and as nearly homogeneous as to wave-length as possible. We must be able also to present this stimulus to contiguous parts of the retina, and to vary each in turn by just noticeable difference steps; and to specify the intensities at every point radiometrically or photometrically or in both ways as the purpose of the work may render desirable. The greater part of the apparatus used for accomplishing this purpose has been in use in Bryn Mawr laboratory for several years for work on color sensitivity.[1] Only such devices were added as were required to adapt it to the needs of the present study.

A description of the apparatus resolves itself into three sections: I. The spectroscopic system, II. The intensity reduction system, III. The photometric system. In order that the detailed description of the apparatus may be more readily understood, the course of the beam of light is given as follows: the light rays from the source pass through the collimator lens, are rendered parallel, pass through the carbon

[1] C. E. Ferree and G. Rand, Journ. Exp. Psych., Vol. I., No. 3, 1916.

bi-sulphide prism and are focused in the plane of the analyzing slit by the objective lens. Beyond the analyzing slit the diverging rays are again rendered parallel by an auxiliary collimating lens, making it possible for reduction purposes to intercept the entire beam of light with the sector disk. Finally the rays are focused directly on the eye with an auxiliary focusing lens. Figure I. represents schematically

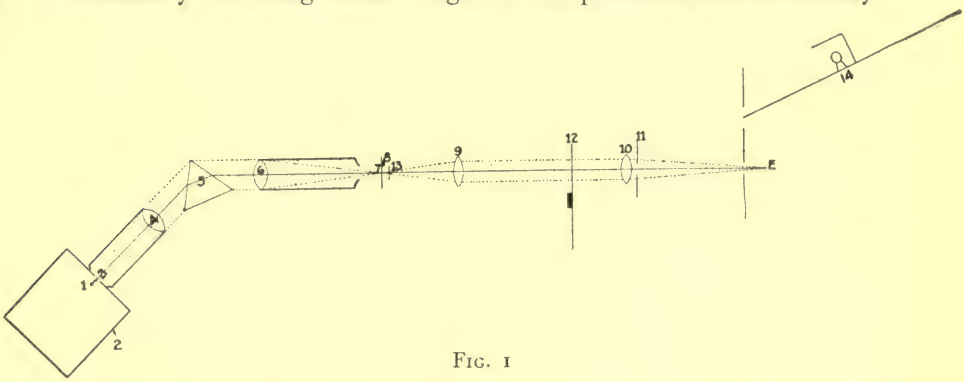


FIG. 1

the arrangement of the apparatus; Figure II. is a photograph of the apparatus as it was actually set up for use.

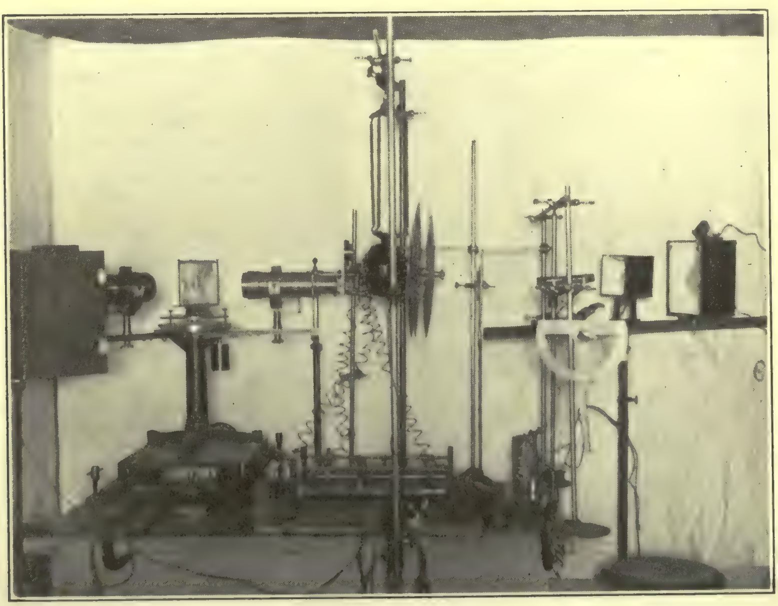


FIG. 2

I. *The Spectroscopic System.*—In detail the spectroscopic system consists of the Nernst filament (1),[1] a special light-proof box (2), made of sheet iron, 15 x 28 x 20.5 cms., with a light-proof ventilator in back and on top. The box is fitted directly on the collimator tube and encloses the Nernst mounting and the collimator slit. It is provided with a door opening all of one side, thus permitting easy manipulation and adjustment of the filament and slit. The advantage of such a box is its efficiency in reducing the amount of stray light without the use of cloths and screens. Such a box can easily be constructed by any laboratory mechanician at very little expense.

The collimator slit (3) is 11 mm. long and is set at a constant width of 2 mm., allowing a maximum flux of light from the filament to illuminate the surface of the prism. The collimator lens (4) is a Zeiss triple achromat, 60 mm. diameter, 180 mm. focal length. The prism (5) is of the hollow glass type, filled with CS_2, 100 mm. high, with a refracting base of 85 mm. and a refracting angle of 60°. With the exercise of a reasonable amount of precaution to keep the CS_2 free from impurities and to maintain a constant temperature in the room this prism has given satisfaction. Emphasis should be placed on the importance of maintaining a uniform temperature in the room when using a CS_2 prism. If this is not done, considerable annoyance is caused by the need frequently to check up the wave-length falling on the analyzing slit. Large changes in temperature cause a change in the refractive index of the CS_2, thus causing a shift in the wave-length falling on the analyzing slit.

The objective lens (6) is another Zeiss triple achromat, 60 mm. diameter and 240 mm. focal length. Both the collimating and the objective lens are mounted in brass tubes, painted black on the inside. The entire beam passes through these, but stray light reflected from the surfaces of the lenses, etc., is to a considerable extent absorbed by the walls of the tubes.

[1] *Ibid.*, page 253.

The analyzing slit (7) is placed in the plane of the spectrum focused by the objective lens. It is set at a constant height of 4 mm. By observation this gave the maximum size image that would fall within the normal pupil. The width of the slit is .4843 mm. The slit is mounted on an oblong brass frame, 12 cms. long and 6.5 cms. wide. On the side nearest the prism are attached holders into which are inserted small gelatine filters especially selected to increase the purity by limiting the light transmitted to the range of wave-length selected for investigation. It is very important in all work on color sensitivity which is meant to be quantitative that light as nearly homogeneous as to wave-length as possible be employed. The presence of alien wave-lengths affects the determination of chromatic sensitivity in 2 ways: (*a*) through the physiological inhibitions and interactions it decreases the amount of color response; (*b*) it increases the value of the energy measurements. On the other side of the brass frame, away from the prism are grooves into which is inserted for radiometric purposes the linear thermopile (8) with its receiving surface facing the direction of the slit.

The energy measurements were made with a linear thermopile having bismuth and silver couples, designed by W. W. Coblenz of the Bureau of Standards. A Paschen small coil galvanometer constructed especially for the thermopile was employed and suitable auxiliary apparatus. This group of apparatus was designed and constructed by request for this laboratory for quantitative work on color sensitivity. A detailed description has been given in a previous article.[1] The apparatus is so planned that energy measurements may be made at the objective slit, at the stimulus opening in the screen situated just in front of the focusing lens and at the eye. A description at one of these places, namely at the objective slit, sufficient to show the method employed is given in the above-mentioned article. It was decided in this study to specify the energy density (*i.e.*, the energy per square milli-

[1] *Psych. Rev. Monog.*, 1917, XXIV., 'Radiometric Apparatus for Use in Psychological Optics,' C. E. Ferree and G. Rand.

meter) at the analyzing slit and at the eye. If desired the energy density at the stimulus opening can be computed from the data given inasmuch as the dimensions of the stimulus opening are given and all of the light from the opening enters the eye. Of these specifications the energy density at the stimulus opening sustains perhaps the most direct and most nearly comparable relation to the energy density of the image formed on the retina, the actual stimulus for the sensitive elements.

As has been stated the apparatus has been so planned that the energy measurements may be made at the eye. However, the intensity of the beam from the objective slit is greatly reduced by the comparatively small stimulus opening needed for work within the fovea. Because of this reduction it is a matter of some difficulty to make the measurements of the small amounts of energy represented in the shorter wave-lengths with a satisfactory degree of precision. The energy was measured directly for the red, both at the analyzing slit and at the eye. In making the measurements the full height of the slit was used and all of the energy emitted from the slit measured and also in the image formed in the plane of the pupil of the eye. A factor was thus obtained which could be applied to all of the wave-lengths at the analyzing slit to reduce them to the proper value for the image in the plane of the pupil. To give the actual value entering the eye, however, a further reduction was necessary. That is, in order to insure against variations between values measured or computed at the pupil and the actual amount of light entering the eye, the light entering the eye had to be made independent of variations in the size of the pupil. Independence of change in the size of the pupil was especially needed in this work because of the large variations in the intensity of the light used. Such control is very easy to secure with the means of presenting the light to the eye used in the apparatus. All that is needed is to keep the image that falls on the eye of a constant size and smaller than the pupil throughout its entire range of variation in the given series of experiments. Not only can the variation be determined in preliminary experiments as a guide to the size of the pupil, but the image itself can be compared with the size of the pupil at every observation. This control of size of image on the pupil is secured by a very convenient attachment for the objective slit by means of which the length of the aperture of the slit may be varied by ½ mm. steps. Space will not be taken here for a detailed description of this device. In brief it consists of two knife-edge jaws moving in the vertical, operated by means of a ratchet and spring. Still finer control could be secured of course by means of a micrometer screw. Constructed in this latter form the device would be available as a means of producing finely graded changes in intensity. The dimensions of the image chosen, the selection having been based on a long series of observations under the conditions of work employed, were 3.475 mm. x .96 mm. The size of this image and the image measured by the thermopile were compared by measuring on a micrometer comparator the images formed on photographic plates mounted in the plane of the pupil of the eye. A comparison of these dimensions gave a correction factor by means of which the energy measured was reduced to the energy entering the eye.

To serve the purpose of the investigation the method of presenting the colored light to the eye had to fulfill the following requirements: (1) a method was required which would give the effect of a homogeneous surface or field of suitable size and shape which could be imaged on the retina; (2) the apparatus should be available for use in the light as well as in the dark room; (3) provision had to be made for controlling the brightness of the field surrounding the color, also of the pre-exposure of the surface to which the eye was exposed immediately preceding the stimulation to color; (4) the method of presentation had to be such as to permit of no admixture of light from the room with the range of wave-lengths needed for the stimulus; (5) it had to be as little wasteful of light as possible else a sufficiently high range of intensities could not be secured. In order to present the stimulus in compliance with the above plan, the rays of light emitted from the objective slit (7) are first rendered parallel by lens (9), Fig. 1, placed at 150 mm., its focal length, from the objective slit of the spectroscope, and then focused by lens (10), focal length 270 mm. directly on the eye at *E*. The focusing lens (10) is mounted 20 cms. from the auxiliary collimating lens (9) in order to permit of the easy introduction of the rotating sector disk (12) used as a means of reducing the intensity of light.

II. *The Intensity Reduction System.*—As has been stated above, if the chromatic response of the eye is to be traced from minimum to maximum by just noticeable difference steps, it is necessary to present to the eye two contiguous fields of color, one to be used as standard to which the other may be compared. Furthermore, we must be able to reduce the intensity to the threshold and to vary each field independently by small amounts for each step in the series. The range of variation must therefore be from minimum to maximum of the chromatic component of the eye's response, and beyond, in order to be able to determine accurately the position of the point of maximum saturation.

The primary reduction was effected by means of three neutral gray gelatine filters, mounted between glass, 2.5 x 2.5 cms. These filters, furnished by the Eastman Kodak Company, have their densities specified by the research laboratory of the Company. By means of the formula

$$\text{Density} = \log \frac{1}{\text{Transmission}}$$

the transmission for each filter was calculated. The densities and their transmission values are:

Filter	Density	Transmission
No. 1	1.3	.050118
No. 2	2.0	.01
No. 4	4.08	.000083
No. 1 and No. 4	1.3 and 4.08	.00000416
No. 1 and No. 2	1.3 and 2.0	.00050118

A special filter holder was designed so that two filters could be inserted in combination. The filters in their holder (13) were placed in front of and close to the analyzing slit, so that the area of the gelatine used should be as small as possible, insuring maximum uniformity of transmission.

Further reductions and fine variations in intensity were effected by means of sector disks, designed especially for J.N.D. work.[1] In brief the set consists of four pairs of disks, with a radius of 19.5 cms., cut from sheet aluminum and painted a dull black. The first pair, giving a range of variation of total open sector from 0°–180° is constructed with a total closed sector of 180°. In the second pair the breadth of total closed sector is 90° and the range of total variation of opening is from 180°–270°. In the third pair the breadth of the closed sectors is 45° and the range of aperture from 270°–315°. In the fourth pair the breadth of the closed sectors is 22.5° and the range of aperture is 315°–337.5°. For J.N.D. work this set is supplemented by single disks of the same breadth of sectors, but of lesser radius. These were made with a radius of 17 cms., thus leaving a margin of 2.5 cms. between the edges of the two sets of disks. The two sets were so adjusted that the edge of the inner disk bisects the stimulus field, thus giving two contiguous fields, one of which was used as standard to which the other was compared. A special protractor,[2] designed for use with these disks, made it possible to measure the open sectors to ¼°.

By using the gelatine filters in combination with these disks, the objection often raised against the use of J.N.D. disks, namely the difficulty of measuring small amounts of change accurately, was to a large extent obviated. Only in a few instances was it necessary to measure less than 1°. The experimental results show in these instances discrepancies that may be ascribed to the errors in measuring. These errors might have been reduced if more filters had been used, but they could not have been foretold until the final calculation of the entire data was complete. To eliminate these experimental errors would necessitate several repetitions of the entire experimental series, and averaging the results.

[1] Jour. Exper. Psych., Vol. I., No. 3, 1916, p. 271.

[2] *Ibid.*, p. 271.

The improvement in the data obtained by such a procedure would hardly justify the long and tedious labor of repeating a J.N.D. series from minimum to maximum for each color several times. Where these discrepancies do occur they are so obvious as not to detract from the value of the general conclusions which are to be derived from these data.

To bridge the gap from one filter to the next, the sectors were set to reduce the amount of light by the same amount as the filter had reduced it. For example, the threshold for red was found with a filter transmission .00000416, sector disk value 33° total open sector. By varying the value of the open sector as was needed, four J.N.D.s in saturation were found requiring a change in open sector to 360°. Filter No. 1 was then removed (transmission .050118) and the value of the open sector reduced until the value of the light transmitted by the sector and the filter No. 4 (transmission .00083) was the same as had been transmitted by filters No. 1 and No. 4 and 360° total open sector. That is, filter transmission multiplied by the sector value = new filter transmission multiplied by the new sector value or

$$.00000416 \times 360 = .000083 \times (X),$$
$$(X) = \frac{.000083}{.0014976}$$
$$= 18°.$$

Thus the total open sector value of 18° with filter No. 4 reproduces the intensity transmitted through filters No. 1 and No. 4 and a total open sector value of 360°. Just noticeable difference steps were again determined until the limit of range with filter No. 4 was reached. Now filters No. 1 and No. 2 were substituted. The equation becomes:–

$$.000083 \times 326° = .0050118 \times (X),$$
$$(X) = \frac{.027058}{.0050118}$$
$$= 54° \; 15'.$$

Settings were made to the nearest quarter of a degree.

III. *The Photometric System.*—The photometric apparat served two purposes in this investigation: (*a*) In one part of the investigation it was decided to make the determinations with the immediately surrounding field of approximately the same brightness as the colored stimulus in order to eliminate the effect of physiological induction in the saturation and hue of the color. For this the photometric apparatus provided both the illumination and the means of controlling the intensity of the illumination needed. (*b*) At the point of maximum saturation both a photometric as well as a radiometric determination of the colored light was wanted, *i.e.*, the radiometric determinations were needed as a basis for the quantitative estimations of the selectiveness of the eye's response and the photometric determination in order to provide a specification of the intensity of maximum saturation in terms commonly used in colorimetry and all practical applications of physiological optics.

In order to secure control of the brightness of the surrounding field it was necessary to provide *as surrounding field a surface of the brightness equaling point by point* the brightness of the colored stimulus as its intensity was changed from threshold to maximum. This was done by selecting for the surrounding field neutral pigment papers approximately graded as to reflection coefficient and by varying the intensity of the light falling on these surfaces by amounts needed to give the brightness match. The selection of the pigment surfaces was made from a series of Hering standard gray papers, graded from 1–50. The needed change of illumination of the surrounding field was secured by means of a lamp moving along a photometer bar. This bar was supported on floor stands with extension uprights to allow of adjustment in height. The range of movement provided by the bar was two meters. The lamps used were a type *C* automobile lamp, 24.73 c.p., operated on a storage battery at .8 amp., and a 100-watt type *C* lamp operated at .85 amp. on the regular line current.

C. Method of Determination.

To render the eye approximately uniformly sensitive a 15-minute period of dark adaptation was taken before each period of work. This condition of sensitivity was maintained throughout the series, as far as compatible with change of intensity of stimulus light, by short exposures (2 to 3 seconds) and frequent rest periods. The method of limits was employed in making the determinations. Several weeks of practice work preceded the final determinations.

As has been stated above, the gross reduction in intensity was made with the neutral filters and the final determinations of the J.N.D. steps were made with the sector disk. The value of the open sector was measured with the vernier protractor and settings were made to the nearest ¼ degree. For the J.N.D. determinations the outside disk was moved so as gradually to increase the value of the open sector until the right half of the field appeared just noticeably more saturated. For the descending series the outside sector disk was opened still further, giving more than a just noticeable difference in saturation, and then gradually decreased until the field appeared uniform. The average of the series of determinations was taken and both sector disks set at this value for the next J.N.D., etc., until a point was reached where the lighter field appeared less saturated. The point of maximum saturation was determined for seven points in the spectrum by six observers and the intensity specified both in radiometric and in photometric terms. The determinations of the J.N.D.s were continued beyond the point of maximum saturation until no further change of saturation with change of intensity could be detected or to the limit of intensity of the apparatus.

The first series (*A*) was made in the dark, with a dark surrounding field, and the second series (*B*) with a surrounding field of approximately the same brightness as the stimulus.

The monochromatic lights used for this investigation were bands from the following regions of the prismatic spectrum: 655 $\mu\mu$, 616 $\mu\mu$, 580 $\mu\mu$, 553 $\mu\mu$, 522 $\mu\mu$, 488 $\mu\mu$, and 463 $\mu\mu$. The wave-lengths transmitted were checked up every 3 or 4 determinations with a Hilger direct vision spectroscope.

Table II. gives the energy density of these lights at the analyzing slit and at the eye.

TABLE II

Wave-length	Energy per Square Mm. in Watt $\times 10^{-8}$	
	At Analyzing Slit	At the Eye
Red 655 $\mu\mu$	7656.81874	23.8776
Orange 616 $\mu\mu$	2679.88621	8.537
Yellow 580 $\mu\mu$	1196.5544	3.731
Y. Gr. 553 $\mu\mu$	348.46526	1.0867
Green 522 $\mu\mu$	287.24839	.89544
Bl. Gr. 488 $\mu\mu$	141.2697	.41405
Blue 463 $\mu\mu$	131.85172	.411

The energy value for each J.N.D. step was obtained by reducing the total energy for the maximum intensity at the eye to the fraction of the total light transmitted to the eye, multiplying first by the per cent. transmission through the neutral filters and finally by the per cent. transmission through the open sector for each step.

D. Discussion of Results

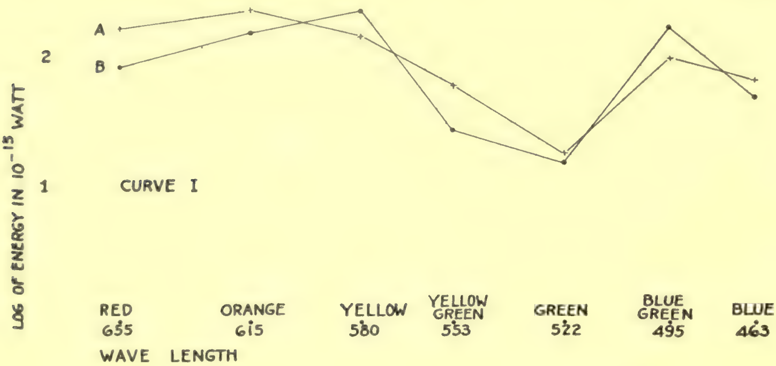
The selectiveness of the eye's chromatic response to wave-length and its change with change of intensity of light may be demonstrated by plotting the amounts of energy required to arouse equal amounts of chromatic sensation for different groups of wave-lengths. That is, we plot the amounts of energy required to arouse the threshold, the fifth, the tenth, etc., J.N.D. to the maximum. At any one of these cross-section points, the relation between the energy quantities for the different wave-lengths demonstrates the selectiveness of the eye to wave-length at that point in the intensity scale. The manner in which this relation changes with increase of intensity indicates how the eye changes its selectiveness with change of intensity. In Table III. is given the energy of

TABLE III

ENERGY OF CHROMATIC THRESHOLD IN 10^{-15} WATT

	Dark	Light
Red	173.34	91.53
Orange	239.90	158.78
Yellow	154.84	227.96
Yellow-green	63.81	29.77
Green	18.59	16.20
Blue-green	96.71	172.76
Blue	69.50	51.05

the threshold of chromatic sensation at seven points in the spectrum. The results are given for experiments conducted in the dark room as well as for those with the surrounding field of approximately the same brightness as the stimulus. From these data have been constructed Curve I., *A* and *B*.



Here the writer has plotted the log of the energy at the color threshold against wave-length. Inspection of the curves will show that the eye is generally more sensitive under the condition with the light surrounding field than in complete darkness, except in the case of yellow and blue-green. This increased sensitivity with the light surrounding field is due to the absence, under these conditions, of the induced white light over the color field, obviating the inhibitive action of the achromatic on the chromatic.

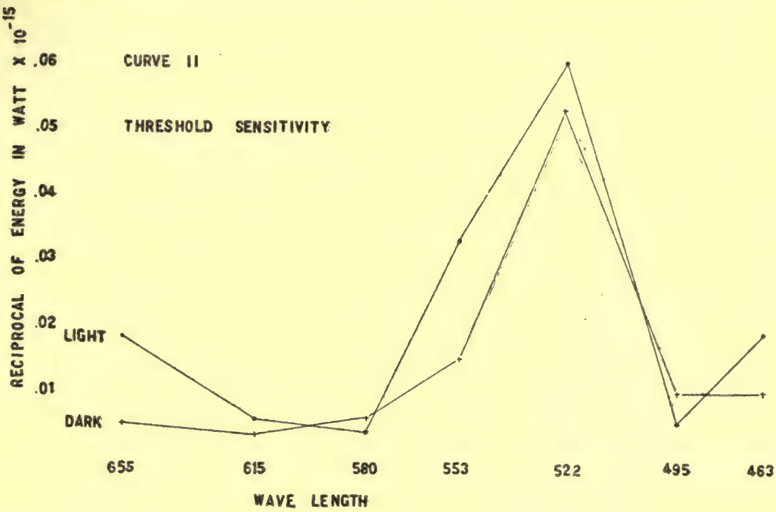
In Table IV. are given the values indicating the color sensitivity of the eye to wave-length at the limen as expressed by the reciprocal of the energy required to arouse the chro-

Table IV

Chromatic Sensitivity at Limen

	Dark	Light
Red	.00576	.01092
Orange	.00417	.00629
Yellow	.00646	.00438
Yellow-gr.	.01567	.03359
Green	.05378	.06173
Blue-gr.	.01034	.00579
Blue	.01431	.01959

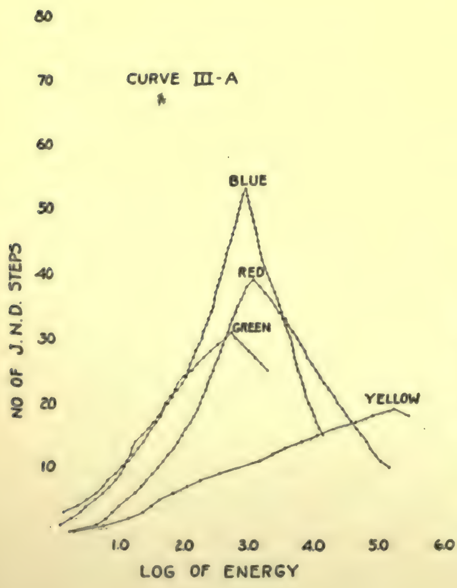
matic threshold. These values are plotted in Curve II., *A* and *B*. It is apparent that the sensitivity at this intensity



is greatest for green. This condition obtains throughout the intensity series except at high intensities in the dark, where the values for blue indicate that the sensitivity has shifted towards the shorter wave-lengths. In this connection it may be interesting to mention a fact, observed not only by the writer but also by others in conducting research with spectrum lights, namely, that the effect on the observer of green light from the region of the spectrum around 522 $\mu\mu$ is extremely unpleasant. The writer experienced great discomfort, headaches, watering of the eye, ocular hyperæmia, excessive flow of lymph across the field of vision, especially at moderate and high intensity. The doctrine which is so prevalent that green light is most pleasant for the eye is certainly false as far as spectrum light is concerned. Some discomfort was experienced in working with blue at very high intensities in the dark, but the effect was not nearly so marked.

Table V., *A* and *B*, is so arranged as to give the energy value of each sensation step for the four primary colors. An inspection of these tables enables us to trace the energy value of each J.N.D. as the intensity is increased. We can pick out the energy required to arouse the fifth, tenth, fifteenth just noticeable difference in saturation for each of the four colors, showing how the selectiveness of the eye's chromatic response changes with change of intensity.

Curve III., *A* and *B*, has been constructed from the data contained in Table V. The writer has plotted the log of the energy required for each J.N.D. of saturation. The number of J.N.D.s are plotted along the ordinate and the log of the energy along the abscissa. Thus, for example, in Curve III., *A*, at the fifth J.N.D. yellow has the highest energy value, red next, then blue and green least, showing that the eye has its greatest sensitivity in the green and the lowest in the yellow at this point in the intensity scale. The horizontal difference between the four curves at any point in the sensation scale indicates the relative sensitivity to wave-length. That is, the eye shows its greatest difference in sensitivity between the red and the yellow. To green and blue the eye is more equally sensitive. As the eye changes its selectiveness to wave-length with change of intensity, the horizontal difference between the curves changes. Beyond



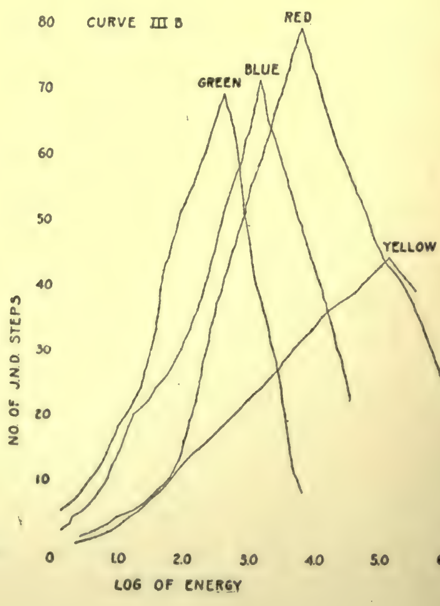


Table V

Dark Surrounding Field

J.N.D.	Red	Yellow	Green	Blue
	Energy in 10^{-13} Watt	Energy in 10^{-13} Watt	Energy in 10^{-13} Watt	Energy in 10^{-13} Watt
1	1.73	1.55	.19	.70
2	4.34	5.69	.31	1.16
3	6.11	14.01	.69	1.80
4	7.85	23.14	1.34	2.52
5	9.81	32.06	2.30	3.06
6	13.58	42.87	3.29	4.30
7	17.91	71.90	4.46	5.57
8	22.57	116.60	5.52	7.05
9	27.21	175.34	6.74	8.90
10	34.19	350.06	9.00	10.41
11	41.30	726.50	11.09	11.91
12	49.51	1,542.47	14.02	13.30
13	59.53	2,389.42	16.88	15.74
14	73.09	3,653.76	20.36	16.50
15	85.77	6,718.89	24.07	18.18
16	97.96	11,956.99	28.36	23.90
17	111.40	21,460.22	33.02	30.16
18	130.79	44,560.22	38.61	37.00
19	149.75	83,670.32	44.70	43.82
20	166.63	M176,167.00	50.96	50.64
21	185.13	315,813.97	58.37	58.62
22	203.05		67.49	66.58
23	222.75		82.41	75.68
24	244.08		97.02	84.52
25	270.21		116.90	97.11
26	299.75		140.55	103.98
27	327.50		172.52	120.67
28	362.00		213.67	136.03
29	393.35		268.20	150.24
30	425.79		354.91	165.12
31	463.43		460.44	180.27
32	508.99		M565.25	195.77
33	560.47		669.52	214.55
34	601.14		819.52	236.17
35	660.09		1,054.92	255.53
36	727.18		1,291.35	276.57
37	815.58		1,609.02	296.52
38	917.25		2,086.15	318.11
39	1,030.23			337.98
40	M1,290.93			357.99
41	1,605.73			390.04
42	1,964.15			419.33
43	2,357.93			447.67
44	2,751.00			475.11
45	3,209.14			510.50
46	3,701.95			544.80
47	4,158.17			578.98
48	5,010.59			616.05
49	5,798.14			651.23
50	6,550.23			694.23

TABLE V (*Continued*)

J.N.D.	Red	Yellow	Green	Blue
	Energy in 10^{-18} Watt	Energy in 10^{-13} Watt	Energy in 10^{-18} Watt	Energy in 10^{-18} Watt
51	7,597.24			741.32
52	8,580.50			791.00
53	9,879.55			849.94
54	11,420.18			907.03
55	13,275.93			M955.50
56	15,524.21			1,015.47
57	17,915.53			1,075.32
58	20,634.21			1,129.43
59	23,579.28			1,198.02
60	28,995.57			1,260.65
61	33,642.96			1,323.49
62	39,208.40			1,391.87
63	45,728.99			1,466.11
64	53,908.61			1,544.54
65	64,680.92			1,625.76
66	75,794.28			1,702.14
67	90,659.13			1,933.95
68	108,464.03			2,106.03
69				2,276.48
70				2,446.12
71				2,675.56
72				2,900.92
73				3,130.37
74				3,359.82
75				3,585.17
76				3,869.93
77				4,154.30
78				4,496.12
79				4,779.70
80				5,064.31
81				5,351.13
82				5,691.20
83				6,032.51
84				6,490.21
85				6,945.09
86				7,508.56
87				8,078.19
88				8,646.12
89				9,555.37
90				10,468.72
91				11,554.50
92				12,701.75
93				14,216.71

TABLE V (*Continued*)

LIGHT SURROUNDING FIELD

J.N.D.	Red	Yellow	Green	Blue
	Energy in 10^{-13} Watt	Energy in 10^{-13} Watt	Energy in 10^{-13} Watt	Energy in 10^{-13} Watt
1	.92	2.28	.16	.51
2	2.74	4.81	.29	.95
3	4.67	8.25	.44	1.40
4	7.25	11.65	.75	1.96
5	9.79	15.64	1.06	2.11
6	17.12	21.17	1.41	3.16
7	24.19	26.30	1.89	3.85
8	31.52	34.46	2.46	4.63
9	39.37	44.34	2.96	5.48
10	51.64	53.87	3.48	6.34
11	62.00	65.87	4.15	7.09
12	69.67	79.99	4.89	7.86
13	78.28	98.31	5.55	8.64
14	87.85	120.00	6.37	9.62
15	94.31	146.56	6.74	10.65
16	101.39	186.20	7.72	11.58
17	107.65	240.52	8.83	12.68
18	115.00	303.24	9.84	13.84
19	123.69	376.41	11.02	15.15
20	131.28	486.31	12.52	16.17
21	140.46	601.12	14.09	18.02
22	147.61	742.28	16.02	23.76
23	156.19	903.91	17.88	29.08
24	163.09	1,118.69	20.26	34.13
25	169.77	1,452.93	22.46	39.24
26	177.78	1,882.02	24.46	47.77
27	185.09	2,270.46	25.55	56.31
28	193.23	2,703.06	26.97	65.53
29	203.07	3,178.82	28.47	73.73
30	212.87	3,843.79	30.00	85.15
31	224.41	4,627.90	31.74	92.15
32	235.82	5,609.35	33.26	101.98
33	247.26	7,083.91	35.02	112.06
34	262.01	8,600.84	36.95	120.25
35	276.75	10,502.55	38.97	134.34
36	291.44	12,406.96	39.95	147.04
37	307.85	14,823.88	41.21	159.74
38	325.86	22,960.67	42.48	172.02
39	349.03	31,354.36	43.67	187.56
40	370.08	40,764.68	45.08	199.88
41	394.75	51,227.62	46.35	212.98
42	422.54	64,309.76	48.36	226.80
43	448.71	79,458.95	49.57	240.84
44	481.23	100,914.09	52.62	253.94
45	520.74	128,617.36	58.13	269.51
46	560.00	M155,793.75	61.11	283.43
47	599.39	188,202.15	65.58	298.28
48	643.48	225,813.09	70.04	317.73
49	670.14	264,547.70	74.52	333.50
50	715.68	316,281.72	78.99	349.48

TABLE V (*Continued*)

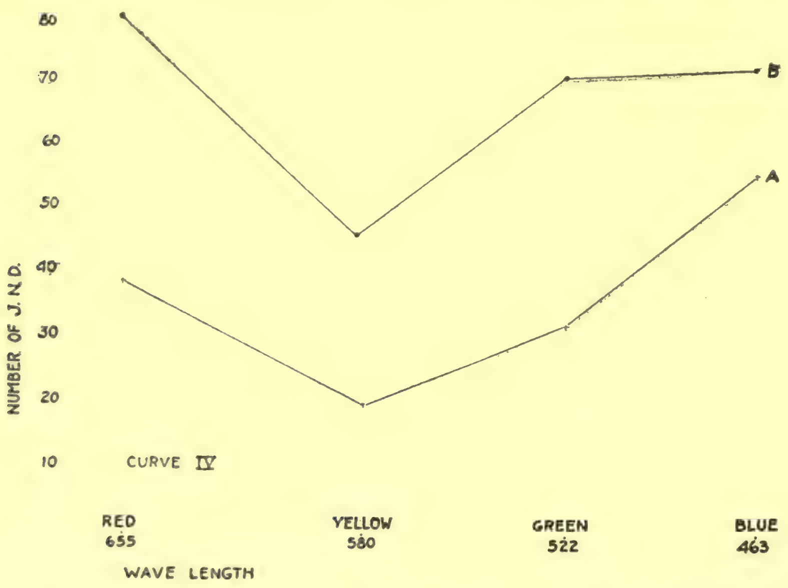
J.N.D.	Red	Yellow	Green	Blue
	Energy in 10^{-13} Watt	Energy in 10^{-13} Watt	Energy in 10^{-13} Watt	Energy in 10^{-13} Watt
51..........	763.60		83.67	368.84
52..........	817.44		90.91	395.74
53..........	856.08		99.85	426.62
54..........	912.04		108.80	460.76
55..........	962.02		119.23	495.58
56..........	1,015.34		130.69	531.33
57..........	1,113.36		141.63	570.71
58..........	1,211.68		155.04	610.81
59..........	1,310.10		169.85	655.30
60..........	1,407.40		184.80	692.19
61..........	1,506.72		199.68	725.06
62..........	1,588.90		213.09	767.67
63..........	1,703.13		228.01	822.01
64..........	1,838.81		244.42	873.21
65..........	1,965.09		265.24	925.64
66..........	2,096.14		286.16	970.71
67..........	2,227.06		307.00	1,028.03
68..........	2,390.65		326.40	1,072.10
69..........	2,387.36		350.10	1,149.37
70..........	2,784.12		M377.06	1,206.30
71..........	2,980.40		402.34	1,265.60
72..........	3,165.60		432.19	M1,325.90
73..........	3,379.45		463.63	1,400.74
74..........	3,635.93		490.47	1,474.47
75..........	3,895.14		520.18	1,544.21
76..........	4,159.38		545.43	1,622.01
77..........	4,520.15		570.00	1,691.54
78..........	4,847.89		591.53	1,757.08
79..........	5,240.25		616.38	1,933.30
80..........	5,666.09		652.79	2,105.21
81..........	M6,123.83		670.00	2,277.23
82..........	6,515.25		696.04	2,445.28
83..........	7,040.77		719.86	2,617.19
84..........	7,538.74		740.67	2,780.36
85..........	8,057.04		758.51	3,015.53
86..........	8,646.52		772.77	3,236.84
87..........	9,335.03		804.73	3,416.74
88..........	9,988.18		831.55	3,593.14
89..........	10,742.72		869.88	3,981.91
90..........	11,452.53		895.71	4,318.72
91..........	12,275.58		927.01	4,665.06
92..........	13,135.09		955.26	4,997.53
93..........	14,279.61		983.60	5,298.20
94..........	15,394.65		1,020.83	5,821.32
95..........	16,408.82		1,065.58	6,308.01
96..........	17,455.74		1,103.02	6,770.26
97..........	18,799.76		1,162.03	7,283.35
98..........	20,007.62		1,219.04	7,791.17
99..........	22,447.47		1,311.18	8,363.51
100..........	25,218.74		1,400.93	8,932.92

TABLE V (*Continued*)

J.N.D.	Red — Energy in 10^{-13} Watt	Yellow — Energy in 10^{-13} Watt	Green — Energy in 10^{-13} Watt	Blue — Energy in 10^{-13} Watt
101	28,177.02		1,490.36	9,612.72
102	31,431.17		1,586.92	10,351.59
103	34,572.47		1,698.46	11,142.32
104	38,151.78		1,811.18	12,063.03
105	42,089.00		1,915.26	12,913.88
106	46,014.95		2,034.36	13,823.24
107	50,271.01		2,160.97	14,740.86
108	55,634.02		2,317.91	15,700.05
109	60,103.57		2,437.15	16,780.96
110	64,836.61		2,556.06	17,755.16
111	70,572.77		2,617.06	18,885.48
112	77,594.89		2,742.52	20,025.27
113	84,170.94		2,845.53	21,166.82
114	93,515.40		2,970.08	22,482.10
115	103,488.85		3,107.36	23,780.96
116	124,497.41		3,241.24	25,147.87
117	150,647.98		3,353.05	26,565.11
118	176,844.56		3,433.52	28,216.77
119	202,222.60		3,613.77	29,944.80
120	232,534.08		3,740.63	31,524.47
121	262,012.88		3,845.54	33,235.75
123	291,439.82		3,969.25	34,881.84
124	320,913.92		4,121.30	
125	356,874.44		4,292.22	
126	392,013.86		4,456.04	
127	429,024.88		4,620.42	
128	471,588.63		4,798.35	
129	517,459.79		5,514.02	
130	563,308.85		5,961.45	
131	619,003.09		6,409.53	
132	687,807.41			
133	756,541.00			
134	825,274.62			
135	913,690.50			
136	1,002,119.13			
137	1,091,038.65			
138				
139				
140				

the 25th J.N.D. step the eye becomes less sensitive to green than to blue. This may be due to the greater inhibitive action of the achromatic on the green. It may be noted also that red and green retain their characteristic colors to very high intensities, as indicated by the slope of the curves beyond the maximum. Blue loses its saturation most rapidly, it has the steepest descending curve. The curves of yellow and green appear to have very nearly the same inclination. The trend of these results is in agreement with observations made by Helmholtz and others on this point.

From the inspection of these curves we can also see the number of J.N.D.s required to reach the maximum. With the dark surrounding field blue requires 55, red 40, green 32, and yellow 21. With the light surrounding field blue requires 71, red 80, green 70 and yellow 46. If the number of J.N.D.s required to reach the point of maximum saturation may be used as an indication of the degree of saturation at its maximum, as Geissler has suggested, then, in the dark, blue would be the most saturated color, whereas with the light surrounding field red would be. To demonstrate more directly the relative degrees of saturation of the four colors under the two experimental conditions Curve IV., *A* and *B*,



has been constructed. Here the writer has plotted the number of J.N.D.s required to reach the maximum against wavelength. The results are in accordance with the writer's estimate by direct inspection of the rank under the two conditions. That is, blue with the dark surrounding field actually appears more saturated at its point of maximum saturation than the other colors. With the light surrounding field red appears the most saturated. Yellow has the least saturation under both conditions.

In considering Curve III., *B*, we find that the general trend of the results is similar. The most significant difference is the greater chromatic sensitivity under these conditions, as indicated by the greater number of J.N.D. required to reach the maximum. Furthermore, although the order of sensitivity remains the same (that is, green highest, then blue, red, yellow), the difference between the colors is greater with the light surrounding field than with the dark. This is shown by the horizontal difference, representing the energy value between the points of maximum saturation for the four colors.

That this type of selectiveness of chromatic response to wave-length is characteristic of the eye and not merely an individual peculiarity of response is shown by Curve V., *A* and *B*. This Curve is constructed from data given in Table VI. In the table are given the values of the energy

TABLE VI

ENERGY OF POINT OF MAXIMUM SATURATION IN 10^{-13} WATT

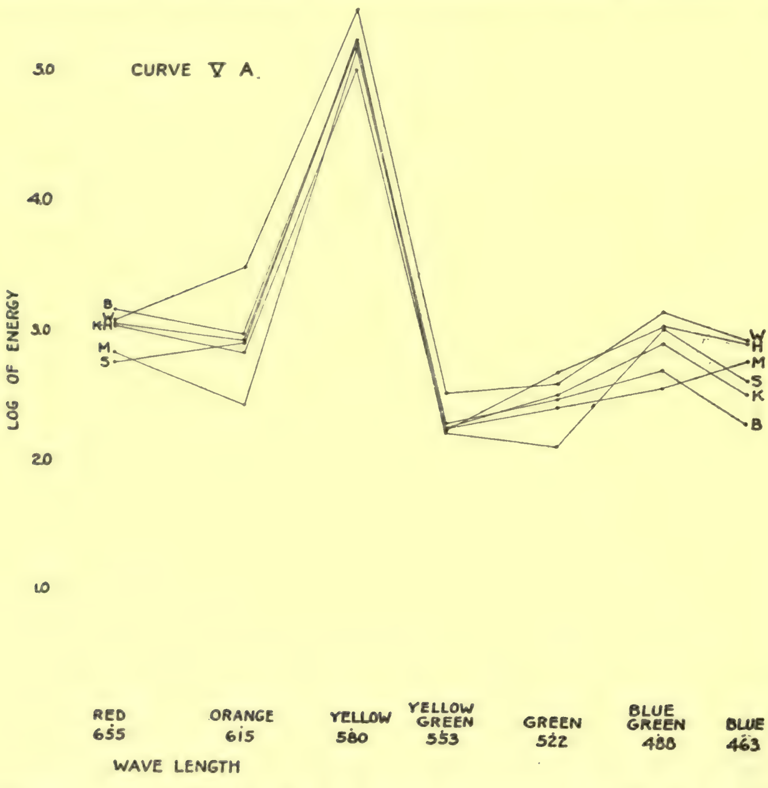
Dark Surrounding Field

	Red	Orange	Yellow	Y-Green	Green	Bl-Green	Blue
S......	628.84	881.40	184,673.16	189.68	156.46	1,291.74	503.97
W......	1,318.07	3,372.17	324,844.52	378.11	452.00	1,648.77	1,024.34
M......	756.38	301.21	180,018.37	203.35	239.05	439.67	714.17
B......	1,588.80	1,042.56	177,932.51	216.91	351.42	595.31	233.55
H......	1,290.93	938.31	174,321.18	198.80	565.25	1,257.46	955.50
K......	1,191.51	753.01	114,848.20	207.90	368.78	946.61	397.45

Light Surrounding Field

	Red	Orange	Yellow	Y-Green	Green	Bl-Green	Blue
S......	3,607.89	1,413.39	185,180.37	366.02	223.50	1,529.62	756.78
W......	5,614.37	6,302.12		55.73	186.27	415.27	515.03
M......	403.91	521.28	177,932.51	222.98	264.53	519.12	614.43
B......	1,986.00	684.13	177,932.51	210.94	273.19	635.15	381.05
H......	6,123.38	556.04	154,132.61	81.33	377.06	1,312.85	1,325.90
K......	463.48	729.77	180,018.37	96.52	149.03	1,416.84	620.34

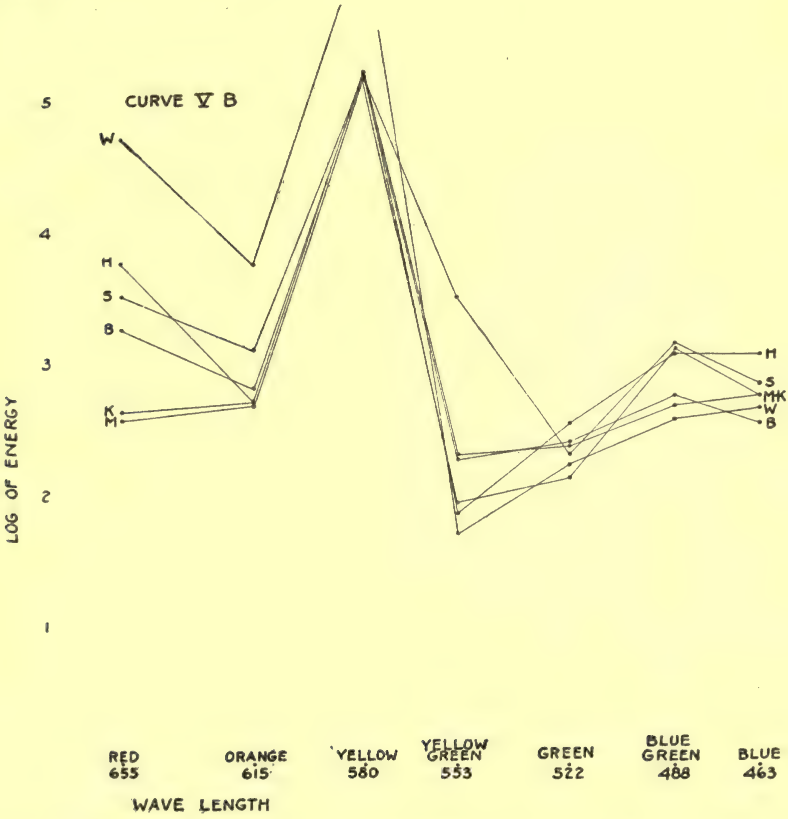
at the point of maximum saturation for each of the seven points in the spectrum for six observers. Curve V. was plotted using the log of the energy as the ordinate against the wave-length along the abscissa. The agreement in the results for the six observers is encouraging. It is closer with the dark surrounding field than with the light surrounding field. For some observers the judgment was easier under this condition. Observer *W*, who, as may be noted in the table, has the highest energy values in the yellow and orange, conspicuously different from the others, may be classed as



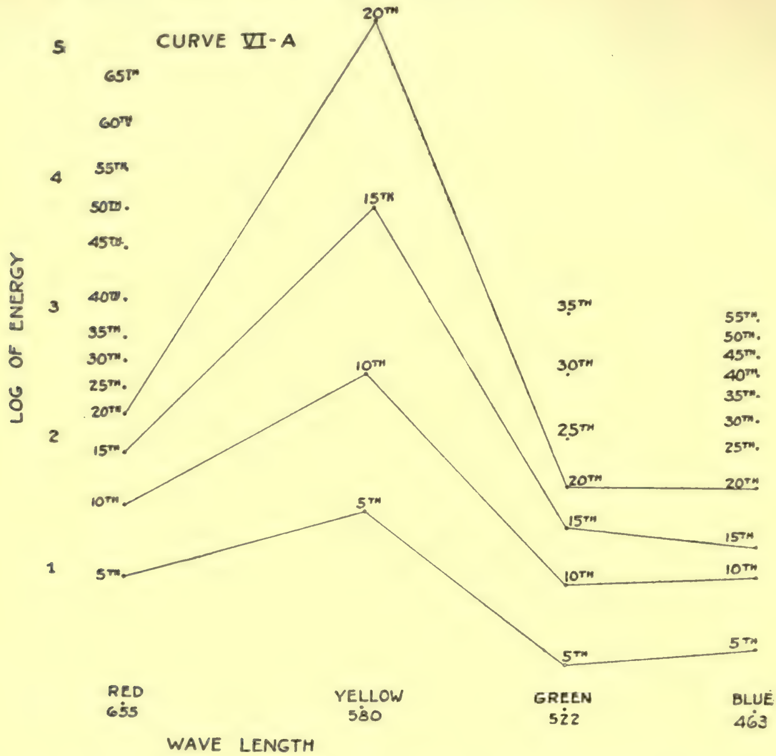
erratic. She complained of being seriously hampered in her judgment of color saturation by the vividness and persistence of after-images. No such difficulty was reported by the other observers. The writer's results are included as observer *H*.

In order to show the change of selectiveness of the eye's response to wave-length with change of intensity we have plotted Curve VI., from the data contained in Table V.

Along the ordinate we have plotted the log of the energy required to arouse the chromatic threshold and every fifth J.N.D. thereafter, for each of the four wave-lengths investigated (abscissa). By inspection of *B*, in the light, we note

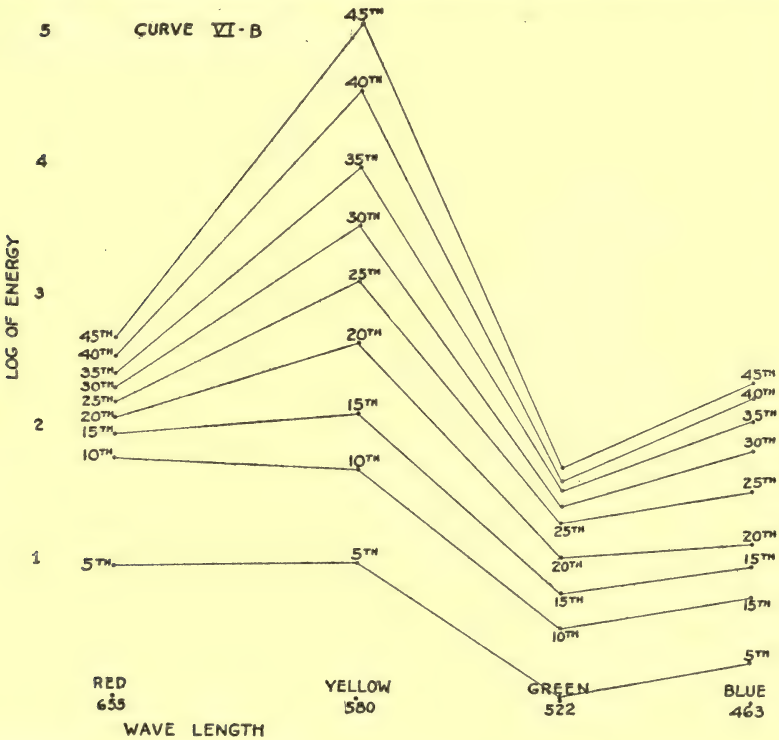


that the change in the eye's selectiveness to wave-length with change of intensity is most noticeable in its response to yellow. A marked difference is noted also in the slope of the lines between the values for green and blue, as the intensity increases, particularly between the 20 and 30 J.N.D. This demonstrates again that the eye loses its sensitivity to blue more rapidly than to green, but more slowly to the shorter wave-lengths than to the long. It is interesting to note, with the dark surrounding field, *A*, after the 10th J.N.D. the relative sensitivity to blue and green is actually reversed. At higher intensities the eye becomes more sensitive to blue than to green.



In Table VII. are given the photometric specifications of the intensity at the point of maximum saturation. The photometric match was made by the equality of brightness method. The colored stimulus light was set at the intensity to give maximum saturation. A magnesium oxide screen was inserted immediately behind the stimulus opening, dividing the field and at such an angle that the light from the standard photometer lamp fell normally upon the photometric surface. The surrounding field was a screen of Hering White. The standard lamp was moved along the photometer bar to give the brightness match.

In the course of the experiments it was observed that as the intensity is increased, the chromatic sensation aroused changes both in hue and in saturation. A knowledge of these hue changes is of considerable importance in colorimetry. For example, it is often considered sufficient to give a colorimetric specification of lights at one intensity alone, in spite

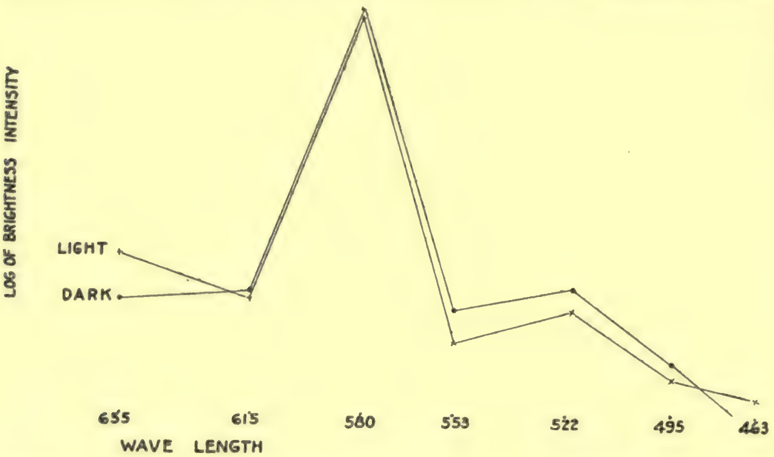


of the fact that both the saturation and the hue of the color changes with change of intensity as well as with the composition. Changes of hue simulating an apparent change in composition of light take place when there is no change in composition, only change in intensity. That is, while specification of the composition of light is independent of intensity, a true colorimetric specification may not, depending upon the method used, be definitive unless it is accompanied by a specification of intensity. In determining the J.N.D. as the intensity of light was increased, the significant changes in hue have been noted, and the points at which they have occurred are specified radiometrically. It was found that in the work with the dark surrounding field the just noticeable difference for hue is larger than for saturation, hence it was

TABLE VII

PHOTOMETRIC SPECIFICATION OF MAXIMUM IN M.C.

	Red	Orange	Yellow	Y-Green	Green	B-Green	Blue
Dark surrounding.....	7.66	8.79	1,318.45	6.39	9.49	2.62	.95
Light surrounding.....	16.87	7.86	1,162.88	3.47	6.21	2.91	1.30



impossible in a progressive series of saturation steps to indicate the exact point in the series where the hue change took place. It is important for all colorimetric interpretation to specify the change under both conditions, because the changes in hue experienced with a light surrounding field are different from those observed in the dark. A still different series of changes is experienced under light adaptation with change of intensity when the surrounding field is lighter or darker than the color field. It was found that at the threshold, with the light surrounding field red appeared as a good red, but at the 10th J.N.D. appeared brownish, at the 30th, brownish-red, not attaining a good red until the 50th J.N.D. At the maximum (80 J.N.D.) the hue may be described as bright scarlet, rapidly acquiring an orange tint (82 J.N.D.), passing through orange pink as the intensity was increased.

Yellow appears olive-green at the threshold, distinctly green at the 6th J.N.D., yellow-green at the 10th, greenish-yellow at the 18th, with still a tinge of green at the 26th, attaining a good yellow at the 36th, which faded into a whitish-yellow beyond the maximum. Green does not manifest such striking changes of hue with change of intensity. It appears blue-green at the threshold, attains a pure green around the 50th J.N.D., becoming progressively more yellow beyond the maximum. Blue comes in as a purple at the threshold, retains a tinge of this even to the 49th J.N.D., then becoming pure blue, and manifesting no hue changes as it passes into white beyond the maximum.

To summarize, then, the data presented for showing the selectiveness of the eye's chromatic response to wave-length and its change with change of intensity, it may be stated that at the threshold was found that the eye is most sensitive to color in the following order (*A*) with the dark surrounding field: green, yellow-green, blue, blue-green, yellow, red and orange. (*B*) With the light surrounding field the order became: blue-green, green, blue, yellow-green, red, yellow and orange. At the point of maximum saturation, with the dark surrounding field the eye was most sensitive to yellow-green, then to green, orange, blue, blue-green, red and finally yellow. With the light surrounding field at the maximum, the sensitivity decreases in the following order: yellow-green, green, orange, blue-green, blue, red and yellow.

NOTE: The Editors alone are responsible for the acceptance of this article, Professor Ferree having advised against its publication.

321

Journal of

Experimental Psychology

Vol. V., No. 6 December, 1922

PERSONALITY STUDIES OF THREE-YEAR-OLDS

BY HELEN T. WOOLLEY, PH.D.[1]

Introduction

There is no need to expatiate upon the lack of scientific information about the mental life of small children. The fact is only too well known. With our growing belief that the experiences of early childhood have a powerful and permanent effect on character and personality, the lack of knowledge about the mental capacities and types of reaction of very young children seems an even more serious omission. Furthermore, we are becoming more and more disturbed by the fact that the care and training of children up to five years is still exclusively in the hands of mothers who come, without specific training, to a task so important that it may determine the success or failure of their children before the organized agencies of the community for the training of children touch them at all. So keen is the realization becoming of the danger of making no provision for the study or training of very young children, either directly or through the training of mothers, that it seems probable that the next large new development of the educational system may take the form of some type of provision for this lack.

If training for all young children, or for all mothers, potential or actual, is to be provided by the community, it is obvious that the public-school system is the only agency adequate to the task. It is possible for private agencies, however, to carry out experiments which may demonstrate

[1] The physical examinations referred to in this paper were made by Dr. C. A. Wilson of Detroit. The laboratory tests of blood and urine were made by Dr. Paul G. Woolley at the National Pathological Laboratories. The mental examinations were given by Miss Elizabeth Cleveland, working at the school laboratory under the direction of the writer. The records of behavior were made by the college students, and various members of the staff who came in contact with the children.

both the value and the practicability of new educational ventures. The Merrill-Palmer School of Detroit is undertaking an experiment in which the two purposes of providing suitable training for young children, and of developing courses of training in the care and management of young children for prospective mothers are made to supplement one another. A nursery school for children between a year and a half and five years of age has been established, and a course of training in the care and management of young children for college students. The college girls who are in residence get their practice work in health and nutrition, and in the psychology and training of young children by acting as assistants in the nursery school. The children in the school are in immediate charge of Miss Emma Henton, from the Gipsy Hill Training College in London. Miss Henton was in charge of the nursery school established at the college under the provisions of the Fisher Act.

Under the plan of organization, each of the thirty children in the school becomes the subject of as careful a study of personality as we find it possible to make. Every child has a complete physical examination, and as thorough a mental examination as we can make with the imperfect tools now at command. In each case a personal history of the child, and a brief family history is recorded. Continuous records of height, weight, feeding, and types of reaction under varied conditions are kept. Needless to say each child stands out as a personality as distinct to those who take the pains to know him well, as adult acquaintances, though we hope not so immutable.

The following studies of two of the three-year-old children in the school are offered first for publication because they seem to us interesting examples of the kind of experience in young childhood which is frequently held responsible for warped, or even pathological types of personality later on.

Mary Jane

Case of Phobia in a Three-year-old Child

Age: Mary Jane entered the school in January, 1922, when she was three years and eight months old.

Physical Condition: Mary Jane's physical condition was perfect. The doctor rated her in the highest class for general condition. Her standards of height and weight were both above those of her age. The Wassermann test was negative and blood count and urine analysis showed no deviations from normal.

Mental Development: On the Stanford Revision of the Binet scale Mary Jane had an intelligence quotient of 133. Her rating on a series of performance tests was in the one hundredth percentile for her age in two of them, in the ninetieth percentile in a third, and in the sixtieth percentile in a fourth. These performance tests were standardized in the laboratory of the Merrill-Palmer School, and will be published soon.

Appearance: Mary Jane is a large, very healthy looking child with fair curls, blue eyes, a round rosy face, and a winning smile that frequently displays her white teeth.

Character and Disposition: Her most outstanding characteristic was her constant happiness, good temper and coöperative spirit. There was never a conflict of any sort with Mary Jane at the school. She was always ready to help with the work and join in the plays. In the little tasks of the school she was competent and reliable. She always knew exactly what she wanted to do when the choice of work was left to her and she stuck to one task until it was finished. Her father reported that she was frequently stubborn at home, but, although she was so very positive in her plans and ideas, we did not find her stubborn at school. She sometimes failed to obey immediately but was never really resistant to reasonable demands. Mary Jane showed signs of being in a certain sense rather nervous. In telling a story she frequently trembled with excitement and her hands shook. It was noticeable that her hands shook when threading beads, but she was never irritable. Her control of her body was excellent. She had a good sense of rhythm and could march or skip to music more accurately than most of the children.

Home Background: The child is of Norwegian descent on both sides, though she herself and her mother were born in this country and her father came here as a child. Her father is a doctor and her mother was a trained nurse before her marriage. Her grandparents were merchants and farmers. At the time of her entrance, Mary Jane was a first and an only child. Her parents are exceedingly interested in Mary Jane, and devoted to her, and on the whole wise in their management.

History of the Development of the Phobia: On March 7, when Mary Jane was three years and ten months old and had been about six weeks in the school, she stood at the top of a flight of stairs one day after the other children had gone down and began to cry. When one of the college students went up to find out what was the matter she insisted that she was afraid to come down the stairs and must have someone take her hand and walk next to the railing. Her eyes looked panic stricken and she trembled. When asked what she was afraid of she pointed to the two spindles of the railing nearest the top post. The only reason she could give for being afraid was that there wasn't any curtain. The child had been up and down these same stairs several times a day every day since entering the school and we were totally at a loss to account for this sudden intense fear. Much to our surprise the fear lasted for about four weeks. Day after day Mary Jane stood at the top after the other children had gone down, full of fear and distress and totally unable to force herself to descend the familiar straircase. She kept repeating amid her tears—"I can come up all right. I am not afraid to come up, but I can't go down. I am afraid to go down." The genuineness of her distress and her inability to understand what had caused it seemed unmistakable to us. We did everything we could think of to overcome it. We talked to the child about the railing and told her it was there to help her and keep her from falling. We finally got her to touch it herself, though it was evidently a fearful ordeal, and gradually by coming down with her and urging and encouraging her we got her to the point where she could come down with comparatively little trepidation. Her mother was totally at a loss to account for the situation. During this period she displayed the same type of fear of a similar banister which she encountered in a department store when her mother was with her. At the end of about four weeks, when we felt that her fear was in a large measure conquered, a vacation of over a week occurred. Since that time the child has apparently completely overcome the trouble.

We were totally at a loss to explain the situation. Our only theory at first was that something had happened to frighten or disturb her when she happened to be looking at that part of the banister, but since she could not explain it and no one was with her at the moment, we were somewhat hopeless about finding out what had really happened. It then occurred to us that possibly the history of the development of phobias in adults might lend some clue this case. In adults the condition seems to be connected with suppressed ideas related to sex. We then recalled the fact that about two weeks after Mary Jane entered the school, on February 7, a baby brother had been born. Mary Jane had talked about the baby at school most affectionately and had displayed a great interest in him. Her mother reported that she seemed delighted with him and had displayed no trace of jealousy. One day when I was in the nursery she came up to me and began talking about the birth of the baby. She said, "When mother went to the hospital she didn't have the baby and when she came back she had it." She repeated this several times with a rather strained look on her face and I felt sure that the whole problem of the baby's birth had worried her. It suddenly occurred to us that her fear of the stairs might be in some way connected with the event of the baby's birth and her experiences in connection with it. Accordingly I sent for her mother to inquire the details. Her mother told me that the baby was born in a hospital and that Mary Jane, on her way to school, was brought to the hospital by her father to see the baby each morning. I inquired whether she went up or down stairs at the hospital and her mother replied that she never walked up stairs because they always came up in an elevator, but that they frequently walked down stairs after the visit because the elevator was not at hand. I inquired if she knew what kind of a staircase it was and she said that she did. The stairs were visible when her door was opened and she could frequently see Mary Jane and her father start down them. She described a staircase with a short flight of five or six steps at the top and a banister with upright posts at the side. This was exactly the situation with our stairs at the school. I was at once struck with the remembrance that Mary Jane had kept saying, "I can come up all right but I am afraid to go down." I also asked the mother whether they had talked with Mary Jane at home about the arrival of the baby. She said they had not mentioned it to her because they thought she was too young to understand about it. She and her husband had several times mentioned the fact that Mary Jane had never asked any questions about the baby's arrival. This set of circumstances seem to me conclusive in proving that the disturbance was due to her concern about the birth of the baby, the fact that in some way she had been made to feel that it was a topic on which she couldn't ask questions at home, and her continued anxiety about it. On coming out of her mother's room, with her mind full of anxiety and distress about the problem of the arrival of the baby which she did not dare voice, the first object which met her gaze was the staircase with its short flight of steps at the top and the upright posts of the banister. Her problem found no solution, but kept on distressing her in secret until it forced some sort of expression. Finding herself again in a situation so similar to that identified with her suppressed anxiety, suddenly and unaccountably to herself, gave rise to her experience of fear. I explained the matter to her mother who promised to talk to Mary Jane at once about the birth of her little brother, and to try to forestall possible future effects from this very vivid experience.

This is the second case of phobia in a young child which I have personally encountered. The other one I did not succeed in analyzing. A physician in Detroit who is a specialist in children's diseases told me that he had met with five or six cases of phobia in his practice among children about three or four years of age. He seemed to have no theory to explain the phenomenon. We were particularly interested in being able to unravel at the time the type of experience in young childhood which is frequently held accountable for the development of mental abnormalities in later years. Very few individuals remember definitely events that occur between three and four and are therefore frequently quite unable to find the clue to their own difficulties without outside aid.

Bobbie

A Case of Mother Complex

Age: On entering school in January, 1922, Bobbie was three years and six months old.

Physical Condition: Bobbie's physical condition was only good. He was above the standard of height and weight for his age, but was rated only 2 by the doctor. He had a previous rachitic history and had had one convulsion before his entrance to the school. During a school vacation, after he had entered, he had another convulsion which terrified his parents. They sent him to the Ford Hospital for a complete physical examination. The doctor reported that he found signs of a previous encephalitis. His Wassermann test was negative and the blood count and urine examination showed no abnormalities.

Mental Condition: On the Stanford Revision of the Binet Scale Bobbie had an intelligence quotient of 105. He failed in two of the four performance tests but since so many of the three-year-olds failed in these tests, this had no consequence other than to show that he was not very superior. In two of the performance tests he succeeded with ratings in the 60 and in the 80 percentile.

Appearance: Bobbie is a rather solemn little boy, large for his age, with a rather large head, stiff stubby light hair, and somewhat bulging forehead. He is awkward and rather unattractive.

Character and Disposition: Bobbie proved to be a difficult child—resistant to suggestion, uncoöperative and unable to play normally with the other children. He was greedy and selfish. At table he stuffed food into his mouth with his hands, not caring what became of the rest just so his mouth was full. He fairly swallowed it whole. It was exceedingly difficult to interest Bobbie in the normal activities of children of his age. Neither was he interested in the other children. His attitude toward them was on the whole one of indifference, with occasional displays of hostility. He never seemed to have any projects of his own. His movements were awkward and rather poorly controlled. He had less sense of rhythm than most of the three-year-olds.

Home Background: Bobbie's parents had neither of them had any education beyond that of the elementary schools. His father is an unusually intelligent man who is superintendent of a large plant in the city. The mother was a factory operative before her marriage. His grandparents were all native born. They were farmers, artisans, and domestic servants. There was one other child in the family —a young baby.

Development of the 'Mother Complex': Bobbie was first brought to school by his father who reported the following conditions. The mother had lost one child previous to the birth of this one. She was a highly emotional and somewhat violent-tempered woman with no interests outside of her own home. When this child came she lavished all her affection on him. He was never away from her and she gave him a very unwise amount of petting and physical display of affection. When his father tried to interfere and secure some wiser discipline it led to family dissension with the result that apparently the boy was gradually set off against his father. The mother had very little real control over him, and when he began to get older and more troublesome she would say to him, "If you don't behave I will tell your father when he comes home and he will punish you." This, combined with the fact that whenever the father attempted any punishment the mother interfered in Bobbie's behalf, gradually produced a condition in the child which was most pronounced when he came to the school. What we noticed was that he was interested in nothing except securing attention and petting from some adult female. Neither the toys, occupational materials, nor the children interested him in the slightest. He would stand around watching the students until he saw one who seemed inclined to pet him and give him some attention; then he would hang on her skirts as long as he was allowed. Each new student or visitor who came to the school was the object of attacks from Bobbie. He whined and fretted and begged for attention all day. In the afternoon when it drew near the time for his father to come for him, Bobbie would frequently begin to say, with a look of complete conviction, "My daddy won't come for me; he won't come." This he would repeat over and over again until stopped. Occasionally he would vary this by beginning to chant, "My daddy's dead; my daddy's dead; my daddy's dead," over and over again. When his father finally appeared Bobbie would rush for him, grab his legs and begin to demand to be petted. This obviously annoyed his father and he would push him off.

Bobbie frequently remarked that he wished he could wear dresses and be a little girl and that he would like to be a woman when he grew up. When somewhat later the little girls in the school made up a game of playing bride and decked themselves with old lace curtains for bridal veils, Bobbie was very insistent that he wished to be a bride too, and utterly disgusted when the students were unable to make the bridal veil stick on his stubby hair. No other little boy in the school displayed any desire to play the part of bride.

We discovered by chance that he seemed to have a general distrust of men. One day he was taken to the house of one of the members of the staff to be brought to school by her. When he saw her brother in the house he appeared to be frightened and objected to going in because "There is a man in there" and demanded to know, in a somewhat hostile manner, who it was. He was told it was her brother, Fred, and that he liked little boys, but Bobbie would have nothing to do with him and clung to Miss Rodgers until Fred had left the house. Later he wished to know to whom the car in the garage belonged and when he was told it was Fred's he again seemed alarmed and wanted to know if he had to be taken to school in Fred's car. He was obviously relieved when assured that he was to be taken on the bus.

He was allowed to sit at Miss Rodgers' desk while she was getting ready for school and was told that he might have some paper and a pencil (which were given him) but that he must not touch anything else. In spite of this injunction he played with a bottle of ink and spilled some. When Miss Rodgers discovered it and rebuked him for it his first words were, "Don't tell Fred, you won't tell Fred, will you?" When his attention was called to spots of ink on his clothing, he said, "That will come out with water."

In spite of his disgusting methods of eating, Bobbie was inclined to be very careful about his clothes, and was worried when he got them dirty. Since his mother did all of her own work, including the washing, she had doubtless put stress upon the virtue of keeping his clothes clean. Several times in the course of a day Bobbie was apt to ask if his clothes were dirty yet, or call attention to the fact that they were still clean. His efforts were really effective, and the end of a day usually found him noticeably cleaner in clothing than the other little boys. His preoccupation with cleanliness of clothing was one thing that prevented him from joining in many of the children's pursuits.

During the course of his stay of five months in the school, in spite of the fact of frequent periods of absence and rather irregular attendance, Bobbie improved very much. He got to the point where he could occupy himself independently with the toys or beads or clay modelling. His incessant demands for attention were abated and his food habits and habits of eating enormously improved. He seemed a fairly normal child at the end of the term, though his great preference for women and his dislike and distrust of men were still evident. Though his father had never failed a single time to come for him, Bobbie was still apt to begin his monotonous chant of "My daddy won't come, my daddy won't come," and occasionally, "My daddy's dead."

The fact that about the time of his entrance to the school another baby was born into the family probably has much to do with Bobbie's improvement. His mother's time and attention and excessive emotional energy were turned off in part in another direction and Bobbie was somewhat freed from the incubus. It seems to us, however, that his character and disposition have been given a slant which it will be exceedingly difficult to overcome completely.

THE COLOR PREFERENCES OF FIVE HUNDRED AND FIFTY-NINE FULL-BLOOD INDIANS

BY THOMAS R. GARTH

University of Texas

Introduction and Problem

One of the ways of looking for racial differences is to seek them in the field of æsthetics. We have objective facts of a more or less complex character in the art and music of isolated races that suggest just such differences.

In fact popular and a large body of scientific opinion seem to hold to the view that there are fundamental racial differences in æsthetics. Now if there are, *i.e.*, if they are really racial, they are *a priori* native. Being native they are to be regarded as the results of the inheritance factors carried by racial germ plasm.

We mean to say that just a difference, as found, does not mean a difference due to physiological heritage. The differences due to mere environment and circumstance, if these really be found, are not the differences sought.

It is the problem of this experiment, if possible, to find just racial differences, those due to physiological transmission. However in the attempted solution of the main problem other subsidiary problems arise, such for instance, as the influence of education.

Deeming it wise to avoid undertaking the solution of one of the more complex problems of racial æsthetics, we sought a very simple phase of the larger problem by confining our investigation to a study of racial differences in color preference.

It would seem best in an experiment which has for its task the determination of racial differences in color preference to break the problem into three parts:

1. To determine an order of preference for the colors used which might be said to obtain and to find preference values for the colors in the race studied.

2. To ascertain if the preferences, or orders, so determined could be said to be native or acquired for the group.
3. To find out if these preferences are peculiar to the particular race as compared with other races tested under similar circumstances and with the data treated in the same way as the race under consideration.

The Subjects of the Experiment

In the spring of 1921 the writer visited the United States Indian schools located at Chiloco, Oklahoma, and at Albuquerque, New Mexico, for the purpose of giving psychological tests to Indians. This expedition was financed through the courtesy of the Committee on Grants of the American Association for the Advancement of Science. Because it would seem that the element of education, a very important factor in the experiment, could be more nearly controlled, students in these schools were sought as subjects. In this comparisons with white subjects would be more nearly legitimate.

The group whose tendencies we wish to study here is one of full-blood Indians having the ancestry of Plains and Southeastern Indians and of Plateau Indians. Of the first of these there were 281 individuals and of the second 278.

For purposes of comparison as many Indians of mixed blood as were available, 174, and a group of 560 whites were tested with the same color material. The mixed bloods considered, were in fact, a mixture of white blood and Indian blood of the tribes represented in the full-blood group. Unfortunately the number of these mixed bloods is too small for the drawing of definite conclusions from any comparison with the full-blood Indians, and the white performance. As they performed the experiment, the mixed and full-blood Indians sat side-by-side in the school room and the experiments with the whites were administered under conditions as nearly the same as those for the Indians as was possible.

The Materials of the Experiment

As to the materials of the experiment they were seven color discs, of Milton Bradley Company's 'standard' colored papers: Red, Orange, Yellow, Green, Blue, Violet, and White. All the colors are printed on what is called 'coated papers.' By 'standard' it is meant by the makers of the papers 'that particular color in the spectrum which is considered by authorities as the reddest red, the greenest green, the bluest blue,' etc. These discs were one-half inch in diameter and were mounted on white cards $1\frac{1}{2} \times \frac{3}{4}$ inches. In addition to the color discs there were record blanks on which each subject recorded name, sex, age, date, name of school, and, in case he was an Indian, his tribe with degree of Indian blood.

The Procedure

It was very clearly explained to the subjects that record blanks were to be passed to them and on these they were to give the desired information called for as to age, sex, etc. After this part of the procedure had been carried out they were next told that it was desired to find out just what colors they liked best of the ones to be found in the little envelopes about to be distributed. When these had been distributed, one to each subject, instructions were then given that each individual should throw out all the colors contained in his envelope, on the top of his desk. Then it was said that each individual should select the color out of the seven that he liked best and place it on the record blank opposite the '1'; the one he liked next best he should place just below it and opposite the '2' on the record blank; and the one he liked third best was to be placed under that and opposite the '3,' etc., and after all the colors had been arranged, which would of course make the one liked least of all come last and that would be opposite the '7,' they were to write the name of each color beside the color itself on the record blank before disturbing their orders of preference. Then it would be expected of them that they place the colors—all seven of them—back in the envelopes. But care was taken to see that each subject followed these instructions and also that there was no suggestion of any kind on the part of the experimenter or his assistants as to what the individual's color preference should be.

Handling of the Data

The record blanks gave the information desired and the data so obtained were tabulated with due regard for classification with reference to race, sex and education.

In the full-blood Indian group there were of both sexes and all ages ranging from 11 to 21 years, as we have said, 559 cases. This makes a homogeneous group and is sufficiently large for our purposes. The question now is as to how we shall go about determining the order of preference for these colors which obtains in this large group of Indians. The tabulation referred to above which is the first draft of the material, by way of illustration, looks something like the following:

Boys—Fourth Grade—Rankings Given the Colors

Subject No.	1	2	3	4	5	6	7	8	9	10	11	12, etc.
Age Yrs.	11	11	12	12	12	12	13	13	13	13	13	14
Rank given:												
Red	1	1	2	1	2	5	3	1	6	3	2	2
Blue	3	3	1	2	3	1	1	2	3	1	1	1
Green	2	4	3	4	1	3	2	4	4	4	3	3
Purple	4	2	5	3	4	2	4	3	1	2	5	4
Orange	5	6	4	5	6	4	5	5	2	5	4	5
Yellow	6	5	6	6	5	7	6	7	5	7	6	6
White	7	7	7	7	7	6	7	6	7	6	6	7

In order to find out the order of preference existing in the data in this first tabulation and to bring same to light, we may do two things: (1) simply average the ranks given to any color, calculate the A.D. and P.E., find the median of the ranks, and having done this for all the colors proceed to arrange our colors according to the ranks of these averages and medians, determining the reliabilities of the differences according to the accepted formula for same; and (2) we may construct a scale of color preference for Indians such as Thorndike constructed for handwriting and Hillegas made for English composition. (See *Teachers College Record*, Vol. II., No. 2, and Vol. XIII., No. 4.)

Both of these methods have their advantages. The first gives us a rough idea of the tendencies of the data and permits of the use of measures of overlapping which is a matter of importance here; the second will make possible the determination of the preferential value of any color under the experimental conditions and permits of a graphic representation of the relative positions which may be assigned to the several colors on a scale of preference.

The Color Preference of Indians

Availing ourselves of both methods, we begin with the first of them, as: Table I. gives the average and median

Table I

Showing Central Tendencies and their Ranks of Group of 559 Full-Blood Indians in Judging Seven Colors

Ranks of	Ave.	A.D.	P.E.	Med.	Ave.	Med.	Per Cent. Attaining and Exceeding Median of Preceding Color
Red	2.5	1.04	.8	2.5	1	1	
Blue	3.02	1.4	1.1	3.3	2	2	(Red) 67
Violet	3.6	1.7	1.4	8.9	3	3	(Blue) 62
Green	4.09	1.1	.8	4.7	4	4	(Violet) 64
Orange	4.7	1.3	1.1	5.3	5	5	(Green) 65
Yellow	5.2	1.2	.9	6.1	7	6	(Orange) 66
White	4.9	1.8	1.5	6.5	6	6	(Yellow) 52

position assigned to the seven colors by the 559 full-blood Indian subjects as well as the sequence of the colors as indicated by the ranking of these. It gives also the overlapping, or measure of per cent. of judgments, of a color attaining and exceeding the median assigned to the color just preceding it. This latter figure indicates whether or not the several sequences as indicated by the respective medians are supported by a fact of difference between them, *i.e.*, as to whether or not they are representative and if not exactly representative whether or not they tend to be so. It will be seen that in this group of full-blood Indians the sequence of the preference for these colors is this, going from most preferred to least preferred: First, Red; Second, Blue; Third, Violet; Fourth, Green; Fifth, Orange; Sixth, Yellow and White. That is, the situation of White and Yellow is undetermined by this handling of the data. The sequence, shown in Table I., tends to be representative where such a sequence is indicated, as tested by the overlapping, with a slight advantage for the Yellow over White; however this advantage is barely evident.

But this method fails to give us all the facts afforded by the data, for we want to know how much stronger as measured is the preference value of one color than another as reckoned from a common reference point. It is just here that this experiment claims to be different in method from other color preference experiments and experiments for determining the relative 'pulling power' of similar given stimuli. We do not feel that this is the place for a discussion of the relative merits of either method of procedure, but we recommend its use for the purpose of measuring the relative strengths of colors, 'selling points,' advertisements, and similar material.

A Color Preference Scale

It may not be necessary to describe the derivation of a judgment scale such as that of Thorndike in his Handwriting Scale and that of Hillegas in his Scale for the Measurement of Quality in English Composition and the writer's Scale for Measuring Method of Study of College Students (see *Pedagogical Seminary*, Mar., 1920), but some account of the method of handling this particular mass of data for determining color preference of Indians may be more or less worth while. It was simply a matter of finding how many times out of 559 one color was placed above the other. Table II.

Table II

Showing Number of Times One Color Was Preferred to Another by the Group of 559 Full-blood Indians

The table reads thus: White was preferred to White never, to Yellow 268 times, to Orange 243 times, etc., out of 559.

	White	Yellow	Orange	Green	Purple	Blue	Red
White	0	268	243	221	208	155	78
Yellow	291	0		147	143	113	79
Orange	316	361	0	224	183	128	97
Green	338	412	335	0	220	180	149
Purple	351	417	376	339	0	229	184
Blue	404	446	431	379	330	0	219
Red	481	480	462	410	375	340	0

shows this for all colors, as Red was preferred to Blue by 340 out of the 559 pure-blood Indians; Red was preferred to

Purple by 375 out of the 559, was preferred to Green by 410. Yellow was preferred to White by 291 out of the total 559. Since this last was only 53 per cent. of the time, it had practically little advantage over White. Table III. shows what per cent. of the subjects preferred any color above another.

Table III

Numbers of Table II. Expressed as Percentages of Total Judgments

To be read: White was preferred to Yellow 47 per cent. of the time, to Orange 43 per cent. of the time, to Green 39 per cent. of the time, etc.

	White	Yellow	Orange	Green	Purple	Blue	Red
White	0	47	43	39	37	28	14
Yellow	53	0	35	26	25	20	15
Orange	56	64	0	40	33	23	17
Green	60	73	60	0	39	32	27
Purple	63	74	67	60	0	41	33
Blue	72	79	77	67	59	0	39
Red	86	85	82	73	67	61	0

The next step in deriving a scale of color preference is to take the facts of Table III. and arrange the seven colors in sequence running from least preferred to most preferred and having done this to transmute the necessary percentages into values of D/M.D. by using the table given on pages 24 and 25 of Hillegas' 'A Scale for Measurement of Quality in English Composition' which was derived from a table found in Thorndike's 'Mental and Social Measurements,' page 228.

The reason for using White at all was with the expectation that, since it is achromatic, it would prove to be a zero point in color preference. But it did not serve the purpose of an absolute zero point, so we took it as an *arbitrary* zero point, and upon that constructed our scale. The differences as expressed in terms of Difference divided by Median Deviation for the several colors are these:

Color		Color	Per Cent. of the Time	D/M.D.
Red	preferred to	Red	0	—
Blue	" "	Red	39	.41
Violet	" "	Blue	41	.34
Green	" "	Purple	39	.41
Orange	" "	Green	38	.40
Yellow	" "	Orange	35	.57
White	" "	Yellow	47	.11

This gives the following scale values:

Color	Value
White	.00
Yellow	.10
Orange	.68
Green	1.08
Violet	1.49
Blue	1.83
Red	2.24

It will be seen that for this group of Indians, there are fairly definitely established preferences for the several colors which stand at almost equal distances on the scale. Violet has more than twice the value of Orange. Red is most preferred of all. It is twenty times as strong as Yellow, three times as strong as Orange, one and one half times as strong as Violet, and one and one fourth times as strong as Blue. Taking Yellow as a base, we have approximately these proportions:

Yellow	Orange	Green	Violet	Blue	Red
1	6	9	13	16	20

This sequence represents a series in arthmetical progression after leaving Yellow. These observations and all those following hold only for the data in hand. We do not know whether or not more data from individuals of such racial stocks as represented here would change the above sequence, however since the group is rather large, it seems to the experimenter probable that this would not change materially the results.

While the above scale gives the order of preference of the seven colors as obtained from the 559 full-blood Indians, the accompanying chart (Fig. 1) gives a graphical representation

A Graphic Representation of Color Preference Scales Showing Positions Assigned the Colors as Determined by

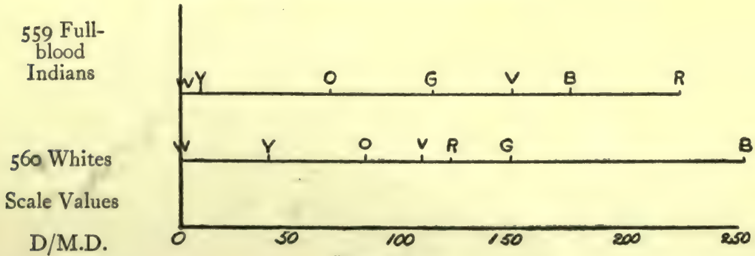


Fig. 1.

of this. In addition there is likewise a graphic representation for a group of white subjects similarly constructed. At the bottom of the chart the values of various points on a scale thus constructed are given. On the line above these we give the positions as derived for the colors on this scale for the group of 559 full-blood Indians.

So much for the sequence of the colors, the reliability of the differences between the members of the sequence and their relative values on a color preference scale for the full-blood Indians. The sequences as indicated tend to be real with the exception of the sequence: Yellow followed by White. The relative preferences of the colors are indicated on the scale.

The Influence of Education on Color Preference in Indians

The second problem which we wish to attack after finding a satisfactory sequence of color preferences for our original group of Indians, is stated in the question: Are the color preferences, here indicated, native or acquired? Unfortunately we are not able to go below the fourth grade in the matter of education with our subjects, for all of them had that much school training in the United States Indian schools. And since we cannot go below that in the present case, we shall have to grant that they possessed the nearest approach to native preference our data offers.

We shall, then, take the color preferences of these subjects as a basis of comparison with the color preferences of students of grades higher up. In order to make this study of the influence of education we have made a division of our 559 subjects on a basis of school grades which gives us five sub-groups, starting with the fourth grade, whose compositions are as follows:

Grade	Number
Fourth	113
Fifth	121
Sixth	117
Seventh	106
Eighth, Ninth, Tenth	102
Total	559

The same measures found for the full bloods in the total group, were found for the educational subgroups into which we have divided it, *i.e.*, averages, medians, measures of overlapping. Table IV. contains an array of these measures and

TABLE IV

SHOWING CENTRAL TENDENCIES AND THEIR RANKS FOR SEVERAL GRADE SUBGROUPS OF FULL-BLOOD INDIANS

Color	Ave.	A.D.	P.E.	Med.	Ave.	Med.	Per Cent. Attaining and Exceeding Median of Preceding Color
Fourth Grade—113 Judges							
Red	2.6	1.4	1.2	2.6	1	1	
Blue	3.1	1.1	.9	3.5	2	2	(Red) 73
Green	3.4	1.7	1.4	3.7	3	3	(Blue) 57
Violet	4.3	1.9	1.6	4.6	4	4	(Green) 64
Orange	4.8	1.1	.9	5.3	6	6	(Yellow) 50
Yellow	4.5	1.3	1.0	5.3	5	5	(Violet) 66
White	4.0	1.3	1.1	6.0	7	7	(Orange) 62
Fifth Grade—121 Judges							
Red	2.2	1.0	.8	2.1	1	1	
Blue	3.1	1.3	1.0	3.1	2	2	(Red) 89
Violet	3.6	1.6	1.3	4.0	3	3	(Blue) 68
Green	4.2	1.6	1.3	4.9	4	4	(Violet) 63
Orange	4.3	1.3	1.0	5.2	5	5	(Green) 59
Yellow	4.4	.8	.7	6.2	6	7	(White) 57
White	4.8	1.8	1.5	6.1	7	6	(Yellow) 49
Sixth Grade—117 Judges							
Red	2.1	.9	.7	2.1	1	1	
Blue	3.1	1.7	1.4	2.8	2	2	(Red) 79
Violet	3.5	1.5	1.2	4.1	3	3	(Blue) 72
Green	4.4	1.1	.9	5.1	4	4	(Violet) 71
Orange	4.7	1.4	1.1	5.3	5	5	(Green) 58
Yellow	5.4	1.0	.8	6.2	7	7	(Orange) 51
White	4.8	1.8	1.5	6.1	6	6	(Yellow) 48
Seventh Grade—106 Judges							
Red	2.4	1.2	.9	2.5	1	1	
Blue	3.2	1.3	1.1	3.6	2	2	(Red) 75
Violet	3.4	1.6	1.3	3.8	3	3	(Blue) 55
Green	4.4	1.2	1.0	5.1	4	4	(Violet) 77
Orange	4.5	1.3	1.1	5.5	5	5	(Green) 63
Yellow	5.2	1.2	.9	6.1	7	6	(Orange) 63
White	4.3	1.8	1.5	6.1	6	6	(Yellow) 50
Eighth, Ninth, and Tenth Grades—102 Judges							
Red	2.1	1.3	1.1	3.1	1	1	
Blue	2.8	1.3	1.1	3.1	2	1	(Red) 50
Violet	3.2	1.5	1.2	3.5	3	2	(Blue) 58
Green	3.9	1.5	1.2	4.4	4	3	(Violet) 66
Orange	4.5	1.2	.9	5.2	5	4	(Green) 73
Yellow	5.6	1.1	.9	6.5	7	5	(Orange) 76
White	4.9	1.7	1.4	6.5	6	5	(Yellow) 50

the orders of sequence of the colors for each educational subgroup.

While the sequences are not exactly the same, there is very little change in that respect as we proceed from the fourth grade sequence of color preference to the highest school grade sequence.

In the fourth grade the sequence is Red, Blue, Green, Violet, Yellow, Orange, White. In the fifth grade almost the same sequence obtains for averages and medians excepting for the Green and the last three colors, in which the results for averages and medians do not coincide for Yellow and White. While they do in the next higher grade, the sixth, where Yellow goes to the lower end of the scale, they do not in the seventh grade sequence. This contest between Yellow and White may be seen ending up in the highest educational group with White and Yellow about even, but Green never recovers its lost prestige. Red is always in the ascendancy excepting in this last group where the median shows Blue occupying the same rank with it. It should be noted that the measures of overlapping are usually sixty per cent. and above. The table will indicate the cases in which this measure is or is not approaching significance. The conclusion to be reached here is that education from the fourth grade up has very slight effect on the sequence of color preferences of full-blood Indians as indicated by this method of handling the data, though it is true that Blue and Violet tend to move up toward the position given Red—*i.e.*, first place; and White tends to displace Yellow in a few instances.

In order to determine the reliability of the differences in color preference, if any is to be found in moving up the grades, measures of overlapping of the distribution of each grade were obtained on the Fourth Grade distribution as a base. This is shown in Table V. If there is any difference its tendency may be thus measured. As for instance, the facts of this table bear us out in saying that Red experiences no change throughout the grades until we reach the last grade group—that of 8th, 9th, and 10th grades, where the overlapping on the fourth grade is significant. The overlapping is significant again in 6th grade for Blue where that color's ranking was better than was the same color's ranking in the 4th grade—*i.e.*, was more preferred. But the overlapping for

TABLE V

SHOWING OVERLAPPING OF SUCCESSIVE SCHOOL-GRADE DISTRIBUTIONS ON THAT OF FOURTH GRADES FOR EACH COLOR FOR FULL-BLOOD INDIANS

	Fourth Grade Medians	Overlapping for 4th Grade on			
		Fifth Grade	Sixth Grade	Seventh Grade	8th–10th Grades
Red..........	2.6	45%	44%	50%	63%
Blue..........	3.5	49	36	56	54
Violet.........	4.6	42	45	37	31
Green........	3.4	70	79	79	63
Orange......	5.3	49	54	59	51
Yellow.......	5.3	72	73	68	74
White.......	6.0	52	52	54	51

TABLE V*a*

SHOWING AGES OF FULL-BLOOD INDIANS FOR EACH EDUCATIONAL SUBGROUP

	Fourth Grade	Fifth Grade	Sixth Grade	Seventh Grade	8th–10th Grades
Ave...........	14.5 yrs.	15.4 yrs.	16.2 yrs.	17.7 yrs.	18.5 yrs.
A.D..........	2.4	1.6	1.5	1.7	1.8
P.E..........	2.0	1.3	1.2	1.2	1.4

that color is not significant elsewhere, not even in the highest grades. The situation with respect to Violet and Green reveals this, that the overlapping for Violet becomes significant (see table) in the 7th grade and increases in significance in the highest grade and that Green shows significant differences in the grades above the 4th, but it is not of a constantly increasing significance. This is likewise true of Yellow. In all, these measures of overlapping indicate that no great change of an increasing nature is brought about in color preference by education except in the case of Violet and Red. Violet increases and Red loses in preference.

COLOR PREFERENCE SCALES FOR EDUCATIONAL SUBGROUPS

As in the case of the group of 559 proper we have derived separate scales for the educational subgroups for the purpose of comparison of 'pull' of colors in the grades. These scales and their graphic representation are given here in Table VI.

TABLE VI

COLOR PREFERENCE SCALES FOR EDUCATIONAL SUBGROUPS—FULL-BLOOD INDIAN CHILDREN

Fourth Grade

Color	Scale Value
White	.00
Yellow	.41
Orange	.48
Violet	.67
Green	1.05
Blue	1.31
Red	1.84

Fifth Grade

Color	Scale Value
Yellow	.00
White	.04
Orange	.23
Green	.57
Violet	.95
Blue	1.36
Red	2.01

Sixth Grade

Color	Scale Value
Yellow	.00
White	.07
Orange	.29
Green	.51
Violet	1.25
Blue	1.90
Red	2.35

Seventh Grade

Color	Scale Value
White	.00
Yellow	.11
Orange	.52
Green	.59
Violet	1.37
Blue	1.44
Red	1.88

Eighth, Ninth, and Tenth Grades

Color	Scale Value
White	.00
Yellow	.26
Orange	1.26
Green	1.60
Violet	2.09
Red	2.39
Blue	2.46

and Fig. 2 (which gives the graphic representation of similarly treated white groups each of about the same size as these Indian groups). The scales only and not the tables from which they were derived are given here for the sake of economizing space.

As we derived these scales for the grade groups, five in number, for fourth, fifth, sixth, seventh grades, each a scale, and one scale for eighth, ninth, and tenth grades combined in one, we discovered that our original arbitrary zero point, the achromatic White, which served very well in the large scale for full bloods had shifted very slightly in two of the educational subgroups, *i.e.*, in the fifth and sixth grades. That is to say, White was preferred to Yellow in both of these educational groups, but it found its place below the Yellow again in the last two of the educational subgroups, the seven thand the 8th–9th–10th grade subgroups. We cannot account for this and do not wish to speculate to any extent on its significance. If it has any significance in being preferred to Yellow, in these two instances that significance is very slight, and this may lie in the fact that the contest of Yellow with the other chromatic samples was so great that it overshadowed the contest of these with just no color, White. It will be observed that Red maintains its position as most preferred in all the educational subgroups until we

A Graphic Representation of Color Preference Scales Showing Positions of Colors for Various Educational Groups of Whites and Full-blood Indians

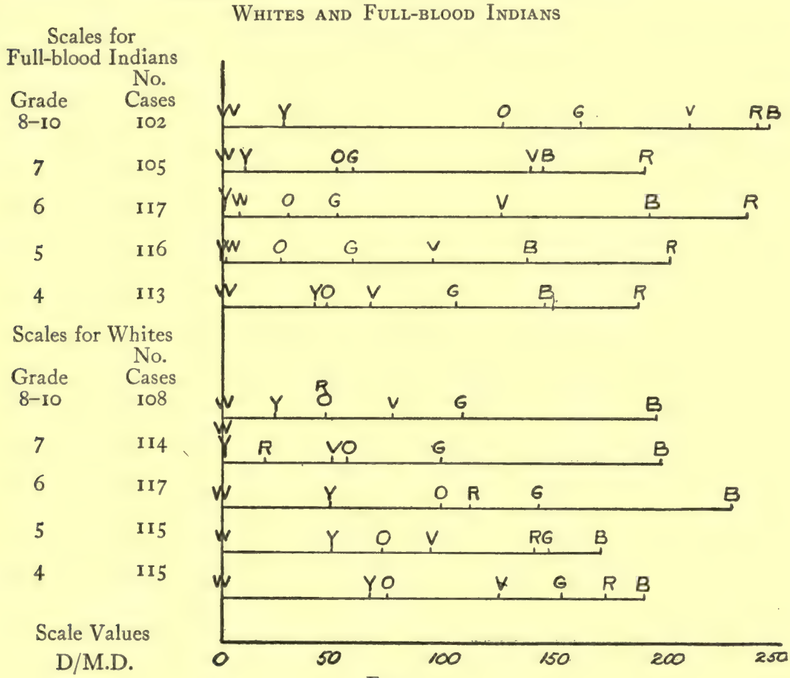


Fig. 2.

come to the most educated subgroup where it barely succumbs to a slight preference for Blue, when we examine the chart and the accompanying scales for each grade subgroup.

Red starts in the fourth grade subgroup with four times the preference given to Yellow and shows nearly nine times that of Yellow in the last two most educated subgroups. In its contest with Violet, it loses some of its 'pull' over that color, starting with three times and ending up with one and one tenth times the 'pull' it had over that color. It starts with a 'pull' equal to one and one half times that of Blue and is slightly inferior in 'pull' to it in the most educated group.

Since the factor of age has been to a certain extent controlled along with education (See Table V*a*.) no detailed study of that factor will be made here. On the average age increases with education.

Sex Differences in Color Preference among the Indians

A question arises as to what influence sex has on the color preferences of the full-blood Indians. An examination of the literature shows that no striking sex differences obtain for whites. Jastrow found a masculine tendency to prefer Blue for first place and Red for second, and a feminine tendency to place Red first and Blue second (2, 3).

In order to answer the question as to whether there is a sex difference among the subjects representing full-blood Indians we have separated our 291 males from our 268 females and found for the two sex groups the facts as represented in Tables VII. and VIII., which give the sequences of averages

Table VII

Showing Central Tendencies and Sequences of Positions Assigned to the Colors by Males and Females of Full-blood Indians

I. *For Males, 291 Cases*

Color	Ave.	A.D.	P.E.	Med.	Rank of: Ave.	Rank of: Med.	Per Cent. Attaining and Exceeding Median of Preceding Color
Red	2.4	1.7	1.3	2.5	1	1	
Blue	3.4	.9	.7	3.3	2	2	(Red) 79
Violet	3.5	1.8	1.4	3.8	3	3	(Blue) 58
Green	8.9	1.6	1.3	4.5	4	4	(Violet) 61
Orange	4.5	1.0	.8	5.2	5	5	(Green) 67
Yellow	5.0	1.1	.9	5.8	6	6	(Orange) 64
White	5.0	2.0	1.6	6.3	6	7	(Yellow) 59

II. *For Females, 268 Cases*

Color	Ave.	A.D.	P.E.	Med.	Rank of: Ave.	Rank of: Med.	Per Cent. Attaining and Exceeding Median of Preceding Color
Red	2.4	1.2	.9	2.3	1	1	
Blue	2.6	1.3	1.0	2.7	2	2	(Red) 62
Violet	3.6	1.3	1.0	4.1	3	3	(Blue) 77
Green	4.3	1.2	.9	5.0	4	4	(Violet) 69
White	4.6	2.0	1.6	5.4	5	5	(Green) 59
Orange	4.8	1.1	.9	5.4	6	5	(Green) 62
Yellow	5.5	1.3	1.0	6.5	7	6	(Orange) 70

and medians and the overlapping of one color distribution over the preceding one in the sequence. Also we have derived color-preference scales for the sex groups which accompany these tables. As to the sequences, they are more or less definite for the sex groups and are practically the same for the males and females with this exception, that the females place White above Yellow and equal in 'pull' to Orange, the males placing it last. The experimenter thinks he can account in part for this high rating of White on the part of the females by the fact of recency, since at the time the experiment was administered the older female students were preparing for graduation and he was warned of this fact by the teacher of domestic science in one of the schools. For that reason it is only in all probability a temporary tendency which spread from the older students down the grades. This is supported by the fact that the lower grade students tend to place White lower in the scale than do the older students of this sex.

Table VIII

Showing Overlapping of Positions Assigned by Males on Female Distribution

Per Cent. of Male Judgments Attaining and Exceeding Median of Female Median Position for Each Color

Red	52
Blue	76
Violet	48
Green	41
Orange	49
Yellow	37
White	61

Table VIII*a*

Scale Values for Male and Female Indian Color Preference

Males (291) Color	Scale Value	*Females (265)* Color	Scale Value
White	.00	Yellow	.00
Yellow	.45	White, Orange	.74
Orange	.86	Green	1.05
Green	1.24	Violet	1.62
Violet	1.43	Blue	2.36
Blue	1.47	Red	2.56
Red	2.67		

An examination of the facts of Table VIII shows a tendency for the males to place Blue lower than females, and Yellow higher and White lower, since the overlapping for the first is 76 per cent. and for the last two colors is respectively 37 per cent. and 61 per cent. However this difference with respect to the first color, Blue, does not alter the sequence in that instance. It merely indicates that the females esteem Blue more highly than do the males, but not to the extent of placing it above Red. An examination of the two color preference scales for the two sexes bears us out in this statement for if we should eliminate White from the males' scale entirely the scale value for Blue would be 1.02 and examination of the females' scale shows that Blue has a preference value of 2.36. The two scales for the sexes show other interesting features as Orange has nearly twice the scale value (eliminating White from the scale) for the females as it has for the males and under the same conditions rather less value with the males than it has for the females as have all the colors. Still these sex differences do not alter the race sequences of the color preferences, we believe.

Racial Differences in Color Preference

In order to answer the third question as to whether or not these full-blood Indians differ from mixed bloods and whites in their color preferences, we shall compare *first*, tendencies of the racial groups and their differences, and, *second*, the color preference scales for these groups. Only in the case of the whites shall we study the influence of education on these representatives of that race, and bring the facts so derived to bear on the array of facts obtained in the previous study of the influence of education on the Indians. This will not be possible in the case of the mixed bloods because of the fact, noted above, of the inadequacy of the number of individuals in this group.

Tables IX. and X. show the averages and medians of the positions assigned to the colors by the group of 174 mixed bloods and by the group of 560 whites as well as the ranks of these averages and medians. Table XI. shows the white group broken up into educational subgroups corresponding to those of the full bloods.

An examination of these various measures will show for

Table IX

Showing Central Tendencies and their Ranks of Group of 174 Mixed-blood Indians in Judging Colors

Color	Ave.	A.D.	P.E.	Med.	Ave.	Med.	Per Cent. Attaining and Exceeding Median of Preceding Color
Red	2.7	1.0	.5	3.01	2	2	(Blue) 56
Blue	2.6	1.2	.6	2.8	1	1	
Violet	3.6	1.6	.8	4.2	3	3	(Red) 62
Green	4.4	1.4	.6	4.9	5	4	(White) 66 or (Violet) 66
Orange	5.1	1.0	.5	5.7	6	5	(Green) 72
Yellow	5.3	1.3	.6	6.2	7	6	(Orange) 62
White	4.1	1.7	.7	4.2	4	3	(Violet) 50

Table X

Showing Central Tendencies and their Ranks of Group of 560 Whites in Judging Colors

Color	Ave.	A.D.	P.E.	Med.	Ave.	Med.	Per Cent. Attaining and Exceeding Median of Preceding Color
Blue	2.2	1.1	.9	2.2	1	1	
Green	3.4	1.3	1.0	3.8	2	2	(Blue) 83
Red	3.8	1.7	1.4	4.2	3	3	(Green) 51
Violet	3.9	1.2	.9	4.6	4	4	(Red) 60
Orange	4.4	1.2	1.0	5.1	5	5	(Violet) 61
Yellow	4.7	1.3	1.0	5.6	6	6	(Orange) 61
White	4.9	1.7	1.5	6.3	7	7	(Yellow) 60

the mixed bloods and for whites, sequences different from that of full bloods, as:

	Rank						
	1	2	3	4	5	6	7
Full Bloods	Red	Blue	Violet	Green	Orange	Yellow	White
Mixed Bloods	Blue	Red	Violet	White	Green	Orange	Yellow
Whites	Blue	Green	Red	Violet	Orange	Yellow	White

Blue and Red are scarcely different for mixed bloods and in the case of whites, Red and Green are so nearly alike that any discrimination will be more or less arbitrary. The reliability of the several sequences as representing real differences may be observed by examining the above designated tables.

It will be necessary now to measure the overlapping of the positions on the scale assigned to the colors by the different racial groups. (See Table XI.)

TABLE XI

A. SHOWING THE OVERLAPPING OF FULL-BLOOD INDIANS ON MIXED BLOOD AND WHITE MEDIANS FOR POSITIONS OF EACH COLOR

	Per Cent. of Full-blood Judgments Attaining Median Position Assigned by:	
Color	174 Mixed Bloods	560 Whites
Red	39	23
Blue	63	83
Violet	47	41
Green	48	66
Orange	43	59
Yellow	51	58
White	68	51
Range	39–68	23–83

B. SHOWING THE OVERLAPPING OF FULL-BLOOD INDIANS ON MEDIAN OF WHITES FOR EACH GRADE GROUP

	Per Cent. of Full-blood Indian Judgments Attaining and Exceeding Medians for Positions of Each Color:				
	Fourth Grade	Fifth Grade	Sixth Grade	Seventh Grade	8th–10th Grades
Red	38	26	14	12	11
Blue	74	76	71	79	56
Violet	51	40	35	21	33
Green	43	63	77	72	67
Orange	64	51	29	63	59
Yellow	50	69	62	61	67
White	34	43	48	54	50

Now if the measures of overlapping of full bloods over whites and mixed bloods may be taken as indicating tendencies toward real differences in the groups, we may say that these comparisons of the blood groups show that for each color there is more tendency toward real difference in color preference between full bloods and whites than between full bloods and mixed bloods because the measures of overlapping are on the whole more significant for full bloods on the white than on the mixed-blood distribution. The overlapping in the former case runs from 23 per cent., in the case of Red, to 83 per cent., in the case of Blue, and is smallest in the case of White, 51 per cent. That is, the differences in preference for Red and Blue are real for the groups—full bloods and whites—and are very much alike for White, Violet, and Yellow. As to the overlapping of full blood on mixed-blood distribution, this runs from 39 per cent. in the case of Red to 68 per cent. in the case of White, with that of Blue 63 per cent. There is more tendency to difference in the case of White than in the case of Blue and Violet, here.

Color Preference Scales for Whites and Mixed Bloods

So much for the determination of indicated racial group differences, but we believe that insofar as the comparison of white and full-blood Indian color preference is concerned, it will be more definitely represented by a comparison of the color preference scales. In the case of the mixed-blood scale, such comparisons are not nearly so legitimate, because of the disparity of the number of subjects, here only 174 in the mixed-blood group, while there are 560 whites and 559 full bloods.

Color preference scales for the 174 mixed bloods and the 560 whites were constructed in the same way that such scales were made for the full-blood Indians. These are given below. Figure 1 affords a graphic representation of the white scale along with that of the full-blood scale, and Figure 2 gives this for the mixed-blood scale. These comparisons are permissible: The white subjects gave to Yellow four times the

Color Preference Scale for Whites—560 Judges

Color	Scale Value
White	.00
Yellow	.41
Orange	.82
Violet	1.08
Red	1.23
Green	1.49
Blue	2.49

Color Preference Scale for Mixed-blood Indians—174 Judges

Color	Scale Value
Yellow	.00
Orange	.30
Green	.75
White	.94
Violet	1.01
Red	1.46
Blue	1.53

value given it by the Indians, reckoning from White as a point of reference. To Orange, they give about the same, according it only slightly more preference. Violet has more 'pull' for the Indians than it does for the whites, while Green 'pulls' harder on the latter than on the former, something like a third more, if we may express these relations in terms of scale positions. Further, Red is esteemed nearly twice as much by the Indians as by the whites, and Blue has something like three fifths of the value for them that it has for the whites.

Because of the smaller group, 174 judges, deriving the Mixed-blood Scale, we do not feel disposed to make very definite comparisons between it and those for the full bloods and whites. But if the scale as it stands had been derived by a group of judges of the same size as the other racial groups we could say, taking Yellow as a reference point in all scales, that the mixed bloods have somewhat less regard for Orange than the whites and much less than the full bloods. The 'pull' of Green, and the 'pull' of Orange stand in about the same ratio with them, that they do with the whites. This is likewise true of Blue and Orange. But they are more like the full bloods when the 'pulling powers' of Violet and Red are taken in relation to Orange. However, we only suggest by this what might be ascertained if the group of mixed bloods were equal in size to that of the other racial groups.

Comparison of Influence of Education on Whites and Indians

We have not yet finished the discussion of racial differences in color preference, for the data affords other possibilities still. We have noted that the sequence of color preferences of full-blood Indians changes only slightly with education if we may take the groups as representative of full-blood Indians. Is this so with the whites? We have seen that the median positions assigned to the several colors by a race group differ in many points for the large groups of whites and Indians as measured by overlapping of Indian distribution on the white distribution. Do these tendencies to differ exist in the educational subgroups? Or does education tend to eliminate these differences? We have observed the tendency for a few colors to change in relation to 'pulling power' as education increases with the full-blood Indians. How does this fall out for the whites? In fact, these, and many other questions may be asked in this connection.

Table XII. gives the central tendencies of positions assigned to the colors by the whites of various grade subgroups, their sequences and the tendencies toward real differences for the colors in a designated sequence. The tables show that Red tends to lose in preference value with education and Green to gain. This is not supported by the findings of some other investigations. See Winch (1). Green starts with third place in the fourth grade, and gains second place in the sixth grade, and maintains this position, while Red falls to fourth place, in this last educational subgroup, from that of second place in the first two grade subgroups. Otherwise there is practically little change in the sequence as found in going from one grade group to the other for the white judges. This is quite different in the case of the full bloods where Green instead of gaining fell away early in the grades and Red tended to remain first instead of falling as it did with the whites. Violet displaces Red for the whites in the most educated white subgroup, but, though it encroaches on it, it never displaces that color in the educational subgroups of full-blood Indians.

Table XIII. shows scales derived for the educational subgroups of whites. Fig. 3 offers a graphic representation of

A Graphic Representation of Color Preference Scale for 174 Mixed-blood Indians

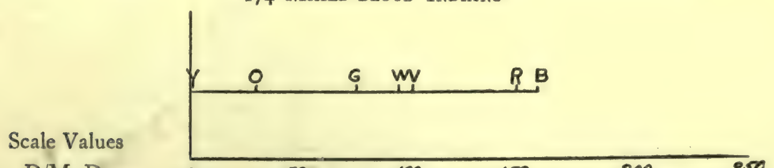


Fig. 3.

the same scales. As to the comparative effect of education

Table XII

Showing Central Tendencies and their Ranks for Several Educational Subgroups of Whites

Fourth Grade—115 Judges

Color	Ave.	A.D.	P.E.	Med.	Ave.	Med.	Per Cent. Attaining and Exceeding Median of Preceding Color
Blue	2.8	1.1	1.0	2.5	1	1	
Green	3.6	1.4	1.1	4.0	3	3	(Red) 72
Red	3.3	1.8	1.4	3.2	2	2	(Blue) 65
Violet	4.0	1.5	1.2	4.5	4	4	(Green) 60
Orange	4.4	1.3	1.2	5.1	5	5	(Violet) 66
Yellow	4.6	1.4	1.2	5.4	6	6	(Orange) 60
White	5.5	1.6	1.4	6.1	7	7	(Yellow) 70

Fifth Grade—115 Judges

Color	Ave.	A.D.	P.E.	Med.	Ave.	Med.	Per Cent. Attaining and Exceeding Median of Preceding Color
Blue	2.2	1.1	.9	2.3	1	1	
Green	3.54	1.4	1.2	4.0	3	3	(Red) 59
Red	3.51	1.6	1.2	3.8	2	2	(Blue) 79
Violet	4.1	1.4	1.1	4.6	4	4	(Green) 64
Orange	4.6	1.4	1.1	5.3	5	5	(Violet) 66
Yellow	4.7	1.6	1.2	5.4	6	6	(Orange) 62
White	5.3	1.6	1.2	6.7	7	7	(Yellow) 66

Sixth Grade—117 Judges

Color	Ave.	A.D.	P.E.	Med.	Ave.	Med.	Per Cent. Attaining and Exceeding Median of Preceding Color
Blue	1.9	1.0	.8	2.0	1	1	
Green	3.5	1.9	1.5	3.8	2	2	(Blue) 89
Red	3.8	1.5	1.2	4.2	3	3	(Green) 59
Violet	4.1	1.4	1.2	4.6	4	4	(Red) 63
Orange	4.9	1.3	1.0	6.3	5	6	(Yellow) 45
Yellow	5.0	1.1	.9	5.9	6	5	(Violet) 77
White	4.9	1.3	1.0	6.3	5	6	(Yellow) 56

Seventh Grade—105 Judges

Color	Ave.	A.D.	P.E.	Med.	Ave.	Med.	Per Cent. Attaining and Exceeding Median of Preceding Color
Blue	2.0	.9	.8	2.1	1	1	
Green	3.4	1.4	1.2	4.0	2	2	(Blue) 83
Red	4.3	1.7	1.4	4.9	4	3	(Green) 66
Violet	4.2	1.4	1.1	5.0	3	4	(Red) 51
Orange	4.3	1.3	1.0	4.9	4	3	(Green) 70
Yellow	4.8	1.4	1.0	5.7	6	5	(Violet) 59
White	4.5	1.7	1.5	5.8	5	6	(Yellow) 52

Eighth, Ninth, and Tenth Grades—108 Judges

Color	Ave.	A.D.	P.E.	Med.	Ave.	Med.	Per Cent. Attaining and Exceeding Median of Preceding Color
Blue	2.1	1.1	.9	1.9	1	1	
Green	3.3	1.4	1.1	3.3	2	2	(Blue) 84
Red	4.3	1.6	1.2	4.9	4	4	(Violet) 57
Violet	3.9	1.4	1.1	4.5	3	3	(Green) 73
Orange	4.3	1.4	1.1	4.9	5	4	(Violet) 57
Yellow	4.9	1.3	1.0	5.9	6	5	(Red) 65
White	5.1	1.7	1.5	6.1	7	6	(Yellow) 65

on the relative 'pull' one color has over another in the two races we noted for the full-blood Indians that Red increases its 'pull' over Yellow, nearly doubling its effect with increase

TABLE XIII

COLOR PREFERENCE SCALES FOR EDUCATIONAL SUBGROUPS—560 WHITE CHILDREN

Fourth Grade Color	Scale Value	*Fifth Grade* Color	Scale Value	*Sixth Grade* Color	Scale Value
White	.00	White	.00	White	.00
Yellow	.65	Yellow	.53	Yellow	.49
Orange	.74	Orange	.72	Orange	.98
Violet	1.22	Violet	.94	Violet, Red	1.09
Green	1.52	Red	1.39		
Red	1.71	Green	1.44	Green	1.39
Blue	1.90	Blue	1.71	Blue	2.39

Seventh Grade Color	Scale Value	*8th, 9th, and 10th Grades* Color	Scale Value
White, Yellow	.00	White	.00
Red	.19	Yellow	.22
Violet	.49	Red, Orange	.48
Orange	.56	Violet	.78
Green	.97	Green	1.08
Blue	1.97	Blue	1.94

of education for the subjects considered, but with the whites, Red loses in relative 'pull' over Yellow; as in the fourth grade it is nearly three times as strong as Yellow, but in the last grade subgroup it is twice as strong only. This is due to the fact that Red was a diminishing factor all along the above mentioned grade scales. The Red started with the whites with a 'pull' one and one half times that of Violet, and ends up with a little more than half the 'pull' of Violet. The fact is, Red always 'pulls' harder in the Indian Scale relatively, as compared with Violet, than it does in the white Scale. The situation of relative 'pull' of Red and Blue as influenced by education, thus indicated, is this for the two races: for Indians, in term of 'pull' for the lowest grade, Red is to Blue as 100 is to 71, but in the highest grade subgroup it is 100 to 102; and for whites, in lowest grades Red is to Blue as 100 is to 110 and for highest grade subgroup it is 100 to 404; that is Red lost 73 per cent., almost three fourths of its 'pull' here. An examination of the graphic representation of the scales will make plain these and similar relations. In addition the scales indicate that as we go along up the educational scales, taking both tables and graphic representations as guides, the whites lose somewhat their appreciation of all colors but Blue, which seems to be very highly valued, being more so with the highly educated, but the full-blood Indians appear to experience no such 'repression' of color preference, on account of education (4).

The Question of Background

In the treatment of the data we have had nothing to say of the effect of background which of course has played its part and whatever has been said of a color's preference is to be taken of that color as on a White background. When we have said 'Red' we might have said 'Red-on-White,' etc. To have given the experiment with colors having no background would have been impractical as a field enterprise.

In justification of the use of the White background we note that Indians use much of it in decorative work, as in feather designs, bead-work, blankets, pottery, etc. These show Red, Green, Brown, Black, etc., on White. But we rarely see the White with absolutely no colorative design. The writer recalls an almost white Navajo blanket with a very few small red and black squares to rejoice the eye. Another blanket was immaculately white to begin with, but on this a representation of an Indian ceremonial of more or less sacred significance was decorated with a procession of human (Indian) figures garbed in bright red and green uniforms, and the whole was set off by a few other colors much less conspicuous. While for the Indians White may be esteemed as a background, the experimental results show that just blank background is not what the Indian, or the white individual, desires.

Summary

1. There is for full-blood Indians a rather definite sequence of color preference for all the seven colors used in this experiment excepting Yellow and White, in which cases the preference for the former over the latter is not very conclusive. There is likewise a definite tendency for a preference sequence

of these seven colors for the whites, but the tendencies as determined for both races, by overlapping and scale positions are not so strong for them as for the full bloods. For the mixed bloods the sequence is even less secure as indicated by overlapping. For the three racial groups the sequences of the color preferences are:

	Rank						
	1	2	3	4	5	6	7
Full Bloods	Red	Blue	Violet	Green	Orange	Yellow	White
Mixed Bloods	Blue	Red	Violet	White	Green	Orange	Yellow
Whites	Blue	Green	Red	Violet	Orange	Yellow	White

From a statistical standpoint, as well as from the standpoint of judgment scales, the data indicates race differences in color preference, between full bloods and whites, and, by the test of overlapping, between the full bloods and the mixed bloods, but the differences are not so great for them as thus indicated.

2. Since the sequences of the colors are almost identical in all Indian educational subgroups, it may be said that this factor, education, has very little influence over the color preference with the Indian. In consequence we may say that for them their preference tendencies, on the whole, are stubbornly native.

3. Full-blood Indians are aware of more decided preferences, in general, than are the whites, in judging colors, if we abide by the objective results, disregarding the two exceptions, *i.e.*, the latter's high esteem for Blue and clear-cut discrimination between White and Yellow.

4. Among the Indians there are no outstanding differences between the sexes to the extent of altering the race differences. What slight tendencies there are for differences in this respect may be accounted for by the immediate influence of education.

References

1. Winch, W. H. The Color Preferences of School Children. *British Journal of Psychology*, 1909, 3, 42–65.
2. Jastrow, Joseph. *Popular Science Monthly*, 1897, p. 361 ff.

3. WASHBURN, MARGARET. *American Journal of Psychology*, Vol. 22, 114.
4. DASHIELL, J. F. Children's Sense of Harmonies in Colors and Tones. JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 1917, 466–475.
5. WELLS, F. L. Variability of Judgments, Essays in Honor of Wm. James, 1908, 548 ff.
6. HOLLINGWORTH, H. L. Experimental Studies in Judgment. *Archives of Psychology*, p. 29.
7. STRONG, E. K., JR. The Relative Merits of Advertising. *Archives of Psychology*, No. 17.
8. THORNDIKE, E. L. Handwriting. *Teachers College Record*, Vol. II., No. 2.
9. GARTH, T. R. How College Students Prepare Their Lessons. *Pedagogical Seminary*, Vol. XXVII., No. 1, p. 90.
10. THORNDIKE, E. L. *Mental and Social Measurements*, p. 228.
11. HILLEGAS. *A Scale for the Measurement of Quality in English Composition by Young People*, pp. 58–60.

A SIMPLE VOICE KEY

BY F. L. WELLS AND J. S. ROONEY,
Boston Psychopathic Hospital

I

The importance of the language mechanism to human behavior gives a special significance to means of reducing it to quantitative terms. Of particular interest are its time relations, as denoting the time relations of more or less complicated mental processes. Graphic methods of studying these functions leave little to be desired from the standpoint of precision. The technique is so elaborate that their application is limited on practical grounds. The more frequent endeavor in psychology has been to get some function of the beginning speech process to operate an electric circuit. Such devices are known generally as 'voice keys.'

The movement that is most generally characteristic of a beginning speech response is an operation of the diaphragm and neighboring muscles, increasing the pressure of the air in the lungs. This will give an easily readable change in the pneumograph curve, but so far as the writers know it has not been used to operate an electric circuit. It is open to the objection that this movement may occur in anticipation, before the mental process it is desired to measure. Such a premature reaction would be difficult to recognize and control.

Another important factor in speech reaction is the opening of the lips. This does not characterize the so-called 'labials' as *f* (labiodental), *b* and *m*, which are formed with the lips closed. It is, however, sufficiently characteristic to have given rise to a special instrument, the lip key. Two arms are held together by the lips, on opening the lips a spring separates the arms, and operates an electric contact. This instrument has a disadvantage already mentioned in connection with the pneumograph. Cattell long since observed that the lips might be opened before the word was 'propositionized.' Hygienic and æsthetic objections are not impossible.

A third factor in speech reaction is the expulsion of breath. This takes place initially in all words save those that begin with 'explosives' (*e.g.*, *p*, *t*, *d*). If the expulsion of air is sufficiently strong, it may be utilized to operate an electric circuit. Its force varies greatly in different phases of speech. It is greatest upon the release of the so-called 'voiceless explosives' (*e.g.*, *p*, *t*, *k*). It is less effective with the voiced explosives (*e.g.*, *b*, *d*). It is comparatively ineffective during the enunciation of other consonants, (*e.g.*, *f*, *l*, *m*). Upon the release of such consonants, however, there may be a more or less effective puff of air by which a sufficiently delicate device of the Cattell or Dodge type is operable. Contact of the subject with the device is apt to be necessary.

A fourth factor in speech is the vibration of the vocal cords. This is present in the vowels, and in the so-called 'voiced' sounds (*e.g.*, *d*, *v*, *m*). It is absent in the 'voiceless' sounds of normal speech, *e.g.*, *t*, *f*, *s*, as it is in whispering. Sensitive devices of the Cattell or Dodge type are able to respond also to the vocal cords, but are probably more effectively operated by the breath stream. A device utilizing the laryngeal vibrations is that introduced by Römer and improved by Dunlap. Aside from sensibility an outstanding advantage of the Dunlap key is that the face is not touched or apparently confined in any way. The subject merely speaks against a metal disk. Dunlap's is the most highly specialized instrument for voice reactions that has been offered. Voice keys of the third and fourth classes are less subject to error through premature reaction than those of the first two. Premature vocalization or release of breath strong enough to operate the voice key is less likely to be undetected.

With regard to these and similar devices, it is appropriate to keep in mind the coördination between the subjective process it is desired to measure and the physiological process through which it has to be approached. Keys of the Cattell type are not readily operated until a free breath stream issues. Keys of the Dunlap type are not readily operated except by free vocalization or a forcible breath. The time between this event and the beginning of the actual speech reaction is included in the measure obtained. The relation between higher mental processes and their linguistic expressions is not always as clear cut as between corresponding mental processes and the operation of telegraph keys. In ordinary usage, the voice key measures a single phase of a particularly complicated motor response. Its use is justified in that only through language and not telegraph keys are specially complicated mental processes workably expressed.

II

The John Hugo Manufacturing Company of New Haven, Conn., market a toy called the 'Radio Rex,' retailing at $2.00. It is made to represent a kennel, containing a little dog. The back of the kennel mounts a small dry cell and an electromagnet, with a pivoted armature, the magnet pulling against a spring. When the dog is pushed back into the kennel the armature is pressed against the electromagnet. When the armature touches the core of the electromagnet, the circuit through the magnet is completed, and the armature is held back, with the stretched spring pulling against it. With the breaking of the circuit, the armature, which is a long lever, kicks forward, causing the dog, in effect, to leap out of the kennel.

The device for breaking the circuit is as follows: on one side of the kennel a small triangular metal plate is suspended from a simple knife edge. In normal position this plate makes contact along its bottom edge from one copper wire to another. These parts are a portion of the circuit operating the electromagnet. A slight disturbance of the suspended plate breaks the circuit between the wires, and releases the armature. The intended source of such disturbance is calling at the apparatus in a fairly loud voice. If, for example, the name assigned to the dog, 'Rex,' be called, the dog can be made to appear.

It is plain that such a device is readily adaptable as a voice key as well as to any other purpose where a circuit is to be operated by a slight mechanical shock. One may arrange a secondary circuit through the armature, which will be broken when the armature is released. In the present instance, this was accomplished by mounting a small copper brush on the base holding the magnet and accessories. When the armature touched the magnet, it also made contact through the brush, which was held against the armature by its natural resilience. With the release of the armature, it was drawn away from the brush by its spiral spring. Incidentally, the brush was dispensed with and both circuits operated between the magnet core and armature. Under these conditions the small dry cell was also eliminated and the entire apparatus operated by the laboratory current. Latency under these latter conditions averaged about 7 sigma, m.v. about 1.5 sigma.

The suspended plate and accessories, the voice key proper, were remounted upon a wooden arm some two feet long which, attached to a standard, could be brought into any convenient position with reference to the subject. A sort of wooden funnel was put around the key to assemble the sound waves but it is not an important accessory. What is gained in this way is probably lost through keeping the subject further away from the apparatus. As with the Dunlap key, the subject touches no portion of it.

Sensibility is conveniently varied by tilting the key and varying the weight with which the plate rests against the copper wires. It can be made more sensitive than is convenient to use, as it trips the armature at odd times, such as the fall of an exposure screen. In ordinary use its standard rests upon the floor of the building, from which it is separated by some cotton wadding. At ordinary working sensibility, a vigorous handclap seven inches from the plate tripped the release eleven times out of twelve. A loud shout nine feet away from the apparatus tripped the release three times out of four. A one gram weight dropped from a height of 4 cm. on a point of the wooden arm 19 cm. distant from the key tripped the release fifteen times out of twenty.

Under experimental conditions the conversational tone of voice cannot be relied on to operate the key. The best operation is secured by instructing the subject to 'call' the response into the device. Some preliminary trials to determine a suitable loudness are desirable. A tone such as one might use talking to a person a hundred feet away is generally suitable. These considerations are sufficient to limit its applicability to controlled association responses. The key operates much more positively with men's voices than with women's.

The key is operable by either voice or breath, but the effective factor, in ordinary experimentation, seems to be the voice. It is risky to use a voice key without some knowledge of how it may be affected by the quality and intensity of the sounds spoken into it. In this connection the following observations are offered:

A series of seven three-letter words was selected, beginning with as many typical sounds, namely *egg* (beginning with a vowel), *gun* (voiced explosive), *hit* (aspirate), *man* (nasal), *peg* (voiceless explosive), *set* (voiceless spirant), *the* (voiced spirant). Experiments were made as follows: first one of the words was exposed, the subject being requested to read it. Then the same word was exposed, the subject being requested to have his organs of articulation all set therefor and to call the word as a *simple reaction* at the first intimation of the fall of the screen. Then the same stimulus was given, the subject making a simple hand reaction with a telegraph key at the first intimation of the exposure. The seven words were given in irregular order until the cycle of observations had been completed five times. Eleven records of such experiments are available, three with one subject. There are thus 55 reading reactions for each word, 55 simple reactions with each word, and 385 simple hand reactions, 1,155 observations in all.

The reading times, and the simple vocal reaction times, are presented in terms of their *excess* over the simple hand reaction. All time units are sigma. All reactions were measured by a ballistic galvanometer of about 50 seconds double swing period, operating as a chronoscope. Measured instrumental latencies are deducted from the readings quoted.

The average hand reaction time in these experiments is 144, m.v. 21. The average excess over this of a simple vocal reaction time is 117 sigma. The average excess for each of the seven words ranges from 112 to 117.2 with all words but *set*, whose excess averages 144. There was obtained for each subject the mean variation of his five simple reactions with each word. The average of these quantities is 34. The average m.v. of the simple hand reactions is 21. It will be remembered that the absolute length of the simple hand reactions is a little more than half that of the simple voice reactions. It would seem that, in general, no special attention need be paid to the initial sound of a word, with voice keys of this type. As represented by the initial of *set* the voiceless spirant gives a somewhat increased time. It is well known that such sounds are ineffective over the telephone. Compound consonants, as in the word *string* for example, would be likely to still further delay the action of the key.

Tendency of the key to trip more promptly when the reaction is in a loud voice involves a source of greater error. Ordinary experience finds the apparatus delaying to break when the response is spoken comparatively low. Two subjects fairly shouted into the apparatus, and their simple voice reactions are much the shortest; but 28 and 37 sigma larger than their respective hand reactions. Their reading reactions also show the least excess. It was impracticable to standardize vocal intensities, but the following observations bear upon the point.

The words 'yes' and 'no' were used as reactions to a sound stimulus. Twenty reactions with each were made in a tone only sufficient to trip the release, and twenty shouted. The comparison is as follows:

Subject A						Subject B					
Yes			No			Yes			No		
	Av.	M.v.		Av.	M.v.		Av.	M.v.		Av.	M.v.
Loud....	161	19	Loud.....	165	14	Loud.....	202	22	Loud....	192	16
Soft.....	258	41	Soft......	249	26	Soft......	268	26	Soft.....	270	36

Some of this difference might be due to more concentrated attention in the case of the 'loud' responses. In the 'soft' responses some attention seems directed to regulating the intensity of the speech sounds. Observations were made through tripping the release by striking a telegraph key adjacent to the voice key. Twenty trials were made with a light blow just sufficient to ensure breaking of the contact. Average and m.v. of these were 29 and 1.5 sigma respectively. Ten trials were made with a smart blow, as hard as could conveniently be struck. Average and m.v. were respectively 27 and .9 sigma. In this case it would appear that the intensity of the mechanical disturbance does not materially affect the quickness of the break. Its quickness was tested with maximal shouts immediately before the apparatus, and at a distance of 6 ft. Ten observations each showed results as follows:

	Subject *A*		Subject *B*	
	Adjacent	6 ft.	Adjacent	6 ft.
Av.	189	259	184	247
M.v.	22	31	21	48

There is seen to be a marked difference in favor of the greater commotion of the voice key by the closer shout. Some of the difference might be due to lessened intensity of the stimulus at the greater distance. Hand reactions were taken, indicating that this factor was relatively ineffective in the following results:

	A		*B*	
	Av.	M.v.	Av.	M.v.
Near	117	13	131	14
Far	123	14	114	08

It would appear that differences of the order of 70 sigma are producible by variations in the intensity of the vocal reactions. Single shocks do not show such difference according to their strength, and the appearance is that a feeble commotion, when continued, 'summates' to an effect sufficient to break the contact. In ordinary use, unaccented portions of polysyllables may hardly be expected to operate the key.

If one recalls the phonetic conditions governing voice keys, one can get some idea of how absolute excess time in simple vocal over hand reactions may be used up. In the case of an initial vowel, the diaphragm must get up sufficient air pressure to operate the vocal cords. For a voiceless explosive (such as *p*) pressure should be already present behind the occlusion but the lips will give slightly before parting. In the case of explosive sounds the length of the neural arcs is more favorable to voice than to hand reactions. In the *p* sound of English there is also a slight unvoiced interval after the explosive before the cords operate, which would further delay the key unless it were operated by the puff of air. Other consonants are similarly ineffective until released, and in the case of voiceless spirants or compound consonants this effect is maximal.

Data are not available to the writers to indicate how these properties in the more specialized voice keys compare with the present simple one. The sensitiveness claimed for the former is not claimed for the Rex. In general serviceability the Rex's simplicity, ease of manipulation and economy make it more nearly equal.

There may be mentioned a few observations made with a breath key, so arranged with a tambour that a slight increase of pressure broke an electric circuit. To get this pressure under the condition of speech, it was necessary to confine the breath stream with a mouth-piece, nor was satisfactory operation obtained except with initial voiceless explosives. On the other hand, both reading and simple vocal times were very much shorter with this device. Vocal reactions with the breath key are of the order of hand reactions in shortness, and the reading reactions are from 100 to 200 sigma shorter than those observed with the 'Rex.' Comparison was made of initial *p* and *k* sounds with evidence that the latter provided the more prompt release. (The *p* sound is, however, more positive in breaking the circuit.) In general, it seems that a key operated by the vocal cords, while especially convenient to manipulate, may include as much as 200 sigma of processes that are not strictly mental. A key operated by breath times the mental process more closely under favorable conditions, but for considerations already named, its use is restricted.

References

1. Cattell, J. McK. Psychometrische Untersuchungen. *Philos. Studien*, 1886, 3, 305–335.
2. Dodge, R., and Benedict, F. G. Psychological Effects of Alcohol. Carnegie Institution of Washington, 1915, pp. 97–101. (Extract.)
3. Dunlap, K. Apparatus for Association Timing. *Psychol. Rev.*, 1913, 20, 250–253.
4. Dunlap, K. The Stopwatch and the Association Test. *Psychobiology*, 1917, 1, 171–175.
5. Dunlap, K. An Improvement in Voice Keys. Journ. Exper. Psychol., 1921, 4, 244–246.
6. Römer, E. Beiträge zur Bestimmung zusammengesetzter Reaktionen. *Psychol. Arb.*, 1896, 1, 566–607.

RESULTS OF VARIATIONS IN LENGTH OF MEMORIZED MATERIAL

BY EDWARD S. ROBINSON AND WILLIAM T. HERON

The University of Chicago

I. Problems, Methods, and Subjects

In planning the present experiment we were primarily interested in discovering what type of relationship exists between the length of memorized material and its susceptibility to retroactive inhibition. Our data, however, also bear very directly upon three other fundamental problems of memory: The nature of the relationships (1) between the length of material and the form of the curve of memorizing; (2) between the length of material and the difficulty[1] of memorizing it; and (3) between the length of material and the amount of it retained after a certain interval. These other problems will be discussed first when we come to consider our results.

Ten subjects served in this experiment. They were graduate and advanced undergraduate students in experimental psychology.

The materials employed were lists of nonsense syllables. Each list contained 6, 9, 12, 15, or 18 syllables. The syllables were exposed by means of a hand-operated memory drum which has been a standard instrument in the Chicago laboratory for several years. The time for the single exposure of a syllable was 2 seconds. The experimenter was made aware of these intervals by means of a mercury cup metronome in

[1] Difficulty being measured in terms of repetitions or time required for complete mastery.

circuit with a head-phone. The metronome was not, of course, in the experiment room.

The *anticipation*[1] method was employed. The entire list was presented once and, after that, the subject, upon being presented with a given syllable, attempted to pronounce the three letters of the next syllable in the list before that syllable appeared. He was not permitted to anticipate aloud more of the list than that syllable following the one which was exposed. Luh made use of the method of having the letters of a syllable, rather than the syllable itself, pronounced, and it is obviously a safeguard against errors of enunciation being mistaken by the experimenter for errors of memory. Each list was presented until the subject had successfully anticipated during a single presentation all but the first of its syllables. The first syllable served merely as a cue. In the learning of a list, then, it may be considered that one less association was formed than there were syllables in that list. The number of presentations necessary for complete learning was taken to be one less than the number of presentations of the list. The correct presentation was a test rather than a practice trial.

We were interested in the retention of the lists as well as in the manner in which they are learned, *i.e.*, the amount of the material recalled at different stages in the memorizing process. This retention was measured in terms of a *recall* score and a *saving* score. Both types of score were obtained for each list learned.

The *recall* score is the percentage of a memorized list recalled when that list is given a first re-presentation some time after the completion of those presentations required for the initial learning of the list. This score can be expressed as follows:

$$\text{Recall Score} = \frac{\begin{array}{c}\text{Number of syllables correctly antici-}\\ \text{pated during first re-presentation of list}\end{array}}{\text{Number of syllables in list} - 1^{2}} \times 100$$

[1] Used by Finkenbinder, E. O., 'The Curve of Forgetting,' *Amer. J. of Psychol.*, 1913, 24: 8–32, and by Luh, C. W., 'The Conditions of Retention,' *Psychol. Monog.*, 1921, 30.

[2] The first syllable which, as we have noted, was merely a cue.

The *saving* score expresses the relationship between the number of presentations required to learn a given list and the number of presentations required to relearn the list some time after the completion of the first learning. In detail this relationship can be expressed as follows:

$$\text{Saving Score} = 100 - \left[\frac{\text{Number of presentations required for relearning}}{\text{Number of presentations required for original learning}} \times 100\right]$$

The subjects were requested to refrain from devising or utilizing any mnemonic schemes to aid them in memorizing.

Before the experiment proper, each subject learned one list of 12 syllables a day for a period of six days, and then relearned the same list after an interval of 15 minutes. This was done to decrease the effect of practice during the experiment. The results of our work indicate, however, that the subjects were still improving in their ability to memorize this kind of material until the end of the experiment. This was to be expected and we guarded against this disturbing effect by arranging the order in which the different subjects went through the different conditions of the experiment in such a way that these conditions were affected by approximately equal amounts of practice. There were ten experimental conditions and ten subjects, and each subject went through these conditions in a different order. This device did not exhaust the number of possible orders, but it insured each condition being first, second, and so on an equal number of times in the series of conditions.

Table I. displays the plan of our experiment.

It can be seen that the results of such an experiment would give data bearing upon the retroactive effects of a constant activity—the memorizing of a list of 12 syllables—upon previously memorized lists of varying length. The normal or rest interpolation was, of course, rest in only a relative sense. It was from this same experimental situation that our data on length and the memory curve, length and difficulty, and length and retention were obtained.

Our experiment contained three cycles. That is, each of our subjects was tested three times under each experimental condition. This gave us a total of 30 values apiece for each

TABLE I

O = Original memorizing of a list of nonsense syllables, 6, 9, 12, 15, and 18 designating the number of syllables in the list.

I = Memorizing of a list of nonsense syllables immediately following the memorizing of a first or original series, and 12 designates the length of this interpolated list, *i.e.*, 12 syllables.

Rest = Casual reading of the daily paper with no attempt to memorize its contents. During this period the subjects refrained from rehearsing O or I.

RO = Relearning of original list of syllables. The interval between the completion of O and the beginning of RO was always 20 minutes. In the case of the practice series, it was 15 minutes, but this was not long enough to make sure that all subjects would complete I.

Condition		Interpolated Period of 20 Minutes	
I	O 6	I 12, Rest	RO
II	O 6	Rest	RO
III	O 9	I 12, Rest	RO
IV	O 9	Rest	RO
V	O 12	I 12, Rest	RO
VI	O 12	Rest	RO
VII	O 15	I 12, Rest	RO
VIII	O 15	Rest	RO
IX	O 18	I 12, Rest	RO
X	O 18	Rest	RO

of these conditions. This would hardly have given statistical accuracy except for the fact that our conditions were arranged in a definite progressive series so that the consistency as well as the absolute magnitudes of differences could be considered in drawing conclusions. In the case of a number of issues we were able to include 60 values under each condition. For instance, in considering the *difficulty* of lists of different lengths, we could include under a single condition the original lists of 6 syllables learned either before rest or before learning an interpolated list.

In no case was a subject tested under more than one condition on a single day. In most cases four tests were made during a week.

II. Length of Material and the Memory Curve

When material is learned by the anticipation method, a record is secured, for each presentation, of the amount that can be recalled (anticipated) up to that point. We have 60 sets of such records for the memorizing of each length of list employed in our experiment. From these data we have plotted the memory curves for the various lengths of material.

In Fig. 1 we have plotted the absolute number of syllables recalled or anticipated against the absolute learning time. The unit for the latter is one presentation of one syllable. Using this unit, one presentation of a 12-syllable list requires as much absolute learning time as, for instance, two presentations of a 6-syllable list, or as two thirds of a presentation of an 18-syllable list. This is as it should be if we are interested in returns for absolute amounts of time.[1] The curves of the figure represent the averages for all subjects and all cycles. Each value on each curve is an average of 60 individual measurements.

Whatever examination is made of such matters as the correlation between *relative* amounts learned and *relative* amounts of learning time, this correlation between absolute amounts learned and absolute amounts of learning time should receive primary consideration. Because of the simpler character of the variables involved, this latter correlation should possess more fundamental importance.[2]

The rate of learning all lengths of the material decreases with successive presentations. The curves show a highly regular variation from that of the shortest list to that of the longest. The shorter the list the more associations are formed per absolute unit of time very early in the learning process, and the fewer are the associations that are formed later. The rates at which associations are formed in the shorter materials are exceeded by the rates at which associations are formed in the longer materials only as the rate

[1] Professor L. L. Thurstone, of the Carnegie Institute of Technology, suggested this convenient method of measuring learning time in terms of syllables presented.

[2] Kjerstad seems to have overlooked this fact. He considers only the former type of correlations. See Kjerstad, C. L., 'The Form of the Learning Curve for Memory,' *Psychol. Monog.*, 1919, 26.

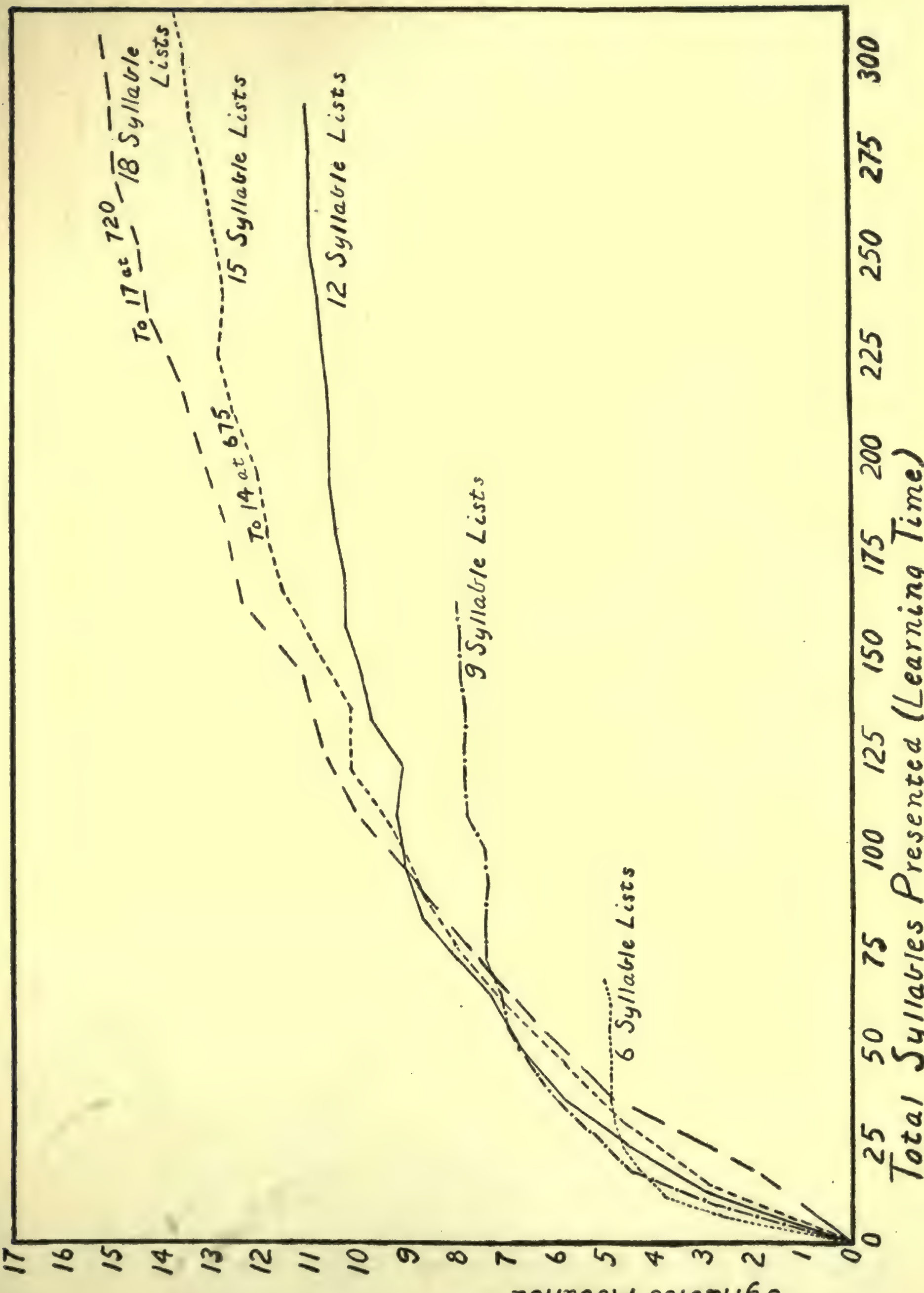

Syllables Recalled
Total Syllables Presented (Learning Time)
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
25
50
75
100
125
150
175
200
225
250
275
300
6 Syllable Lists
9 Syllable Lists
12 Syllable Lists
15 Syllable Lists
To 14 at 675
To 17 at 720 — 18 Syllable Lists

FIG. 1

of learning the shorter materials decreases with the approach of complete learning. The fact that the absolute rate of learning longer materials eventually exceeds that of shorter materials has an obvious explanation in the fact that the limit of possible associations in the case of the shorter materials is reached first. But the initially more rapid rate of learning for the shorter materials has no such obvious explanation. Why should two presentations (12 units of learning time) of a *6-syllable* list result in the formation of a greater number of associations than one presentation (also 12 units of learning time) of a *12-syllable* list?

A possible explanation of the fact before us is as follows: How much will be learned in a given period of time is affected in two almost contradictory ways by the length of the material before us. If there is a very small amount of material before us we can naturally learn only a small amount in a given period of time. In order to learn the maximum amount, then, in the time at our disposal, it is necessary to have at least as much material to work upon as is possible for us to learn. Probably we can profitably work with a larger amount of material than we can learn. There may be, as has been suggested, a tendency to learn easier things first. In that case we should, for maximum results, have enough material before us to give us some range of choice as to what is to be learned. (This choice need not be explicit.) But it is probable also that we may have too much material before us for best results. In that case we may suffer, not because of an exhaustion of the possibilities for association, but because there are so many possibilities that some of these interfere with those associations which might easily be fixated under ordinary conditions. Although this seems to us the most feasible explanation of the slower rate of learning for the longer lists, it is not the only one. Consider the fact that two presentations of a 6-syllable list give greater returns than one presentation of a 12-syllable list. Factors which might be effective here are the difference in study time per syllable under the two conditions, and other differences in the methods of study; *e.g.*, an element of repetition enters into two presentations of 6 syllables which does not enter into one presentation of 12 syllables.

Turning to the interpretation of our curves, we may suppose that with all of our lists there was, during the early period of the learning, more than the minimum amount of material necessary for getting the maximum results out of that early period. It is also likely that the longer the lists the more possibilities they offered for interfering associations. After the earliest portion of the learning time, however, the exhaustion of the possibilities of the shorter list more than compensates for the possibilities for interferences within the longer lists, and the absolute rates of learning become more rapid for the longer materials.

We might summarize the curves of Fig. 1 by saying that, within the limits of our experiment, the shorter the material the greater the rate of learning, until the possibilities for forming associations become considerably exhausted. As we have pointed out, there may be lists shorter than our 6-syllable lists which would give a relatively slow initial rate of learning, but we did not have such lists in our experiment.

Up to this time we have been considering absolute amounts learned in absolute amounts of learning time. The curves of Fig. 1, because they approach different absolute limits at different absolute rates, are difficult to compare in a relative way. For this reason we have approximated Kjerstad's method[1] of dividing into the same number of parts the different lengths of time spent in learning the lists of different lengths and also of dividing into the same number of parts the total amounts learned in the case of the lists of different lengths. Kjerstad divided the time of learning into sixths and the amount learned into per cents. We have divided both time of learning and amount learned into per cents. The results are shown in tabular form (Table II.) instead of in a set of curves for the reason that the five curves happen to be almost coincident. In order to obtain these values, however, we plotted in percentile form curves representing the relationship between certain fractions of the total learning time and

[1] *Op. cit.*, pp. 25–27.

certain fractions of the amount learned. The values shown in Table II. were read off from these curves.

Table II

Percentages of Lists of Different Lengths Learned in Different Percentages of the Total Time Spent Learning Those Lists

Percentages of Total Learning Time	Length of List				
	6 Syllables	9 Syllables	12 Syllables	15 Syllables	18 Syllables
5	30	34	30	32	28
10	52	51	46	52	44
25	81	76	72	80	74
50	97	93	88	99	95
75	98	97	96	99+	98
100	100	100	100	100	100

These percentile figures (Table II.) confirm a contention made by Kjerstad and borne out by his results that the form of the memory curve is independent of the length of the material memorized. Remarkably similar proportions of the material to be learned are learned in similar relative stages in the learning processes. What differences occur seem to be unrelated, at least in any continuous fashion, to the matter of length.

III. Length and Difficulty

The relationship between the length of material and the number of repetitions required to memorize it has had considerable attention. The main body of evidence from experiments conducted by Ebbinghaus,[1] Meumann,[2] Henmon,[3] Lyon,[4] and Kjerstad,[5] indicate that longer materials require more repetitions for memorizing than do shorter materials. But there is considerable disagreement as to the rate at which repetitions must be added with additions to the length of the material. According to Ebbinghaus, Meumann, and Lyon,

[1] Ebbinghaus, H., 'Memory,' 1885, tr. by Ruger and Bussenius, 1913.

[2] Meumann, E., 'The Psychology of Learning,' 1912, tr. by Baird.

[3] Henmon, V. A. C., 'The Relation Between Learning and Retention and Amount to be Learned,' J. Exper. Psychol., 1917, 2, 476–484.

[4] Lyon, D. O., 'Memory and the Learning Process,' 1917.

[5] Kjerstad, *op. cit.*

early increases in length are accompanied by more marked increases in repetitions than are later increases in length. Henmon and Kjerstad did not obtain this result, due, perhaps, to the fact that the shortest lists used by them contained from 10 to 12 syllables, while the rapid increase in repetitions referred to occurred between lists of 8 syllables and those of 12 syllables.[1] This initial rise in the number of repetitions is usually such that there is an increase not only in *repetitions*, but also in *repetitions per syllable*. With further increases in the length of the material there continue to be increases in *repetitions*, but not necessarily in *repetitions per syllable*.

Our results on this point are given in Table III.

TABLE III

AVERAGE NUMBER OF REPETITIONS PER SYLLABLE OF LISTS OF DIFFERENT LENGTH. N FOR EACH VALUE = 20, EXCEPT FOR THE AVERAGES OF THREE CYCLES WHERE $N = 60$

Length of List	6 Syllables	9 Syllables	12 Syllables	15 Syllables	18 Syllables
Cycle I	0.67	0.90	1.11	1.20	1.21
Cycle II	0.62	0.82	0.85	0.98	0.99
Cycle III	0.45	0.77	0.79	0.80	0.73
Averages for Three Cycles	0.58	0.83	0.92	0.99	0.98

Two out of three cycles of this experiment agree in showing an increasing number of repetitions per syllable with increasing length. In all cycles there is a marked increase in repetitions per syllable as the length is increased from 6 to 9 elements. In all cycles, this increase grows less as the length is further increased. In one cycle, the third, there is an actual decrease in repetitions per syllable as the longest length is reached. These findings agree with the majority of those obtained with nonsense syllables in showing a sudden increase in the number of repetitions per syllable with the early increases in the length of the material, and a less pronounced increase in repetitions per syllable or, perhaps even a decrease, with later increases in the length of the material. Our data point more positively toward a continuous increase

[1] Kjerstad, *op. cit.*, p. 80, has given a table summarizing the findings of various investigators on the relation between length and repetitions.

in *repetitions per syllable* with increasing lengths of material than do those of other investigators. Whether or not this continuous increase is obtained may easily be a function of any one of a number of minor differences between the conditions of other experiments and our own. For instance, our results show this continuity only during the first two cycles which indicates that familiarity with this type of learning may have something to do with the matter. Weber and Knors are said to have obtained a continuous increase in repetitions per syllable only with wholly unpracticed observers.[1] While at no time during the actual experiment were our learners wholly unpracticed, the general direction of the two sets of results is the same.

The fact that later increases in the length of material do not always add to the required repetitions per syllable has at times been taken to mean that longer material can be memorized with relatively less effort than shorter material.[2] This would be true if *one repetition* stood for a constant amount of time or effort, but it does not. As a list of syllables becomes longer, each repetition requires that much more time or effort. If, then, one extra repetition is added for each added syllable, it means, since the added repetitions are longer and longer, that the effort required to master the material is increasing, not as its length, but faster than its length. It is undoubtedly difficult to think clearly of variations in effort in terms of such a variable unit as *one repetition.* Lyon and Kjerstad have both taken account of this fact and have shown the correlations between length and *time* taken to memorize. In considering these correlations in our own data and also in those of certain other investigators, we have again followed the procedure, described in our section on memory curves, of measuring *time* in terms of the total number of units of material presented.

In Fig. 2 we have plotted the relationships between length of material and units presented before learning is achieved for the data of Ebbinghaus, Meumann, and Henmon for

[1] See Meumann, *op. cit.*, p. 275.

[2] See Meumann, *op. cit.*, p. 274 ff., and Henmon, *op. cit.*

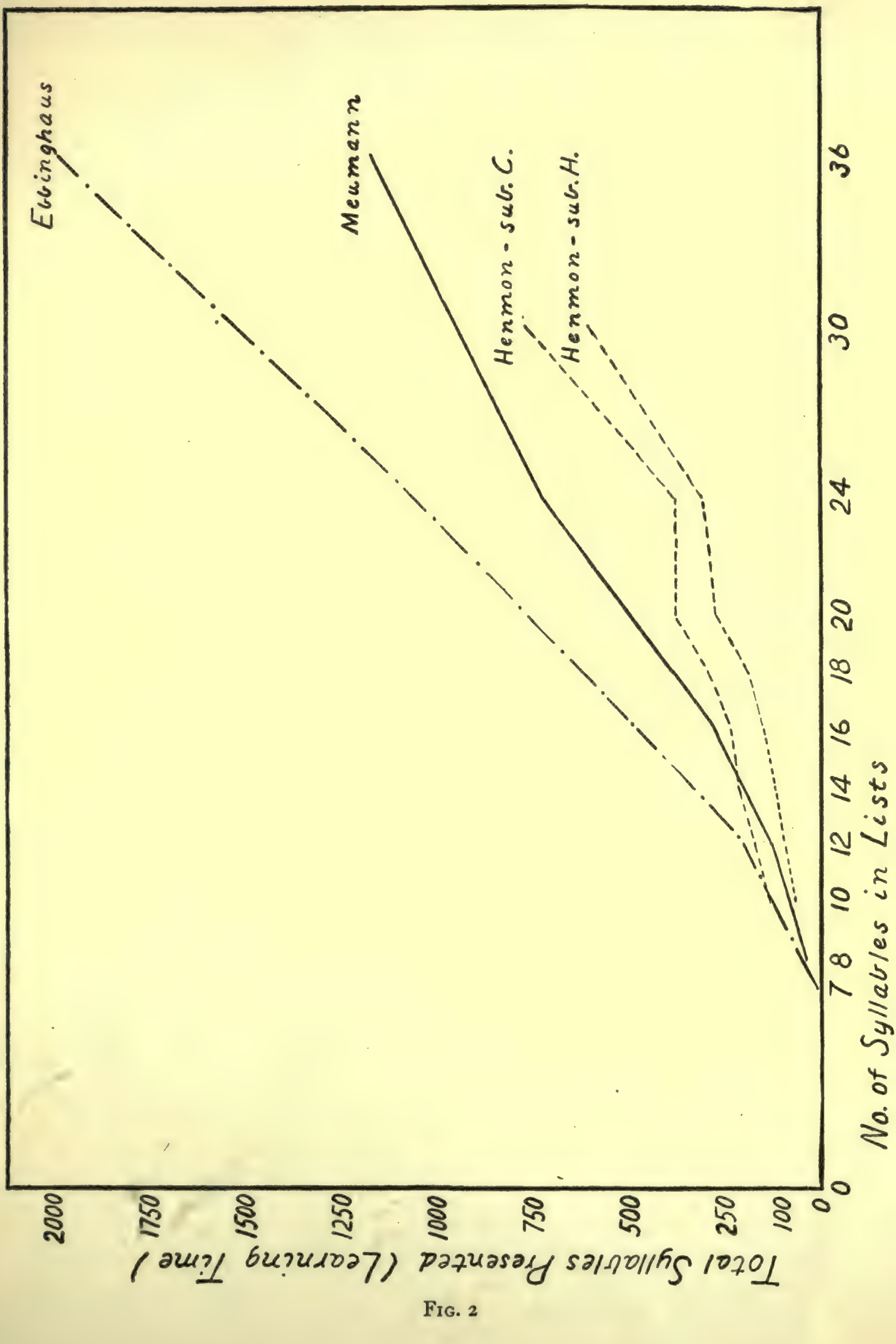

Ebbinghaus
Meumann
Henmon - sub. C.
Henmon - sub. H.
No. of Syllables in Lists
0 7 8 10 12 14 16 18 20 24 30 36
Total Syllables Presented (Learning Time)
0 100 250 500 750 1000 1250 1500 1750 2000

Fig. 2

nonsense syllables.[1] The two Henmon curves and that of Ebbinghaus show increasing amounts of time (total units presented) with increasing length, and this increase takes place at an increasing rate. The Meumann curve indicates a positive acceleration in the early portion of the time-length curve and a negative acceleration later. Lyon has plotted his data in similar fashion. For practically all of his material there are positively accelerated curves expressing the relationships between learning time and length. Binet and Henri[2] have also given such a relationship in the case of memory for numbers. In Table IV. we have presented the time-length relationships for Henmon's data on meaningful material.

Table IV

Av. of Three Subjects—1, 2, 3, 4, and 5 Stanzas of 'In Memoriam'

About 25 measures in toto for each condition.

Stanzas	Repetitions	Total Stanzas
1	3.5	3.5
2	6.3	12.6
3	8.6	25.8
4	10.0	40.0
5	12.2	61.0

One Practiced Subject—1, 2, 3, 4 Stanzas of 'Childe Harold'

Stanzas	Repetitions	Total Stanzas
1	8.3	8.3
2	9.3	18.6
3	10.0	30.0
4	9.1	36.4

Nine Unpracticed Subjects—'Childe Harold' (?)

Stanzas	Repetitions	Total Stanzas
1	4.6	4.6
2	9.7	19.4
3	11.8	35.4
4	15.9	63.6
5	15.3	76.5

One Subject—100-, 200-, or 300- Word Passage from Arnold and Huxley

Words	Repetitions		Total Words	
	Arnold	Huxley	Arnold	Huxley
100	5.5	7.4	550	740
200	6.7	7.9	1,340	1,580
300	6.7	7.4	2,010	2,220

[1] Such curves are roughly comparable as to general form even though we do not stop to inquire whether *one presentation of one syllable* means the same in terms of *time* from experimenter to experimenter.

[2] Quoted by Henmon.

(Henmon, himself, gives this relationship in terms of repetitions-length only.) The curve for "In Memoriam" would show a continuous positive acceleration, those for "Childe Harold" would be first positively and then negatively accelerated, and those for the prose material from Arnold and Huxley would have a slight negative acceleration.

Figure 3 shows the time-length relationship for our own data. The average curve and the curves for two out of three of the cycles are positively accelerated. The one for the last cycle is first positively and then very slightly negatively accelerated. It may be that this negative acceleration, appearing as it does in the last cycle, is in some way a product of familiarity with this type of material. There is general agreement among our own data and those of other investigators that the time-length relationship has an early positive acceleration. In some cases this is continuous to within the limits of the experiment, in some cases it is followed by linearity, and in some cases it is followed by a negative acceleration.

But the typical thing is the early positive acceleration, or increase in relative difficulty with increasing length. In no case has early negative acceleration, or decreasing relative difficulty, been demonstrated. This fact must be due, in part at least, to the particular lengths of material studied. But we should not expect any continuous negative acceleration, because that would ordinarily imply that after a certain point in length is reached, additional increases in length would not be accompanied by additional increases in difficulty. It seems a sensible expectation that in its later stages the time-length relationship will be linear or positively accelerated and that any negative accelerations which have appeared are temporary.

IV. Length and Amount Retained

There are not a great many results on the relationship between length of material and the rate at which it is forgotten. The evidence is probably somewhat in favor of slower forgetting for longer materials.[1]

[1] See Henmon, *op. cit.*, for a review of evidence on this point.

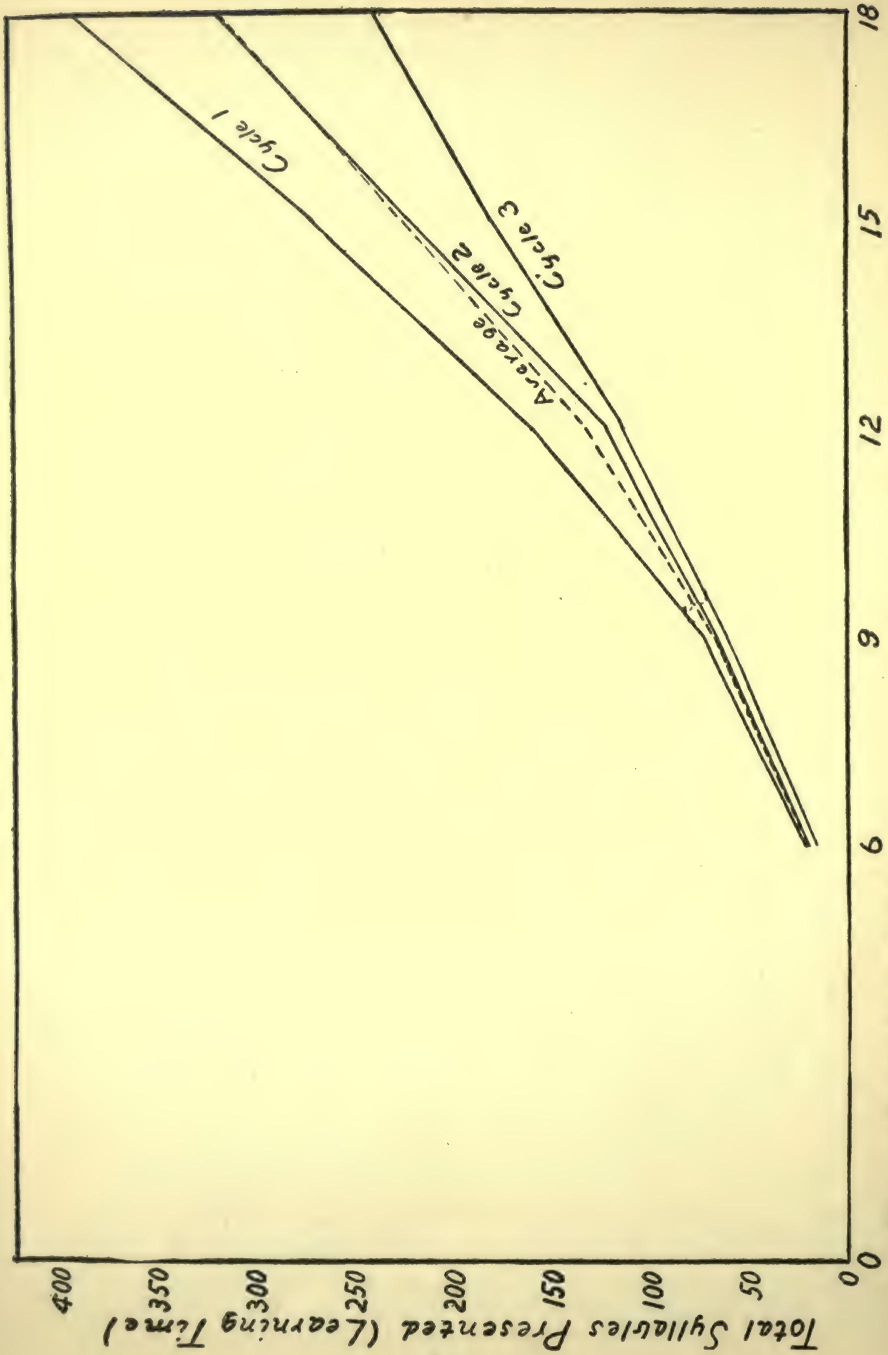

Cycle 1
Cycle 2
Cycle 3
Average
No. of Syllables in Lists
0
6
9
12
15
18
Total Syllables Presented (Learning Time)
0
50
100
150
200
250
300
350
400

FIG. 3

Table V. gives the relative amounts *recalled* and Table VI. the relative amounts *saved* of the lists of different length after a period of 20 minutes spent in newspaper reading. The values are averages of the performances of all of our subjects.

TABLE V

PER CENTS. *Recalled* AFTER 20 MINUTES

n = 10 for each value except averages where n = 30

Length of List	6 Syllables	9 Syllables	12 Syllables	15 Syllables	18 Syllables
Cycle 1	60.0	71.2	75.4	74.2	85.3
Cycle 2	70.0	78.7	79.0	76.2	79.4
Cycle 3	84.0	85.0	81.8	80.7	80.5
Average of 3 cycles	71.3	78.3	78.7	77.0	81.7

TABLE VI

PER CENTS. *Saved* AFTER 20 MINUTES

n = 10 for each value except averages where n = 30

Length of List	6 Syllables	9 Syllables	12 Syllables	15 Syllables	18 Syllables
Cycle 1	51.3	74.4	80.7	84.2	88.1
Cycle 2	71.4	78.5	71.5	80.8	88.6
Cycle 3	83.3	83.5	82.2	77.2	82.1
Average of 3 cycles	68.7	78.8	78.1	80.7	86.3

In general these results show a relatively more rapid forgetting, or poorer retention, for the shorter lists. It is only in Cycle 1, and according to the *saving* method, however, that this relationship is continuous. According to both methods of scoring, the relationship is if anything reversed by the third cycle. Apparently, practice in memorizing nonsense material may at least decrease the relative rate of forgetting shorter materials. This possibility should be further investigated.

V. LENGTH AND SUSCEPTIBILITY TO RETROACTIVE INHIBITION

The degree of retroactive inhibition has been found to vary with the similarity between the originally memorized material and the interpolated activity or *work*,[1] with the

[1] Robinson, E. S., 'Some Factors Determining the Degree of Retroactive Inhibition,' *Psychol. Monog.*, 1920, 28.

degree of learning,[1, 2] with the conditions of learning,[3] with the subjects employed,[4] and with certain other factors.[5] The present experiment affords data bearing on the relationship between the length of memorized material and its susceptibility to retroactive inhibition.

Table VII. shows the differences between the percentages

TABLE VII

DIFFERENCES BETWEEN PERCENTAGES OF DIFFERENT LISTS *Recalled* UNDER *Rest* AND *Work* CONDITIONS—ABSOLUTE INHIBITION

n = 10 for each value except averages, where n = 30

Length of Lists	6 Syllables	9 Syllables	12 Syllables	15 Syllables	18 Syllables
Cycle 1	8.0	33.7	30.0	26.4	29.5
Cycle 2	6.0	38.7	44.5	28.6	22.4
Cycle 3	36.0	46.3	36.4	32.2	15.3
Averages of 3 cycles	16.7	39.6	37.0	29.1	22.4

PERCENTAGES ABOVE DIFFERENCES ARE OF AMOUNTS *Recalled* UNDER *Rest* CONDITIONS—RELATIVE INHIBITION

n = 10 for each value except averages, where n = 30

Length of Lists	6 Syllables	9 Syllables	12 Syllables	15 Syllables	18 Syllables
Cycle 1	13.3	47.3	39.7	35.5	34.5
Cycle 2	8.5	49.1	56.3	37.5	28.2
Cycle 3	42.8	54.4	44.5	40.2	19.0
Averages of 3 cycles	21.5	50.3	46.8	37.7	27.2

of the original materials recalled after a *rest* (newspaper reading), during which the subjects abstained from rehearsing the original material, and after *work*, (after memorizing an interpolated list of 12 syllables). All of these differences are positive and indicate greater efficiency after interpolated *rest*. This table also shows the percentages which these differences are of the percentages of the lists of different length recalled

[1] Heine, R., 'Über Wiedererkennen und rüchwirkende Hemmung,' *Zsch. f. Psychol.*, 1914, 68, pp. 161–236.

[2] Robinson, *op. cit.*

[3] Tolman, E. C., 'Retroactive Inhibition as Affected by Conditions of Learning,' *Psychol. Monog.*, 1917–18, 25.

[4] Webb, L. W., 'Transfer of Training and Retroaction,' *Psychol. Monog.*, 1917, 24.

[5] Such as the method of measuring retention (see Heine, *op. cit.*)

after *rest*. In other words, we have here inhibition expressed in both absolute[1] and relative amounts. Table VIII. contains

TABLE VIII

DIFFERENCES BETWEEN PERCENTAGES OF DIFFERENT LISTS *Saved* UNDER *Rest* AND *Work* CONDITIONS—ABSOLUTE INHIBITION

n = 10 for each value except averages where n = 30

Length of Lists	6 Syllables	9 Syllables	12 Syllables	15 Syllables	18 Syllables
Cycle 1	3.6	16.4	15.6	14.8	12.0
Cycle 2	32.0	22.4	11.5	10.0	21.0
Cycle 3	40.0	25.9	23.7	16.7	15.2
Averages of 3 cycles	25.2	21.6	16.9	13.8	16.1

PERCENTAGES ABOVE CONDITIONS ARE OF AMOUNTS *Saved* UNDER *Rest* CONDITIONS—RELATIVE INHIBITION

n = 10 for each value except averages where n = 30

Length of Lists	6 Syllables	9 Syllables	12 Syllables	15 Syllables	18 Syllables
Cycle 1	7.0	22.0	29.3	17.6	13.6
Cycle 2	44.8	28.5	16.1	12.3	23.7
Cycle 3	48.0	31.0	28.8	21.6	18.5
Averages of 3 cycles	33.3	27.2	24.7	17.2	18.6

the same items as derived from the *saving* scores. Both tables are based upon the performances of our group as a whole.

Considering both tables, the most consistent tendency is in the direction of absolutely and relatively less inhibition with increasing length of the disintegrated material. The most marked exception to this generalization lies in the fact that in two out of three cycles according to *recall* and in one cycle out of three according to *saving* the least inhibition is present in the case of the shortest, *i.e.*, the 6-syllable, list. The other exceptions to the general tendency are not regular enough to be considered very seriously.

The general tendency toward decreasing susceptibility to retroactive inhibition with increasing length is one that

[1] Of course, the absolute amounts are themselves relative in the sense that they are percentages of the original materials.

might have been expected in light of the fact that the longer materials are forgotten more slowly or are, in other words, less susceptible to the disintegrating effects of agencies less inhibitory than the memorizing of an interpolated list of syllables. We found, however, that the more rapid forgetting of the shorter materials is probably a function of lack of practice in this type of memorizing. We might expect, then, that the greater susceptibility of the shorter lists to retroactive inhibition would also tend to disappear with practice, but it is at least as pronounced in the last cycle as in the first. It is also difficult in the light of our facts in regard to the rate at which the different materials were forgotten to explain why in certain cases the very shortest lists were least affected by retroaction.

With the facts at our disposal we can do little more than point out the general tendency toward decreasing susceptibility to inhibition with increasing length and the striking exception, rather too consistent to be accidental, supplied in certain cases by the material of shortest length. Evidently susceptibility to retroactive inhibition decreases with length only after a certain minimum length has been passed. It would be interesting to know whether this inverse relationship between susceptibility to retroactive inhibition and length has also an upper limit.

One theory in regard to retroactive inhibition explains this phenomenon as due to the transfer of elements of the original memorizing over to the situation of learning the interpolated list. This is, then, alleged to embarrass the recall of the original material.[1] If that is the case, we might expect to find the learning of the interpolated list favorably or unfavorably affected by the previous learning of an original list and we might also expect the amount of that effect to vary in some way with the amount of inhibition present for the lists of different length. We have therefore calculated the average number of repetitions required to memorize a 12-syllable interpolated list after learning 6-, 9-, 12-, 15-, and

[1] For discussions of the transfer theory or theories of retroactive inhibition see De Camp, J. E., 'A Study of Retroactive Inhibition,' *Psychol. Monog.*, 1915, 19; Webb, *op. cit.;* and Robinson, *op. cit.*

18-syllable original lists. These averages are 10.1, 10.5, 10.6, 9.9, and 11.4 respectively. The average number of repetitions required to memorize a 12-syllable list when not preceded by the learning of another list was 11.0. In four out of five cases, then, there is a slight positive transfer from the memorizing of the original list to the memorizing of the interpolated list. Furthermore, with the exception of one case, there is a continuous decrease in the amount of this transfer with increasing lengths of the first list, *i.e.*, with those conditions where inhibition was, in general, less marked. One should, however, hesitate to accept these facts as a confirmation of the transfer hypothesis. The amount of transfer demonstrated is in every case small and at least it does not seem sufficiently marked to furnish a cue for the complete explanation of the marked inhibition present under the conditions of this experiment.

VI. Summary of Results

1. The memory curves for all of the lengths of material studied show a general negative acceleration. These curves vary regularly according to the lengths of material represented. The shorter the material, the more rapidly do the earlier stages of its learning proceed and the earlier the stage of relatively slow progress is reached. In other words within the limits studied, the more unlearned material facing the subject, the fewer associations will be formed in the early portions of the total study time. Later, when the possibilities for new associations offered by the shorter material are beginning to give out, more associations per unit of time will be formed where the material is longer.

2. The amounts of any list learned after certain amounts of study may be considered as percentages of the total material in that list learned in certain percentages of the total time required to learn that list. Curves plotted on the basis of such a consideration of our data show a remarkably close coincidence. This finding confirms a similar one by Kjerstad.

3. The number of repetitions required for complete learning increases rapidly with early increases in the length of

material and more slowly with later ones. This is in accord with previous findings. Our results show, not only a continuous increase in repetitions with increasing length, but also a continuous increase in *repetitions* per syllable. This is not in agreement with previous results, which have indicated an early increase in *repetitions per syllable* followed by a decrease.

4. The procedure of considering the relationship between length of material and repetitions required to learn it is not very satisfactory for determining the relationship between length and difficulty, because the *repetition* is a unit of effort which, itself, varies with the length of the lists employed. We have shown the relationship between length and a constant unit of effort, *i.e.*, one presentation of one syllable, both for the data of previous experiments and for our own. The trend of these results indicates a general negative acceleration. That is, as material becomes longer, it becomes relatively more difficult. Meumann's results furnish an exception and, according to Henmon, this relationship is not continuous. Logically, however, it is what we might expect.

5. Our findings agree with most earlier ones in indicating that shorter materials are more rapidly forgotten than longer materials. Familiarity with the experimental situation for some undetermined reason seems to remove this effect.

6. In general, there is a decreasing susceptibility to retroactive inhibition with increasing length. That there may be a lower limit to this law is indicated by the fact that in a number of instances the shortest lists employed were least affected by retroactive inhibition.

7. Twelve-syllable lists, learned immediately after other lists, are learned slightly more quickly than those learned without a previous list. This positive transfer is most marked where the first lists are shorter. But the effect is not great enough to be considered seriously as a demonstration of the place of positive transfer in retroactive inhibition.

DISCUSSION

THE STIMULUS ERROR, A REPLY

BY SAMUEL W. FERNBERGER,

University of Pennsylvania

Professor Weiss[1] objects to a conclusion to an experimental study of mine to the effect that "this experiment seems to show the futility of properly interpreting the statistical results of a purely behavioristic study without the control of the introspective report." [2] The study in question was the determination of the difference limen to lifted weights under the three different instructions to judge: (1) the weights (stimulus error); (2) pressure sensations and (3) kinæsthetic sensations in the wrist. Statistically these different instructions gave relatively different values for the difference thresholds.

It is claimed that these experiments are not purely behavioristic because no attempt was made 'to establish laws according to which weight discriminations are made by some social or anthropological unit.' It is unfortunate, perhaps, but in an experiment of this sort one seems to be limited to the use of the graduate student and the recent Ph.D. as one's 'anthropological unit' and, frankly, I find them more interesting as a unit than '. . . post-office clerks, store-keepers, housewives, etc.,' and much more readily obtainable about a psychological laboratory. It has always seemed to me that one, in reading an experimental paper, considered the experimental setting,—Watson (to whom one usually turns in considering Behaviorism) insists upon it.[3] Our 'stimulus error' setting is as exact as we could devise from the method suggested and implied by Watson (pp. 37) when he speaks of "asking the subject as we do in the verbal report method: *Is this light brighter or dimmer than another light*, or, *Is this tone lower or higher than that tone?*" (italics Watson's) and in his further discussion of the stimulus and difference thresholds (p. 52).

[1] Weiss, A. P., 'The Stimulus Error,' J. of Exper. Psychol., 1922, 5, 223–226.

[2] Fernberger, S. W., 'An Experimental Study of the "Stimulus Error,"' J. of Exper. Psychol., 1921, 4, 63–76.

[3] Watson, J. B., 'Psychology,' 1919, p. 28.

Unfortunately we have found elsewhere that different individuals react differently under such instructions as well as the same individual at different times, so that they will give rather large differences in threshold values. It has been shown that these differences are not due to changes in the experimental setting or in the sense organs themselves but are due to differences in the attitude of the subject—if by that we are agreed to mean the manner in which the subject accepts and understands the instructions.[1] It would seem, therefore, that a study of the way a subject accepts the instructions and the differences in reaction due to the differences in attitude thus created would be a proper form of behavioristic research.

It may be the ideal that any subject, animal or human—'scientist, post-office clerk, store-keeper, housewife' or even a poor graduate student—would have aroused, always and invariably, the attitude which the experimenter meant to imply by the instructions. But, unfortunately, one cannot be sure that this is the case and so one has to use various experimental controls to be sure that he has obtained the desired result. Unfortunately, I am not aware of any experimental control of attitude or thinking or 'implicit sub-vocal response' other than by introspection, in the present state of our knowledge at least. I should be glad to learn of some other method less laborious and less tedious and more 'objective' than systematic experimental introspection for determining whether a subject is now judging pressure on the finger tips and now kinæsthetic sensations in the wrists *when you told him to judge the weights.*

Weiss insists that introspection 'as an end result is hopelessly inadequate to reveal the inheritance factors, the environmental factors and the training factors, which are the causes of a given set of reactions.' I am quite in accord with him in this statement. I certainly realize the limitations of the introspective method and its inadequacy in all but its own sort of problem. But objective methods and introspective methods may be used and should be used to complement one another wherever possible—if for nothing more than as an experimental control. And it turns out in an

[1] Fernberger, S. W., 'An Introspective Analysis of the Process of Comparing,' *Psychol. Mono.*, Whole No. 117, 1919.

Fernberger, S. W., 'The Effect of the Attitude of the Subject upon the Measure of Sensitivity,' *Amer. J. of Psychol.*, 1914, 25, 538–543.

George, S. S., 'Attitude in Relation to the Psychophysical Judgment,' *Amer. J. of Psychol.*, 1917, 28, 1–37.

experiment, such as the determination of the difference limen to lifted weights in the anthropological unit of the post-graduate student in psychology at Clark University in the year 1921, that such an introspective control of the way the subjects interpreted the instructions proved to be necessary. And, from other work, we believe that it is necessary when dealing with other anthropological units.

Weiss does not criticize the experimental results of this study but merely the implications which have been drawn from them. The experiment was devised to study the effects of attitude on the psychometric functions. It was found that attitude had an effect in graduate students. If such an effect is present in graduate students, trained in scientific method, it is reasonable to speculate further that attitude will be much more likely to have an affect as one changes from the housewife to the post-office clerk, because these non-laboratory types are less amenable to constancy of reaction.

Weiss also insists that introspections 'are abnormal supplementary speech reactions of the same nature as illusions, dreams, etc.' If by abnormal, he means unusual in that they occur in a few individuals who have been specially trained, I would agree. But it would also seem to me that any scientific observation whatsoever or any scientist must be similarly abnormal. But I hope that his inclusion of introspections with 'illusions, dreams, etc.' is a verbal slip and is not to be taken literally.