# THE

# Psychological Monographs

## Report of the Committee of the American Psychological Association on the Standardizing of Procedure in Experimental Tests.

Committee:

Charles Hubbard Judd
Walter B. Pillsbury
Carl E. Seashore
Robert S. Woodworth
James R. Angell, *Chairman*

Published by the Association.

# CONTENTS

# PREFACE

It has often been felt that the American Psychological Association ought actively to undertake, as well as merely to encourage, systematic contributions to research and the increase of scientific knowledge. In response to this sentiment the Association at its meeting held in New York City, December 27, 28 and 29, 1906, authorized the "creation of a permanent committee of the Association, to consist of five members, which shall act as a general control committee on the subject of measurements. It is recommended that this committee undertake two general lines of work, organizing as many subcommittees as it shall see fit, and calling to its assistance such outside help as it may desire: first, the determination of a series of group and individual tests, with reference to practical application, and second, the determination of standard experiments of a more technical character."[1]

The present report is the work of the committee appointed in execution of this resolution. The subjects assigned for investigation to the several members of the committee will be found representative of the two main lines of inquiry approved by the Association. They are as follows: C. H. Judd, tests on motor activities; W. B. Pillsbury, determination of intensity of sound; C. E. Seashore, discrimination of pitch; R. S. Woodworth, difference threshold in color tone, and free and controlled association; J. R. Angell, determination of mental imagery.

In accordance with the provisions of the resolution already partially quoted, the committee invited Professors R. M. Yerkes and J. B. Watson to report on tests for color vision in animals. Professor Judd invited the assistance of Professor

---

[1] A somewhat similar committee on physical and mental tests was appointed by the Association in 1895, and reported its recommendations at the annual meeting held at Boston and Cambridge in 1896.

Raymond Dodge in formulating tests on eye movement in connection with the study of nervous fatigue and mental disease. Professor Woodworth secured the coöperation of Dr. Lyman Wells in his work on association, and Professor Angell has been assisted in his work by Dr. M. R. Fernald.

The publication of the reports has been delayed much beyond original expectation because the work has proved unexpectedly arduous and difficult. Nevertheless, the committee has made explicit reports of progress at each of the annual meetings of the association since its appointment, and Professor Dodge's work after being thus reported has already appeared, being finally published in conjunction with Professor Diefendorf.[2]

If the funds appropriated to the purpose had permitted, the committee would probably have chosen to print all their material under one cover. This was not practicable, however, and the present method is accordingly accepted, despite its drawbacks. The reports thus far arranged for are to appear as follows: In the present volume—

Methods for Determination of Intensity of Sound, W. B. Pillsbury.

Measurement of Pitch Discrimination, C. E. Seashore.

Determination of Mental Imagery, J. R. Angell.

In the Animal Behavior Monographs, Yerkes' and Watson's report on tests of the color vision of animals.

In the monographs of the Psychological Review, Woodworth's and Wells' report on association tests.

Professor Judd's work on motor tests has been unavoidably interrupted and it is not possible to speak with certainty of its appearance.

None of these reports is to be considered final. Not only are there outstanding issues to be further studied, but after the reports have been subjected to criticism and practical tests by members of the association and other psychologists, the committee hopes, if continued in office, to print a more authoritative supplementary report in which the final judgment

---

[2] *Experimental Study of the Ocular Reactions of the Insane from Photographic Records,* Brain, 1908, Vol. 31, pp. 451–492.

of the committee as a whole will be registered in favor of one or another form of apparatus, method of procedure, etc. The committee will welcome all specific criticisms and suggestions.

# I

## METHODS FOR THE DETERMINATION OF THE INTENSITY OF SOUND

### W. B. Pillsbury

It seems to be the function of a report of this kind to consider the general principles that govern the problem, to compare and criticize the various methods that have been used and suggested, and to offer suggestions based on tests of the different methods available.

The object of any device to measure the sensitivity of the ear must be to produce a vibration of the tympanic membrane of measurable amplitude or intensity. The intensity of the sound wave at the ear will depend upon two general factors: the intensity of the sound at the sounding body and the laws that govern the distribution of the sound waves about that source. To know how intense the sound will be at the ear one must know how great the amplitude of vibration is at the sounding body, and the distance from the ear together with the way in which the sound spreads from the source, or one must have some means of measuring the intensity of the tone at the ear itself. A complete discussion involves: (1) A study of the different methods of inducing sounds and of measuring the intensity of the sounds that are produced. (2) A study of the laws that control the distribution of sound waves and their intensity at different distances. (3) Or one must devise some methods of measuring the intensity of the tone at the ear.

Perhaps the method of inducing a sound most used is by falling bodies, either freely falling or by the pendulum. Practically all of the more important psychological investigations have been carried on by that method. It is desirable therefore to determine as carefully as possible the controls that are necessary in this method and the means of measuring

the intensity of the tone. This is the more important, as
there has been some controversy in the literature over the
relation between the height of fall and intensity. Starke
asserts that the intensity is dependent upon the product of
the mass and the height, Vierordt that it varies as mass
times the height raised to the exponent $k$ and $k$ for him is
0.6. Tischer came to the opinion that $k$ varies with the height
and with the mass and that in consequence no law may be
asserted. Fechner also expressed the opinion that $k$ was
variable and must be determined for each different substance
of ball and plate.

Obviously the first problem is to determine the relation of
the intensity to the height of fall. I attempted this in three
ways, but the simplest and only one that was carried through
at all successfully was by measuring the amplitude of vibration
excited in a fork by the impact of a rubber ball falling from
different heights. The results may be seen in the curves
that are published herewith (fig. 1). In general it will be
seen that the amplitude increases much more rapidly than
the square root of the height. Each set was taken on dif-
ferent days. The amplitude was measured directly with a
reading microscope.

To show the relation, the square roots of the heights have
been plotted below the curve plotted from the amplitudes of
vibration measured directly with the microscope. Each point
is the average of from fifteen to thirty determinations. The
mean variations were from .2 to .5 scale divisions. They are
stated at the bottom of the curve for each point. It will be seen
that the curve of amplitude of vibration lies between the
curve of the square roots and a straight line. As the intensity
of the sound is proportional to the square of the amplitude
of vibration this means that the sound increases more rapidly
than the height but less rapidly than the square of the height.
It should be added of course that this result will not neces-
sarily hold for any other ball or any other receiving surface.
If the work in the determination of Weber's law in the
Leipzig laboratory had been made with this fork, the result
would be quite different from the formulae given in the

literature. What the actual formula for their apparatus is no one can know now.

In order more nearly to approximate the conditions of use of the fall phonometer, I attempted to measure with Wien's resonator one component of the noise made by dropping balls of wood and hard rubber upon different surfaces. The surfaces chosen were not particularly suitable to give perfect tones, but were some objects about the laboratory large enough to give the deep tone of the resonator, 120 VD. One was an ordinary wooden box, the other a galvanized iron basket. They were always placed upon a spring couch padded with cloth to prevent transmission of the tone through the floor or other solid objects. Of course there were other tones present than the one measured, and it is possible that the different components were favored by the different heights, but uncertainty in the quality of a tone is nearly as grave as uncertainty in intensity. The results are given in figs. 2 and 3. It will be seen that with conditions as nearly alike as they could be made from day to day there are enormous differences both in the absolute intensities of the tone and in the character of the curve. Some show practically a straight line or an increase in proportion to the square of the height, others are more nearly proportional to the height. These statements both rest upon the assumption that the intensity of the tone is proportional to the square of the rough measurements of the deflections seen in the manometer apparatus. Each value in the curve is the average of five readings taken in immediate succession. The irregularity of the different readings is probably responsible for the irregular fluctuations in certain parts of the curves. It may be argued that my materials are not to be compared with the ebony or ivory or steel plate that is ordinarily used in the fall phonometer. This I would readily grant, but would insist that everything that we know of vibrating plates anywhere indicates that the arrangement of nodes depends upon the point of the plate that is struck and the quality of the tone depends upon the arrangement of the nodes. Then too there is nothing in the character of the board in the bottom of our box that would make the

tone change from day to day, that is not also found in the plates
ordinarily used.   The curve of sound intensity has never
been worked out for an instrument of that sort and until it
is, one can know nothing of the sound values that are repre-
sented by the height of fall.

The explanation of the disparities is to be conjectured only,
but every source of error that we had to contend with is
represented in the actual experiment in which the ear re-
places the resonator.   First any plate vibrates in segments
and has different partial vibrations.   Occasionally variations
in a single result were undoubtedly due to the weakness of
the component that we were measuring.   The pitch of the
complex noise would be noticeably different when the read-
ing was found to be unusually low or high.   Anyone who has
worked with the ordinary ebony plate fall apparatus will
recall that frequently a tone will be very different in intensity
and quality from the usual ones.   The well known fact that
the same intensity induces different effects at different pitches
would probably suffice to account for the different appar-
ent intensities of these tones.   In addition to the different
conditions of reflection, the presence of different bodies
and the position of the resonator with reference to the walls
undoubtedly plays a very important part.   One of these came
to our notice in this experiment.   It had grown close and a
door was opened very slightly.   The opening was two meters
or more from the source of sound and about the same distance
away from the line between the source and the resonator.
Nevertheless the readings at once dropped very markedly
and continued lower until the door was closed.   All of these
irregularities indicate that while the relation of intensity to
height of fall is fairly close in the tuning fork, that for the
apparatus that is ordinarily used the variations are too large
and irregular to enable one to obtain results that shall be
constant, to say nothing of obtaining absolute measurements.
Since the pendulum implies all of the fundamental principles
of the freely falling balls, in addition to a source of error from
the tone of the suspending rod, it is unquestionably equally
unreliable.

Attempts have been made to determine the absolute limen by the use of falling particles. It is assumed that the energy of the falling body may all be transformed into sound and that the product of the height and the mass will indicate the energy that affects the ear provided suitable corrections are made for the distance from the ear. Theoretically this is open to two objections. Certainly not all of the energy is transformed into, sound. Much is lost in heating the ball and the plate. How much is left can only be a matter of conjecture. Estimates of the energy transformed into sound under other conditions vary from 1/1000 for the telephone to 1/15 that Wead estimated for this tuning forks. Webster's more recent estimates for a more economical transformer of energy than the ball and falling plate would place the loss nearer the higher figure. At the most the energy computations would give only the maximum values and whether the true limen were 1/10 or 1/1000 of that no one is at present in position to say. In addition to this disadvantage of not knowing what the results mean, the sound produced is open to all of the variations that the larger ball was seen to show in the experiment just described. Early in this investigation an attempt was made to calibrate a Lehmann acoumeter by determining the distance at which a small ball of cork or pith might be heard when it fell different heights. No regular results could be obtained. First the different reflections at different positions in the room made it impossible to draw any results from the distance. The sound might be fully as intense or even more intense at the more remote as at the nearer distance. The same height of fall often gave widely different results. This was undoubtedly due to the relation of the position of impact to the nodes of the plate. Combined, these factors made the method absolutely valueless. The theoretical and practical disadvantages combined make the falling body method worthless even for student exercises.

A second method is the tuning-fork method that Wead devised and which in a modified form is a test widely used by the acoustician. It is much more promising theoretically since it offers a more accurate measure of the absolute inten-

sity and is more generally applicable. Wead's[1] method was
to measure the amplitude of vibration of the fork and to
calculate the energy of the vibration from the dimensions
of the fork and Young's modulus for steel. He tested the damp-
ing of the vibrations and the weight necessary to bend the fork
and found that the formula held in practice with surprising
accuracy. To determine the potential energy in the König $Ut_2$
fork the following formula suffices. The potential energy in
each prong equals $23.94 \times 10a$ Ergs.[2] In the actual experi-
ments $a$ (the amplitude of vibration) is read directly by a
microscope and the energy is computed from the formula.
The experimenter takes the reading at the instant the sound
disappears, and the intensity for the ear is computed from the
energy at the fork and the distance of the observer.

Two difficulties may be pointed out in connection with the
method. First the absolute energy transformed into sound
waves is not known, and cannot be accurately separated from
the energy spent in heating the fork and in communicating
tremors to the supports. Wead had some results that indi-
cate that $1/15$ was transformed into sound but did not regard
the results as conclusive. The second source of error is even
greater. It is the difficulty in determining how the sound will
be distributed. Even in the open the ground will reflect or
absorb the energy differently, and in a closed room it is prac-
tically impossible to say what the law of distribution is.
There will be nodes due to reflection from the walls and other
interferences and reinforcements that cannot be determined.
In the open the wind will make a considerable difference in the
distribution. In spite of the disadvantages the results ob-
tained by the method agree fairly well with those obtained
by other methods as will be seen when the values that have
been determined are reported.
The general formula is

$$Y = \frac{b \, d_\pi^3 \, E}{8l^3} \, a^2$$

[1] Wead: *Am. Journal of Science*, Se. 3, Vol. 26, p. 177.

$v$ = energy of fork; $b$ = width, $d$ = thickness, $l$ = length of fork, and $E$ = Young's modulus.

For psychologists the method has the added advantage of using instruments that are everywhere available. Every laboratory has a tuning-fork and a microscope and that is practically all that is necessary. As compared with the fall phonometer or sound pendulum it gives a tone of known pitch and of constant pitch. The variation of sensitivity with pitch makes any attempt to use noise uncertain. Then too the constancy of pitch will obviate the uncontrollable pitch variations of the block that were seen to make the free falling ball method valueless. Other instruments have the further disadvantage that one cannot change the intensity at will except by changing the distance from the ear.

The ear specialists use another modification of the tuning-fork method that has some advantages. Instead of measuring the amplitude directly they measure the time that the fork may be heard by a defective ear and compare it with the time that the same tone is heard by a normal ear. It has been determined that the amplitude of vibration of the fork at two times will be proportional to $e^{-th}$ in which $e$ is the base of the natural logarithm, $t$ is the time during which the tone is heard and $h$ a constant that depends upon the rate of damping of the fork. This must be determined for each fork. If $t$ is the time during which the fork is heard by the ear to be investigated, and $t_n$ the time it will be heard by the normal ear, the limen $H$ is expressed by the equation $H_\rho = e^{-2h(t_n - th)}H_n$. It might be objected that one of these measurements will always be subjective. To this one can reply that the variation in the normal ear is probably less than the variation in the objective conditions under the most favorable circumstances. If the two ears be placed exactly in the same place the space error of the distant stimulus will be avoided which should fully compensate for any change in the sensitivity of the standard ear.[1] It is also possible to determine the absolute sensitivity of the normal ear.

---

[1] Schaefer: Nagel's *Handbuch d. Physiologie*, Vol. III, pp. 493–495.

One of the most convenient instruments for determining the limen of hearing or for making any tests that require absolute values is the telephone. The telephone may be easily changed in pitch or in intensity and very slight gradations of the tone may be obtained. The convenience of the telephone in use rests mainly upon the fact that the intensity of its tone varies directly with the current and that the current may be measured when the sound itself is very weak. In practice it is necessary merely to establish the relation between a given current and some measurable intensity of the tone and then to measure the strength of the current when the tone disappears. The strength of the tone at the limen may then be calculated directly. The direct relation between strength of current and intensity of tone has been established for faint tones by Lord Rayleigh and Wien.

To determine the energy that is given off from the plate of the telephone one must obviously know the size of the vibrating surface, and the amplitude of vibration of each part. Wien[1] expresses the energy in the formula

$$A = \frac{C\,p\,o}{2\,k}\,\Delta^2$$

$A$ is the energy that passes through a square centimeter of surface in a second, $c$ the rate of transmission of sound, $p$ the tone of the plate, $k$ the index of specific heat and $\Delta$ the relative pressure amplitude. $\Delta$ may be expressed in the equation

$$\Delta = 0.147\,\frac{k}{c^2} \cdot \frac{(2\,n\,\mathrm{N})^2\,R^{2^{a\,o}}}{\rho}$$

where $n$ is the time of excitation, $N$ the rate of the tone, $R$ the radius of the vibrating plate, $a$ the amplitude of vibration at the middle of the plate and $\rho$ the distance between the ear and the plate. In practice one may measure the amplitude of vibration of the plate directly by attaching a fine glass rod

---

[1] Max Wien: Ueber die Empfindlichkeit des Menschl. Oohres etc. Pflüger's Arch., vol. 97, p. 1.

to the center of the plate with sealing wax and measuring the amplitude with a microscope with micrometer eye-piece. Measuring the current offers some difficulties if one uses an alternating current. Wien made use of a dynamometer but I have found it more satisfactory to use either the Edelmann Saiten-galvanometer or to use Pierce's method of a rectifier and ordinary galvanometer. The Edelmann instrument takes advantage of the tendency of a wire carrying a current to move in a magnetic field. It consists of a fine wire mounted in the field of a permanent magnet. The movement of the wire may be read with a micrometer eye-piece or in certain of the instruments may be photographed. The advantage of the instrument is that the moving part has a very small mass and the oscillations induced by an alternating current are as easily read as the deviation from a constant current. A disadvantage emphasized by Pierce is that the own tone of the wire favors one rate of vibration over others.

The rectifiers of Pierce depend upon the fact that certain crystals have a much higher resistance in one direction than in another to small currents. I had such a rectifier constructed of molybdenite and found that an ordinary D'Arsonval galvanometer was as sensitive to the alternating current as the smaller Edelmann instrument. The rectifier is much cheaper and easier to make or obtain than the other instrument, and fully as convenient to use. In use, then, it is only necessary to determine the amplitude of vibration of the telephone for one intensity of current that gives an amplitude that may be conveniently read, and that may also be easily measured by the instrument at hand, and then determine the current that gives a liminal tone. The liminal amplitude may then be computed and from that, the energy that is given off by the telephone and the amount that is expended upon the membrane of the ear.

We have also to consider the difficulty in determining the part of the tone that actually falls upon the ear, which gives as much trouble here as elsewhere. Wien obviated this by closing the telephone by a metal cap and connecting a tube with the opening that might be inserted in the ear. It is possi-

ble to assume that all of the energy affects the ear under these circumstances. The error in this assumption is probably less than on any assumption that one can make of the distribution of sound where the source and the ear are at a distance from each other.

One other difficulty that the telephone presents is that the breaking of the circuit is always followed by a marked click that is more noticeable than the continuous tone. The intensity of the click depends upon two factors. First the regular vibrations do not ordinarily release the plate altogether. Residual magnetism holds the diaphragm nearer the pole of the magnet than its position of rest during the entire period of its vibration. This may be observed through the microscope. Then, too, the forced vibration due to the current probably does not cause so great a vibration amplitude as the natural vibration of the plate that appears when the current is cut off. More important than either, however, is the fact that the natural pitch of the plate (usually about 900 vd.) corresponds to a rate of vibration to which the ear is more sensitive than the lower pitches ordinarily used in the laboratory experiments. Whatever the cause, the disturbance due to the click is very marked. Dr. Shepard, working in my laboratory, found that the click could be lessened to the point of not being noticed if two tones were superimposed upon the telephone. He used the commercial current of 60 cycles and a 250 VD fork. The physical basis for the effect is obscure, but the empirical effect is obvious.

Taken all in all the telephone is probably the most satisfactory instrument to use for sound experiments in the psychological laboratory. It can be made to give a constant sound that can be varied at will in pitch, in intensity and duration. It can be applied to the measurements of intensity in all fields, is invaluable in experiments in rhythm when combined with an interrupting apparatus and has a number of uses in subsidiary experiments. In addition, the apparatus required is less expensive than many of the pieces sold as instruments for the measurement of audition and the parts can be used in many other ways. One needs the telephone

itself, a galvanometer, a rectifier if one uses an alternating current, a tuning fork, otherwise, resistances that may be adjusted and a microscope for measuring the vibration of the plate. An induction coil that may be used as a step-down transformer is also convenient in saving resistance. All of these pieces are needed for other experiments and the cost of those that are not likely to be on hand will be less than fifty dollars.[1]

In addition to these instruments that have been tested one might mention the instruments that have been used by Lord Rayleigh, by Webster, and by Toepler and Boltzman. Lord Rayleigh[2] used cans of ferrotype or tin made to vibrate by an electro-magnet. They were used to measure the relative sensitiveness of the ear to different tones and no attempt was made to determine the absolute intensity. The cans for different pitches were made similar in all of their parts. Intensity of the tones was determined by measuring the amplitude of vibration of the rims by a microscope. The magnets were within the cans and supported from the bottom. The alternating electric impulses were communicated by electric tuning forks. As the use of the instrument was restricted to measuring relative intensities extended description need not be given in this connection.

Very similar in principle is Webster's[3] instrument. In brief it consisted of a resonator made to speak by vibrations of a plate of ferrotype that in turn was made to vibrate by an electric tuning fork. The ferrotype plate closes one end of the resonator and is connected with the tuning fork by a metal rod. The amplitude of vibration of the diaphragm may

---

[1] Seashore's Audiometer may be mentioned as an instrument that makes use of the telephone. For Wien's method it would need some means of measuring the vibration of the telephone plate and would also be more convenient if the resistance could be changed by smaller gradations. It would also need a tuning fork or some method of interrupting a current to give a pure tone. In many respects it is more convenient to use an ordinary resistance box and telephone with the forks and other auxiliary apparatus that are at hand in the laboratory.

[2] Lord Rayleigh: On the Relation of the Sensitiveness of the Ear to Pitch, Investigated by a New Method. *Phil. Mag.*, 14, 1907, p. 596.

[3] A. G. Webster: On the Mechanical Efficiency of the Production of Sound. *Boltzmann Festscrichrift*, 1904, p. 866.

be read directly by a microscope. The instrument was used by Webster to measure the energy used and transmitted by sounding bodies, but would serve admirably as a reliable source of sound.

Toepler and Boltzmann measured the concentration of the sound in a closed tube by an interference method. It is more complicated than the methods described and need not be considered here.

In concluding this series of methods it may be interesting to compare the results so far obtained. The most striking result, as well as the one that shows most difference in the values, is the difference in sensitiveness to sounds of different pitches. The results of Toepler and Boltzmann,[1] Wead, Wien, Lord Rayleigh and Zwaardemaaker and Quix are given in parallel columns. The first column gives the rate of the tone in question and the column under the name

| N. | T. and B. | Wd. | Wn. | R. | Z. and Q. |
|---|---|---|---|---|---|
| 32 | | | | | 550 |
| 50 | | | $4 . 10^{-6}$ | | |
| 64 | | | | | 37 |
| 96 | | | | | $2,8 . 10^{-1}$ |
| 100 | | | $7 . 10^{-9}$ | | |
| 128 | | | | | $2,7 . 10^{-1}$ |
| 181 | $3 . 10^{-6}$ | | | | |
| 192 | | | | | $4,6 . 10^{-1}$ |
| 200 | | | $3 . 10^{-11}$ | | |
| 256 | | $83 . 10^{-6}$ | | $8.5 . 10^{-9}$ | $5,5 . 10^{-2}$ |
| 384 | | $28 . 10^{-7}$ | | $6 . 10^{-9}$ | $3,4 . 10^{-2}$ |
| 400 | | | $3 . 10^{-14}$ | | |
| 512 | | | | | $1,97 . 10^{-3}$ |
| 768 | | $31 . 10^{-7}$ | | | $2,5 . 10^{-4}$ |
| 800 | | | $7 . 10^{-15}$ | | |
| 1024 | | $11 . 10^{-7}$ | | | $2,7 . 10^{-4}$ |
| 1536 | | $22 . 10^{-6}$ | | | |
| 1600 | | | $1 . 10^{-15}$ | | |
| 2048 | | $71 . 10^{-7}$ | | | |
| 3200 | | | $5 . 10^{-16}$ | | |
| 6400 | | | $3 . 10^{-15}$ | | |
| 12800 | | | $5 . 10^{-14}$ | | |

[1] Toepler and Botzmann: *Pog. Ann.*, 141, 1870, p. 317.

the energy in ergs necessary to excite the ear.[1] It should be added that Zwaardemaaker and Quix used tuning forks in their determinations, and the values I have given are the results as corrected for certain errors by Wien in the second appendix to his article cited above.

The most striking feature of all of these experiments is the variation in the results. This variation, it will be noticed, is not only in the absolute values but also in the relative sensitivity to different pitches. The values do not seem to depend upon the method alone. It should be noted, too, that after Wien published his results that showed surprising differences in the limen for different pitches, Lord Rayleigh repeated his tests with the bells above mentioned and got results that showed that the limina for tones from 512 to 85 are related as follows, expressed with the limen for 512 as unity: 512, 1; 256, 1.6; 128, 3.2; 85, 6.4. He adds that he can see no reason for the disparity between his results and Wien's. Very obviously the results need to be gone over, but it seems to be no part of a report such as this to undertake the work. A new determination would have only the value that the methods and reputation of the worker give it and the older workers have apparently left nothing to be desired either in care or in the distinction of the men themselves.

Another method of varying the intensity of sound has been suggested and an instrument embodying it has been put upon the market. I refer to the audiometer of McCallie. I had hoped to have an opportunity of testing the instrument, but up to this time none has presented itself. In essentials it consists of some source of sound in a box and the intensity is varied by changing the size of the apperture that emits the sound. A priori the main objection to the plan is the fact that the intensity of sound will not vary directly with the area of the opening. Lord Rayleigh has shown that the sound will be very little affected by changing the width of the slit as compared with the length, where the slit is narrow. A change

---

[1] The values in the columns Wn. and Z. and Q. assume that below 3500 VD two vibrations suffice to excite a tone, that seven are needed at 6400 and twenty at 12800, while in the others the figures indicate work done on the ear in a second.

in length from .5 inch to .28 inch compensated a change in the
width of the slit from .004 to .020 inch. This proportion would
certainly not hold for wider slits but the law for the relation
between the area of opening and intensity of tone is not known
and would probably be fairly complicated. The instrument
also uses a noise rather than a pure tone, which in view of the
results given above renders the measurement uncertain.[1] Rub-
ber tubes are used to carry the tone to the ear as in the pho-
nograph. This has the advantage of removing any uncer-
tainty of the nature of the reflection from walls and the dis-
tribution of sound that constitutes an important source of
error in the experiments that make use of free distribution
through the air.

Of the sources of sound discussed here the telephone is the
most convenient, and probably the results are as accurate as
any. Next comes the damping method applied to the tuning
fork, the others in a third group. It should be remarked that
all of the methods make certain assumptions in the calcula-
tion of absolute values that may be in error and if in error,
would serve to account for some at least of the divergence
in results. Obviously a highly desirable advance in acous-
tics is the removal of these sources of error and the develop-
ment of methods more suited to direct measurements.

Aside from this group of methods that depend upon the
measurement of the sound emitted from the source, others
are or should some day become possible that depend upon the
measurement of the sound at the ear. The factors that make
for errors in the calculation of the sound that passes through
any given area at a distance from the source, even if the in-
tensity of the sound at the source is accurately known, are
many; some of them have been indicated above. The deflec-
tions and reflections induced by walls and obstructions are
obvious. To these one must add the interference between the
original and the reflected wave that tends to produce a dead
region at a little distance from the wall. To these uncertain-
ties of reflection one must add the fact demonstrated by Web-

---

[1] Professor Angell tells me that the sound has a tonal quality although it is described
in the catalogue as a noise.

ster that the nature of the surface has a marked effect on the character of the transmission. He found that a sound could be heard three times as far over water as over a lawn. He argues that the grass acts as a black body in absorbing sound. The probability that other surfaces may show similar differences and the complete absence of knowledge on the point may well serve to enforce caution in calculating the amount of energy that passes through any area at a distance from the sounding body. One must look with a measure of skepticism upon all methods that have made such calculations and these include practically all the results of the men whose work has been cited above. Wien's results alone are derived with the ear placed so as to receive the entire energy of vibrations, and even his receptacle may permit certain losses.

These considerations make desirable some means of measuring the sound wave that actually falls upon the ear. This might be accomplished were it possible to obtain an instrument delicate enough to measure the force exerted or the work done by the vibrations. Two instruments approach this degree of sensibility, the telephone of Pierce[1] and the receiving instrument of Webster. The Pierce instrument measures the electric current induced in the telephone circuit by the displacement of the plate of the telephone by the sound wave. The apparatus for measuring the current was described above. This with Pierce's adjustment was barely adequate, but it is no doubt possible with practice to get a more sensitive adjustment. Webster's instrument has not been described in print, and I have never seen it, but I am told that in essentials it consists of a glass plate that is made to vibrate by the sound wave and the slight movements of the plate are measured by an interference method. How delicate it may be I do not know. It is highly desirable that one or the other of these methods should be perfected to the point that will permit them to check the results obtained by other methods. Even if one could measure the energy where the ear is to be, when the ear replaces the measuring instrument the conditions would be changed in some

[1] G. W. Pierce: A Simple Method of Measuring the Intensity of Sound. Proc. Am. Acad., vol. 43, p. 377.

small degree, the head and body would reflect sound differently from the plate or telephone.  This error would undoubtedly be much less than those that are made in assuming a law for the distribution of sound in an open space, but still would hardly be negligible.  The method when developed will probably be more accurate than any that depends upon the energy developed by a source of sound, and if the promise is not altogether realized it will at least provide an invaluable check upon the other methods.

## RECOMMENDATIONS.

We may close with a series of specific suggestions based upon the tests and the examination of the literature.

1.   One should always use a tone as pure as possible as the stimulus.  A noise is nothing other than a tone of unknown pitch or a complex of tones of unknown pitch.  So much of the sensitiveness of the ear depends upon the pitch that if that be unknown or neglected the result is uncertain.  The preference of most psychologists for the use of noise seems entirely without rational support.

2.   All methods that depend upon the impact of falling bodies give uncertain results.  (a) The sound is always a noise.  (b) The pitch of the noise can not be kept constant. (c) No known relation exists between the energy developed and the amount transformed into noise.  While the relative intensities vary with the height of fall, the method gives no means of obtaining absolute measurements.

3.   The most accurate and convenient instrument for absolute measurements is a telephone actuated by a tuning fork or alternating current.  Some method of measuring the strength of current and a micrometer for measuring the amplitude of vibration of the plate is also required.

4.   Slightly less expensive and simpler is the equipment for using the tuning-fork method.  A fork with stable mounting, and a micrometer, or a fork whose rate of damping has been determined is all that is requird.  The results are probably less accurate than those obtained by the telephone, but that is not altogether assured.

# THE MEASUREMENT OF PITCH DISCRIMINATION: A PRELIMINARY REPORT.

By C. E. Seashore.

This report was called for by the Association in the belief that the systematic criticism and trial of current methods and means, and the statement of the essential implications, would economize effort for future workers and further the prospects of practical applications of the test. The measurement may be made by so many different kinds of apparatus of various degrees of worth, by so many methods of procedure more or less adequate, under so many hypotheses more or less specious, for so many purposes more or less legitimate, that the situation is very complicated.

This preliminary report is limited to a bare outline of the nature of the problem with tentative, positive recommendations in regard to the procedure.

The outline features here presented have served and should continue to serve as an aid in concentrating research on the most essential problems. Among the problems already taken up in the Iowa Laboratory are the following: the physical constants and variables in the tuning fork, the string, the reed, the pipe, the tone variator (bottle), and the siren; the adaptation of different types of instruments to different needs, and different psychophysic methods; the evaluation of one method in terms of another, *e.g.*, the method of the "average of the middle third" in terms of the method of right and wrong cases; the validity, *i.e.*, the degree of certainty or constancy in the measurement; the effect of intensity, duration, length of interval, order of trial, tone quality (timbre) and prolonged trial (fatigue), on the discrimination; the relation of consciousness of difference to actual difference (very surprising); the lower limit of tonality; the relative

discriminative sensibility for pitch within the tonal range; the relations of pitch discrimination to other musical capacities, *e.g.*, the sense of rhythm, ability in singing and playing, the perception of dissonance and the appreciation of music; also its relation to musical education and musical environment; the possibility of improving the capacity with practice; the establishment of norms; the relation of the cognitive to the physiological threshold; variation with age, sex, and general intelligence and special intelligence; and pedagogical applications. This report is based essentially upon the result obtained up to date in the investigations just named; general acknowledgment to these workers is here given. Volume VI of the *University of Iowa Studies in Psychology* will be devoted to the publication of some of these investigations.

These are large problems. A few have been solved in part, but most of them require a long time. This announcement of problems should in no sense be regarded as preëmpting the field. What we need is elimination of useless effort, coöperation, and stimulation of interest on the part of capable investigators. The committee therefore most cordially urges those who are engaged in, or can undertake, research in this direction to correspond with the chairman of this sub-committee in order that the work may be correlated. There will be no restrictions or interference, but it may enable us to avoid known snags and duplication of effort, and should ensure effective collaboration.

The sub-committee presents this preliminary report as an aid to the better formulation of work in the subject. The extended reports of research on these problems will be published whenever and wherever convenient, and it is to be hoped that, after a few years, the results may be summarized into a more comprehensive and authoritative report—a sequel and, in part, a result of this preliminary report.

There are many and varied demands for the measurement of pitch discrimination: *e.g.*, in the decision about whether or not to start a child in a technical musical education; in the classification of pupils for the purpose of class instruction; in anthropological measurements; in psychological measure-

ment as a means of determining the specific laws of pitch hearing and production, as well as in the working out of the general laws of mental life such as in attention, fatigue, association, suggestion, imagery, memory, automatism, and affective tone; and in the countless applications of these psychological and psychophysical facts to a theory of esthetics, with reference to the application and the expression of music as well as to the pedagogy of musical instruction.

## I.  General Account of Apparatus.

Pitch discrimination is usually measured by determining the least perceptible difference in the pitch of two successive tones.  Without actual measurement, the pitch discrimination is often estimated quite accurately, *e.g.*, by observing the accuracy in singing and in the tuning and playing of certain instruments, as well as in the ability to analyze musical clangs. The measurement may be made by means of tuning forks, stringed instruments, reeds, pipes, sirens, etc.  The relative merits of these instruments and the means for using them should be set forth in a later report.

The measurement of pitch discrimination has a short history.  While musicians have always wrestled with the problem in a general way, the significance of accurate measurement has not been comprehended until in recent years.  It may truly be said that the discovery of the real significance of the measurement of pitch discrimination came gradually with the discovery of the sources of error which enter into the test as worked out in the psychological laboratories.

Much time and energy have been wasted and false returns have been announced from experiments with unreliable apparatus.  Instruments for one kind of work have been used for another.  There has been no satisfactory coöperation in the designing of instruments and in the securing of critical tests before putting the apparatus on the market.  The makers have not had the right encouragement or discouragement.

## *A. Factors in Apparatus.*

The principal variables, the control of which determines the reliability of the apparatus, are: (1) form of the vibrating body—fork, reed, string, etc., (a) with reference to the selection from these fundamental types of vibrating bodies, and (*b*) with reference to the favorable construction of each within its own style; (2) the quality of material and workmanship; (3) accuracy of tuning; (4) mode of energizing; (5) mode of reinforcement and purifying; and (6) constancy of conditions. With these requirements in view, one must take into consideration, among other factors, the following.

1. *Reliability.* The measurement is made in terms of double, or complete, vibrations (vd.)[1] As a most serviceable standard of reliability in the apparatus and accuracy in tuning, we recommend $\pm.05$ vd. for increments from .5 to 2 vd.; $\pm.1$ vd. for increments from 3 to 5 vd.; and $\pm.2$ vd. for increments larger than 5 vd. This concerns the reliability of the original tuning, or the ability to set or tune, as well as the reliability of the apparatus, *i.e.*, freedom from progressive or constant errors.

We must distinguish between rough and accurate tests. One is as legitimate as the other in its place. We must also distinguish in the latter between tests of persons who have a high threshold and those who have a very fine, low threshold. This discussion of apparatus is concerned with accurate tests in which small sources of error would be disturbing.

2. *Availability.* In addition to being reliable, the apparatus must also prove available, *e.g.*, obtainable as regards cost, adapted to the necessary method which serves a specific purpose, and adequate to the present needs as regards time saving, accuracy, etc.

The various types of tuning forks, strings, reeds, bottles,

---

[1] Much confusion in terminology and in the interpretation of records has resulted from the fact that in some scientific work, notably the French, the single vibration (vs.) has been used as a unit, and both single and double are spoken of as vibrations. It is therefore desirable to use the differentiating abbreviation vd. in this country where the double vibration is the unit.

and sirens, each have some legitimate place. Thus the tone variator is our only available instrument for the measurement of gradual change in pitch; a well built monochord is very convenient for rough tests in a class; and a pair of electrically energized tuning forks with selective resonators is very satisfactory for the testing of an individual or group of individuals whose thresholds are known to correspond to the tuning of the forks, and when a sufficient number of trials can be made to justify the use of the method of right and wrong cases. But the first two of these are not reliable for accurate work and the last is not available for the tests most used.

We therefore recommend that, before the records are made, the adaptation of the instrument to the measurement in question be carefully considered.

3. *The Tonal Register.* In many cases it is necessary to make measurements at different levels of pitch. This may fix for us in part the choice of apparatus. Some instruments are adapted for very limited range and others for an extensive range of pitch. In statistical and other comparative tests, it is desirable that we should accumulate data with reference to some standard pitch that is decidedly favorable and can command general adoption.

In view of the available apparatus, the agreeableness of the tone, the fact that it represents the middle of the register of tones most used, that it is in relatively the most sensitive register, and that it is the international standard of pitch, we recommend that standard tests be taken at a′, 435 vd.

For the sake of securing uniformity, although the matter is quite arbitrary, we also recommend that the series of increments upon that standard be chosen above rather than below standard.

4. *Increments.* In a discrimination test of this sort, the change of pitch must of course be an absolute step, not a gradual change. The range of increase should be so chosen as to cover all cases in normal groups. .5 vd. for the smallest increment and 30 vd. for the largest increment at the standard of 435 vd. is adequate for most purposes.

Within this range the arbitrary steps chosen in extensive

experiment seems to require (1) as small difference as would have any significance for a group test in consideration of economy and trustworthiness, and (2) intervals of relatively equal psychophysical significance.

The increments should be chosen in some geometrical ratio, but the strict geometrical ratio would necessitate superfluous steps—too small increments at either end of the series. Furthermore, there is a real advantage in maintaining the increments in whole units without using fractions of a vibration. We therefore recommend as economical and serviceable, the following series of increments above the standard a′, 435 vd.: .5, 1, 2, 3, 5, 8, 12, 17, 23, and 30 vd. It will be observed that, except for the first two steps, this is an arithmetical progression of the second order.

It is probably just as valuable for us to know whether the threshold is .5 vd. or 1 vd. for the fine ear as it is to know whether it is 5 vd. or 8 vd. for a medium ear, or 23 vd. or 30 vd. for a very inferior ear. This equality is of course only an empirical approximation, for convenience and economy.

5. *Sounding.* In view of the fact that timbre, intensity, duration, direction of sources, etc., must be approximately constant; that pitch must be accurate and the tone constant during the time it is sounded; and that there is grave danger of identifying the tones; the mode of energizing the vibrating medium becomes a delicate matter. Many forms of apparatus are ruled out because they have some obstacle to effective energizing.

We therefore recommend that only such apparatus be used as will enable the experimenter to produce in rapid succession two tones which are practically alike in all respects except pitch.

6. *Timbre.* The richer the tones the more opportunity there is for discrimination to fasten more or less unconsciously upon some difference in the character of the tone as a means of identification. Thus, two tuning forks fixed upon wooden resonators invariably assume individual peculiarities by which they may be identified; two strings bowed in succession are open to the same objection. Even aside from this error

of identification, impure tones are probably more readily distinguished than pure, unless the impurity is of such nature as to be distracting, as in a badly bowed string. This relatively greater ease in discrimination for rich or impure tones is probably due to aid from other characters than pitch, and therefore illegitimate. There are two ways of dealing with it: one is to use a selective resonator, *e.g.*, A Koenig or a Helmholtz resonator, or a water resonator; the other is to transmit the vibration through a common medium, *e.g.*, by telephone. The telephone acts remarkably well as a selecting instrument, when used as in ordinary telephoning, and the timbre of the tone produced in the single receiver becomes quite uniform, when the sound comes from approximately uniform sources, but the tone is not as pure as when heard from the resonator directly.

In view of these facts we recommend that the test be made with the purest tone available. As far as we know now, this is best produced by an unmounted tuning fork reënforced and purified by means of a selective resonator.

7. *Resonance.* The sound is most favorable for pitch discrimination when it is just loud enough to be clearly heard without effort. Loudness depends chiefly upon the resonance of the mounting and the surroundings of the vibrating body. It is largely in the reënforcement that the tone gets its individual character; and the increased resonance usually results in increased richness of tone, which is a source of most serious disturbance in this test.

If a single individual is to be tested, a tuning fork, for example, may be held close to the ear with good result without a resonator. For tones lower than 75 vd. the fork thus presented should have hard rubber disks, about 8 cm. in diameter, mounted on the outside of each prong. Such disks strengthen the tone satisfactorily. But, for the ordinary work, with tones near the middle of the register we recommend, as in 6 above, the use of the selective resonators with good tuning forks unmounted.

8. *Danger of Identification.* One of the most insinuating and persistent obstacles—one most frequently overlooked—

is the possibility of identifying one or both of a pair of tones by other means than pitch difference pure and simple. The commonest means of identification are timbre (purity) loudness, characteristic difference in facility of handling, and location of the sounding body. The situation is doubly complicated by the fact that if the observer merely imagines, or has an illusion to the effect, that he can identify a tone by one of these accessories, this is just as fatal to the test as if he actually did identify. And it is further aggravated by the fact that the more effort we make to secure the desired uniformity, *e.g.*, absolute equality in intensity of tone, the more of a temptation we set the observer for pouncing upon this as a means of identification. And, worst of all, the identification may work itself out subconsciously, and indeed it ordinarily does, without the observer being aware that the accessory factors play any rôle, in the estimation of pitch.

For these reasons we are frequently forced to discard apparatus which is reliable in itself and may have other advantages. Our general solution is to get as much uniformity as possible without making it constant, *e.g.*, the intensity of a tone is kept approximately uniform by placing the successive forks at approximately the same point before the resonator. But there is a constant slight variation which cannot be predicted or verified. The same applies to timbre, duration, mode of presentation, etc. The observer should understand that although these factors are kept fairly uniform, they are not constant and that he can find no means of identification in them because they vary fortuitously within small limits. The apparatus must therefore be constructed and manipulated so that this condition can be carried out, *i.e.*, the accessory variation shall be the very minimum, but the observer shall have the positive instruction that the variations are not peculiar to any one tone.[1]

[1] An illustration of this difficulty is found, e.g., in using an automatic hammer to strike the string or fork. The hammer produces more uniform intensity than we can produce by a stroke of the hammer in the hand; but after a few trials the observer will detect some constant peculiarity of that stroke, and from that moment the test is vitiated. And even if the observer does not detect any difference the chances are that he will be influenced subconsciously by automatism in the direction of identifying tones.

We therefore urge that no test be undertaken, or record accepted, unless such precaution has been taken with reference to apparatus, method, and efficiency in manipulation as to guarantee that this danger of identifying one of two tones falsely has been ruled out.

## B. *Apparatus Rejected.*

As stated above, a monograph on the availability and the reliability of the instruments which have been used in this test is in preparation and will be published with measurements in detail. Each instrument is tested in several respects by means of the tonoscope, which makes ready and accurate measurement of all instruments by the same means to the accuracy of $\pm.02$ vd.[1] The measurements have been carried far enough to lead us to reject, for our present purpose, the reed, the piston pipe, the bottle (tone variator), the siren, and the string. All of these are available for rough work and each has some superior feature which may justify it for some special purpose. We must, however, await the publication of the details of the examination before assigning each to its place. In the meantime any one who is engaged, or proposes to engage in research with any of these instruments, may obtain available information by corresponding with the writer.

For our present purpose, all these instruments are rejected (1) because they do not give a sufficiently pure tone, and (2) because, under the conditions ordinarily employed, they are not sufficiently accurate.

We urge that no serious research be undertaken in pitch discrimination unless the experimenter either finds authoritative standardization of his instrument on record or himself undertakes careful tests on the instrument before using it. This is a commonplace rule in psychology but it is as frequently violated as it is commonplace.

[1] This is in rapid reading. For finer adjustment than $\pm.02$ vd. it is more convenient to use the beat method.

## *C. Apparatus Recommended.*

In view of the above and other considerations we recommend as the best available and most serviceable apparatus, a set of twelve unmounted tuning forks, a Koenig adjustable resonator, and a simple sounding rod.

The forks best suited for the purpose are of the grade marked 22807 in the catalogue of Max Kohl, Chemnitz, i, S., Germany, and will be known as the "standard pitch discrimination set." The set includes two standard and ten increment forks. They are tuned to the accuracy recommended in Sect. 1 above, for the increments recommended in Sect. 4 above at the standard recommended in Sect. 3 above. They may be imported duty free for about $41.00. The same set of forks untuned (all tuned to 435 vd.) may be imported duty free for about $34.00.[1]

A cheaper and smaller, yet serviceable, grade of forks is advertised as No. 1730 in the catalogue of the C. H. Stoelting Co., Chicago, and may be obtained tuned approximately the same as the above set for about $18.00. The same forks may be obtained untuned for about $12.00.

It is essential that the forks shall be fairly heavy, and of uniform size and shape (except for tuning) and of good material and workmanship. There are many styles and grades of forks on the market which would answer the purpose.

The resonator is a stock piece of physical apparatus, marked in the catalogues "Koenig adjustable resonator, mi³–la⁷" and may be imported duty free for about $6.50.[2]

For a sounder, use 1 inch square rod covered by very heavy soft rubber tube and fastened in a horizontal position firmly to some heavy metal support, so that the forks can strike on one edge of the rod.

---

[1] These forks are 14.5 cm. long, with prongs 9.5 mm. wide and 5 mm. thick, and perfectly uniform in shape.

[2] In lieu of this a serviceable resonator may be improvised by taking a graduate or a plain glass tube about ten inches long and 1¼ inches in diameter and pouring water into it until it is tuned to the right pitch. Or, a brass tube may be cut to the proper length and one end corked. Such devices give a good quality of tone but it is difficult to get the tone loud enough for a large room.

An opaque screen should be provided to prevent the observer from seeing any movements of the experimenter.

### D.  *Directions for Tuning.*

Since the experimenter must at all times be ready to test and, if necessary, correct the tuning of the forks; and since the difference in price is considerable, it is generally a good plan to buy the untuned forks.

Raise the pitch of a fork for each of the ten increments by filing equally, symmetrically and squarely on the ends of the two prongs—not on the side of the prong as is usually done. Time by counting beats between the filed fork and the standard. For final tests count the beats for as long a period as they can be heard. As one cannot count more than three beats per second with certainty the 5vd. fork may be tuned from the 3vd., the 8 vd. from the 5vd., etc., Avoid the sources of error mentioned above and following, *e.g.*, temperature, force of the blow, the place of striking, etc. In order to reduce the effect of sympathetic vibrations from one fork upon the others hold the forks before the resonator as far apart as may be compatible with audibility of the beats. Test to the accuracy recommended in Section 1 above.

### II.  Precautions to be Observed in the Tests.

### A.  *With Reference to Apparatus.*

1.  *Temperature.*  Use each of the two standards alternately for short periods, and hold all forks with fingers near the end of the stem of the fork.

The pitch of a fork falls by .00011 vd. for each degree centigrade of rise in temperature, according to Lord Rayleigh. Larger differences than this are quoted by experimenters, but it is probable that the effect of the temperature varies with the pitch, size, form, etc., of the fork; perhaps also differently with means and extremes of temperature. A 435 vd. fork just taken from boiling water is more than 4 vd. lower in pitch than the same fork just taken from ice. This error is,

however, practically eliminated: (1) by keeping all the forks
in the same atmosphere, and (2) by using the two standards
alternately to reduce the cumulative effect of heat from strik-
ing and from contact with the hands, and (3) by holding all
forks near the end.

2.  *Position of the Blow.*  Hold the fork so that it strikes the
sounder at the middle of the prong and squarely on the side.

A blow near the middle of the prong starts the prong to
vibrate most favorably—faster and more regularly than when
struck either at the tip or at the shoulder.  An extensive
series of tests with one fork gives the result: middle, 434.38
vd.; tip 434.22 vd.  The blow at the middle of the prong gives
the best quality of tone.

Ideally, both prongs should be snapped so as to start
synchronously in true phase; but this is impracticable, there-
fore we adopt the plan of striking only one prong, but under
uniform conditions.  The vibration frequency of the fork is
irregular until the two prongs vibrate symmetrically.

3.  *Force of the blow.*  Strike the sounder uniformly with
as light a blow as will produce a distinctly audible tone through
the resonator.

A light stroke produces the smoothest tone, a tone that
is most favorable for true pitch discrimination, and a tone
that remains uniform for relatively the longest period; a
strong blow produces a harsh and changing tone and injures the
forks.  The blow should be a light tap with free rebound, in
order that the initial impetus given the prong may not be
modified by continuous pressure or repeated contact.

4.  *The Sounder.*  Make sure that the sounder is soft,
well rounded and firm, with a  minimum of resonance.

A springy sounder prevents the proper rebound of the fork.
Any harshness or other unpleasantness in the thud is likely
to serve as a distraction, coming as it does just at the moment
attention is to be focussed.

5.  *Adaptation Period.*  Allow at least one second after
striking before presenting the fork before the resonator.

Even under favorable circumstances, the first sound of
the fork is rough, and the vibration frequency varies until

the two prongs vibrate symmetrically. This makes it inadvisable to sound a fork resting on a resonator, by striking or blowing.

6. *The Resonator.* Suspend the resonator so that it can vibrate freely; present the fork quickly and uniformly by bringing it straight toward the mouth (not from the side) of the resonator, so that the upper end of the side of one prong covers the opening; withdraw it in the same line.

If the resonator were laid on a solid, *e.g.*, a table, that body would add a disturbing resonance. Presenting the fork from the side gives a rough edge to the resonator tone and may give a clue to identification of forks. A quick and firm movement of the hand insures a clear cut beginning and end of the tone. Ordinarily the fork should be held as close to the resonator as it can be without any danger of touching; yet it is possible to hold it so close that a recognizable change in quality of the tone is produced.

The wooden resonator is discarded because it has a very rich and variable accessory resonance.

7. *Testing the Forks.* Test the forks frequently and keep record of the changes made in a given set of forks. Small forks like these may change vibration frequency by handling. Hard and irregular striking is probably the greatest source of the derangement of forks. A tuning fork should be handled like a tuned string,—with care. This is especially true of the lower forks which represent the small increments. Wherever accurate measurements are made, the lower forks should be tested frequently. In testing, all the variables must be eliminated. No fork should be filed until it is perfectly clear that it is faulty, and then only after its relations to the rest of the series has been determined.

### B. With Reference to Procedure.

1. *Intensity.* Make the two tones equally loud—just so loud that they can be heard distinctly without effort.

One of the most serious obstacles we encounter in this test is the effect of intensity or loudness upon pitch. An immature

observer is likely to identify high and strong to some extent.
A more developed or trained observer may reverse the associ-
ation and show a tendency to judge the weak tone high. For
this reason, various contrivances have been designed to secure
uniformity in the intensity of tones. We here recommend that
the experimenter shall practice, and simply trust his own ear
and hand for approximate uniformity. If a fork is sounded,
either too weak or too strong, the trial should be repeated. This
method is adopted because it prevents the observer from
attaching importance to heard intensity differences. When
mechanical contrivances are used to produce uniformity in
intensity, the observer either hears or fancies he hears a
difference which is consciously or unconsciously taken as a
cue for the identification of the fork.

Extensive experiments show (1) that both trained and un-
trained observers may be influenced by intensity in their
pitch judgment; (2) that although there is a tendency among
the untrained, especially the ignorant, to judge the loud tone
the higher, it may work either way; (3) that the same individual
may show one tendency at one time and the reverse at another;
(4) that for trained observers the two tendencies are about
equal; and (5) that the tendency is more serious for large
than for small intensity differences. Introspection shows
that this confusion rests largely on motor tendencies or
motor images. We associate high and strong with strain—
the reversal can in some cases be traced to a correction, con-
scious or unconscious, based on knowledge of this danger.

Experiments show that the just perfectly clearly perceptible
tone is most favorable for accurate results. It is ordinarily
purer than a stronger tone and favors concentration. Experi-
menters must guard against a very common tendency, usually
unconscious, to facilitate the discrimination by making the
tones loud; and untrained observers usually desire (unwisely)
a loud tone.

2. *Duration.* Sound each tone about 1 second. The
most favorable duration of tone is about $\frac{1}{2}$ second. This
cannot be obtained in a simple method of manipulation; and
the variation with duration from $\frac{1}{4}$ sec. to 1 sec. is not large.

The ratios found by experiment are: $\frac{1}{4}$ sec., 73 per cent; $\frac{3}{8}$ sec., 82 per cent; $\frac{1}{2}$ sec., 84 per cent; and 1 sec., 82 per cent. We may therefore content ourselves with as short a tone as can be produced consistently and economically with clearness and uniformity. In a practiced experimenter it approaches one second, and this should be kept constant. Here again both the experimenter and the observer have a tendency to crave a longer duration.

For most accurate work, in which duration is a factor to be varied, or is otherwise important, mount a revolving disk in front of the mouth of the resonator. Control the speed of the disk and cut slits in appropriate sectors to regulate the duration of tone and the time interval between the two tones. In this case it is necessary to have one resonator for each fork. The fork may be held by the hand in the usual way. This also has the advantage of forcing the trials in rapid succession and that favors the discrimination.

3. *Time Interval.* Hold one fork in each hand and strike them in rapid succession; then present them to the resonators so as to make the time interval between the two tones about one second.

Discrimination for successive stimuli always involves the memory element. The curve of tonal memory, as is well known, is a parabolic curve showing that the accuracy of memory falls off very rapidly, immediately after the first second of interval. But for intervals from $\frac{1}{16}$ sec. to 1 sec. there is practically no difference, as is shown by the following ratios: $\frac{1}{16}$ sec., 80 per cent; $\frac{1}{8}$ sec., 75 per cent; $\frac{1}{4}$ sec., 81 per cent; $\frac{1}{2}$ sec., 78 per cent; and 1 sec., 75 per cent. It is therefore important that the time interval be kept fairly constant and that it should not exceed 1 sec. The most favorable interval may depend in part on temperament and training of the observer.

4. *Order.* Determine the order of presentation of the higher and lower forks by chance; use keys prepared beforehand by tossing a coin. Modify the "chance order" of the key so that there shall be no more than three trials of the same order in succession.

If two forks of the same pitch are sounded in rapid succession  some have a tendency to call the second higher, others lower.   This error is eliminated by alternating the order of the standard and the compared forks in this chance order.

It is important that this order should be determined objectively.   On account of the community of ideas and a natural tendency to anticipate, arbitrary choice of order by the experimenter would lead to guessing "what he will do," and expectation, which would influence the judgments very seriously.

5.   *Uniformity.*  Avoid identifiable uniformity.  Many devices are available for securing uniformity automatically. Thus, a pair of electrically driven forks, with necessary contact. may be used, as the resonator does not speak to the accessory sound.  Or, if the accessory sounds are considered disturbing, the forks at the resonator may be driven "tandem" by another pair of forks in a distant room.   The electrically mounted forks may be held, unmounted, in the hand.   Automatic hammers may be so mounted on the forks as always to strike the fork in the same place and manner and with the same degree of force.   The forks may be so mounted that they swing automatically into position before the resonator.   A resonator may be mounted permanently before each of the automatically struck or electrically driven forks.   The duration and the interval may be regulated by a rotating disk or pendulum.

These and many other devices have been tried with the general result that all except the last (the regulation of duration and time interval) have been rejected.   It required more skill to operate one of these devices *safely* than to follow the simple plan recommended.   The ever present and practically insurmountable danger is that of identification of a tone by some permanent character other than pitch.

6.   *Fatigue.*  Adapt the length of the test to the endurance of those tested.

A test of this sort is novel and requires continuous conscious effort and is, therefore, surprisingly "fatiguing."   The ear itself does not fatigue in the same sense as the eye fatigues,

but the test requires an unusual concentration of attention, and therefore causes a general exhaustion.

Extensive experiments on this specific problem show: (1) that, even through the most severe continuous and uninterrupted strain for two hours, a normal, reliable, adult observer (a) feels no fatigue or strain in the ear, and (b) shows no progressive loss of keeness of discrimination; (2) that there are marked periodic fluctuations in capacity which usually correspond to felt distraction, ennui, or lack of effort; (3) that what is generally known as fatigue in this experiment is primarily an unjustifiable feeling of restlessness or blaséness coming from the habit of change and, secondarily, discomfort from position; (4) that, in normal observers, the discrimination tends toward automatism and becomes easier so far as effort is concerned during uninterrupted progressive adaptation; (5) that in some persons the strain brings on a headache; (6) that the test is far more straining on the person who is not familiar with it than one who is; (7) that after extreme exertion there sometimes follows a reaction in the form of a feeling of exhaustion, especially if the observer has worked in a continuous "heat" unconscious of his bodily self; and (8) that in such an endurance test good results are obtainable with ease in proportion to the absence of distraction. In short the strain of the test is "wearing" on the observer and, although the progressive adaptation is favorable to the ease and accuracy, it is difficult to secure continued application for long periods.

Therefore with primary children the test should not be carried beyond ten or fifteen minutes of continuous testing at a time. With grade and high school children, a half-hour test is most favorable, and even with adults two half-hour tests are more favorable than a one hour test.

7. *Attention.* Secure the most favorable form of attention, the secondary passive; and favor economic distribution of attention by proceeding rapidly and regularly.

It is all important that the observer should trust the primary impressions, and that systematic warning should be given in such a way as to sustain the most favorable form of an

attention wave. The variation of a fraction of a second from the time at which a stimulus is legitimately expected is likely to modify the record. In a test of this sort it is always assumed that the judgment is made under the conditions of maximum efficiency and attention.

8. *The Charge.* Impress the individuals or groups to be tested with the seriousness of the test and their accountability for every judgment.

There is always a certain amount of lethargy in a group of individuals, and the group feeling tends to lessen individual responsibility. It is, therefore, necessary that a very specific charge shall be made to the group at the beginning of the test. Call attention to individual responsibility, the personal nature of the record, the significance of attention, the demand of absolute integrity etc.—and do it with a vim. Be specific and firm. A charge of that sort may lower the average record for the group very materially. We must always assume that an effective charge has been made.

9. *The Judgment.* Always require the form of judgment which is characteristic of the method of right and wrong cases, *i.e.*, the observer is limited to two judgments (higher or lower); or, if the third judgment (equal) is allowed, the records of that equal judgment must be distributed in accordance with the rules of the method of right and wrong cases.

This rule is made mandatory by the fact that (1) the illusion of difference is very insistent and (2) that there is no close correlation between different degrees of certainty and actual difference near the threshold. We know rightly with a fair degree of certainty whether or not we hear a sound, or whether it is strong or weak, but not so with pitch of sound. We have distinct convictions that we hear difference in pitch often when there is no difference and when the difference is below the threshold.

### III. CLASSIFICATION AND ORDER OF TESTS.

All tests may be divided into (1) preliminary and (2) final. Each of these may be made as (1) group or (2) individual

tests. The preliminary group test is heterogeneous, the final homogenous. The heterogeneous group test is one in which a group of individuals, *e.g.*, a grade, a class, a type, or a community, is tested *en bloc* for the purpose of classifying the individuals in the group as a preliminary to the final test, or for rough and rapid permanent classification. The homogenous group is one in which the individuals have been grouped together on the basis of approximate equality in pitch discrimination, as determined by some preliminary test. For the final test we may employ a well recognized method; for the latter it is necessary to use an improvised method. Both are described in the following section.

The extent and order of the test should be determined in view of the purpose to be served, time available, means at hand, etc. For thorough examination of large numbers we recommend the following order of procedure:

1. A one-hour test of the heterogeneous group[1].

2. A one-hour "final" test of homogeneous divisions of the group, based upon the classification obtained in 1.

3. Homogeneous group tests with full introspection, *ad libitum*.

4. Individual interview, check trials, and extended "final" individual test.

## IV. Directions for Tests

A. *Preliminary.* When the first, or only test is to be made on a heterogeneous group, set aside one hour of time for the actual test. Supply each individual with a cross ruled paper, designating the vertical columns by numbers from 1 to 20, and the horizontal columns by the letters, A to J. Call the increments A, B, C, D, etc., beginning with the largest. Thus, A denotes a difference of 30 vd.; B, 23 vd., C, 17 vd.,

---

[1] The limit to the numbers of persons that may be tested at one time depends upon a number of considerations such as discipline under control, the loudness of the tone, the homogeneity of interest in the group, the acoustic qualities of the room, the skill of the experimenter, etc. Under favorable circumstances, one can readily handle a group of 200 to 300 persons at one time for a preliminary test.

etc. Illustrate to the group how the forks are sounded and what is meant by pitch.

Then give instructions to the following effect: I will sound two tones in quick succession. You are required to record whether the second is higher or lower than the first. There will always be a difference in the pitch, but if you can not hear it, you must guess. If the second tone is higher, record H; it if is lower, record L.

Give a few preliminary trials in which all are required to speak L or H instead of writing it. Follow this by some illustration of differences in intensity until all distinguish clearly between intensity and pitch. Urge them to trust the first impression, and warn them not to despair when they feel in doubt. Suggest that they keep their eyes closed when listening. Explain that before each trial you will give them warning by designating the square in which they are to record. Thus: column 1, A; B; C; etc., and in the same manner in succeeding columns.

In determining the order of the sounds, follow a key made out beforehand, as directed above, to determine whether the standard shall be sounded first or second. This operation of the law of chance should be explained to prevent guessing or probabilities.[1]

All being prepared, give without interruption one set of trials, A to J, and then allow a one-half minute rest. Take as many sets of trials as the time permits. A skillful experimenter can take about eighteen sets in an hour. If the record is not to be used in any important way, read off the key slowly, after completing five sets, and require each individual to check errors by drawing a bold line through the letter which is wrong. The experimenter must adapt methods to conditions. For example, each observer may be allowed to check his own record or they may be required to exchange. The experimenter must also use his judgment in deciding whether or not it is necessary to counter-check the records.

---

[1] It should be explained in particular that if the coin has fallen head up, *e.g.*, two times in succession the probability is still exactly equal for head and tail.

Request each observer to make note of anything that may aid in the interpretation of the record and its significance. For certain purposes it is desirable to have systematic information about each individual, as regards musical training, attainments, environment, appreciation, etc. The following questionnaire has been found serviceable.

Name ................Age....Date..........Hour....Class.....

Please give as specific and detailed information as possible in regard to your:
  I.  MUSICAL TRAINING
        1.  In public schools
        2.  Private vocal lessons (when, where, how long, etc.)
        3.  Private instrumental lessons (when, where, how long, etc.)
 II.  MUSICAL ENVIRONMENT
        1.  Instruments in your home, and their use
        2.  Musical encouragement at home (trained voices in family)
        3.  Opportunities for hearing music of any sort (specific)
III.  MUSICAL EXPRESSION
        1.  Favorite selections you can sing (by ear? by note?)
        2.  Favorite selections you can play (by ear? by note? instruments?)
        3.  Singing or playing in public (parts, occasions, etc.)
 IV.  ENJOYMENT OF MUSIC (WHAT ACTUALLY APPEALS TO YOU)
        1.  Vocal (solo, quartette, chorus, opera, popular songs, classics, religious, secular)
        2.  Instrumental (solo, symphony, band, etc.)
        3.  Characteristic effects of music (mental, physical)

We begin with a large step because that gives the observer a clear grasp of the nature of the difference and starts him with confidence in his judgment. We take a set at a time (vertical column) instead of giving successive trials on the one step (horizontal column), because the short set makes a favorable distribution of strained attention. We take the whole set of ten steps each time primarily because we are dealing with a heterogeneous group of observers, and secondarily in order to approach the threshold under uniform conditions. Twenty sets really amount to twenty trials, each

set being a trial to see how far down on the scale of incre-
ments the observer can hear correctly. Given the range of
increments, and the allotted time, the task is to secure as
many trials as may be needed for a fairly reliable record
within this minimum period of time.

The computation of the records must be left to the ex-
perimenter. Run through the checked records and place a
tally mark for each set in the horizontal column above the
one *in which the first error occurred* in a set. The completed
tally will then show the distribution of the thresholds for
sets in terms of the smallest increment for and above which
all judgments are correct.

Compute the *medians*. The formula for the median may
be found in Titchener's *Instructors' Manual*, Quantitative,
pp. 8, 9 and 361.

Supplement this numerical record by notes on internal
evidence in the record, such as modes, progressive changes,
confusion, mistakes in record, etc., and, when fully justified,
enter a numerical estimate of a correction. The record thus
corrected should be more reliable than the original.

For the series of threshold values thus obtained (before
the correction, if any is made) the standard deviation, or the
probable error, or both, should be calculated. For direc-
tions see same reference to Titchener, pp. 9–11. But these
calculations are very laborious. For ordinary work we
therefore recommend as a measure of variation in the record
the use of the mean variation (m.v.) computed as follows:
Regard the difference between successive steps as equal
psycho-physic steps and, with the increment which is nearest
to the median as a base, multiply the number of cases which
are one step from the base by 1, the number that are two
steps away by 2, the number that are three steps away by 3,
etc.; divide the sum of these products by the total number
of cases (sets).

This record on a heterogeneous group should be regarded
(1) merely as a preliminary to more accurate tests in homo-
geneous divisions of the group, or (2) as a rough record,
serviceable when exactness is not required. It is a "quick
method."

*B. Final.* Take in one division all observers who have approximately the same threshold, as determined by preliminary test; use one, two, three, or even four[1] increments, according to the closeness of the grouping, and proceed by the method of right and wrong cases to find the threshold for 75 per cent correct cases.

*C. Individual Tests.* In making individual tests the preliminary procedure here recommended for the group may be followed; but in most cases it is better to make a quick informal skirmish to find the approximate threshold and then proceed by the method of right and wrong cases, using a single increment.

## V. Interpretation.

### A. Distribution of the Records for a Group.

1. *The Normal Curve.* The normal curve of distribution of the records for a preliminary group as determined by the preliminary method described above is shown in Fig. 1. (From data collected by George H. Mount and Franklin O. Smith.) The curve is based on the records of 781 undergraduate university students, 296 men and 485 women.

In the interpretation of this curve it should be remembered that: (1) most of the records (581) are in terms of the "average of the middle third,"[2] in place of the median; (2) no correction of this has been made on the ground of internal

---

[1] For a first homogeneous test on large numbers it is convenient to take three groups as follows: Group A, including all those whose thresholds lie above 9 vd.; Group B, including all whose threshold falls between 9 vd. and 3 vd., inclusive; and Group C, including all whose threshold lies below 3 vd. Later, for extended tests, these groups; may be subdivided.

As a rule, one hundred trials in each instrument is the very minimum for the number of trials that may be trusted. Increments for which the record shows less than 65 per cent or more than 90 per cent right cases should not be used.

[2] The average of the middle third was obtained by rejecting the lowest third and the highest third of the observations and counting the "average of the middle third." For 183 cases in which the median and the average of the middle third were both computed, the former gave 7.4 and the latter 7.2 as a central value. The results are therefore quite comparable.

evidence in the individual record;[1] (3) no increment of less than 1 vd., or more than 30 vd., was used; (4) the division of groups for the curve is made midway between increments; (5) the observers were without practice (the test was made in an hour) in this test, but all accustomed to psychological experiments; and (6) the observers represent a select body, as admission to the University represents a survival process.

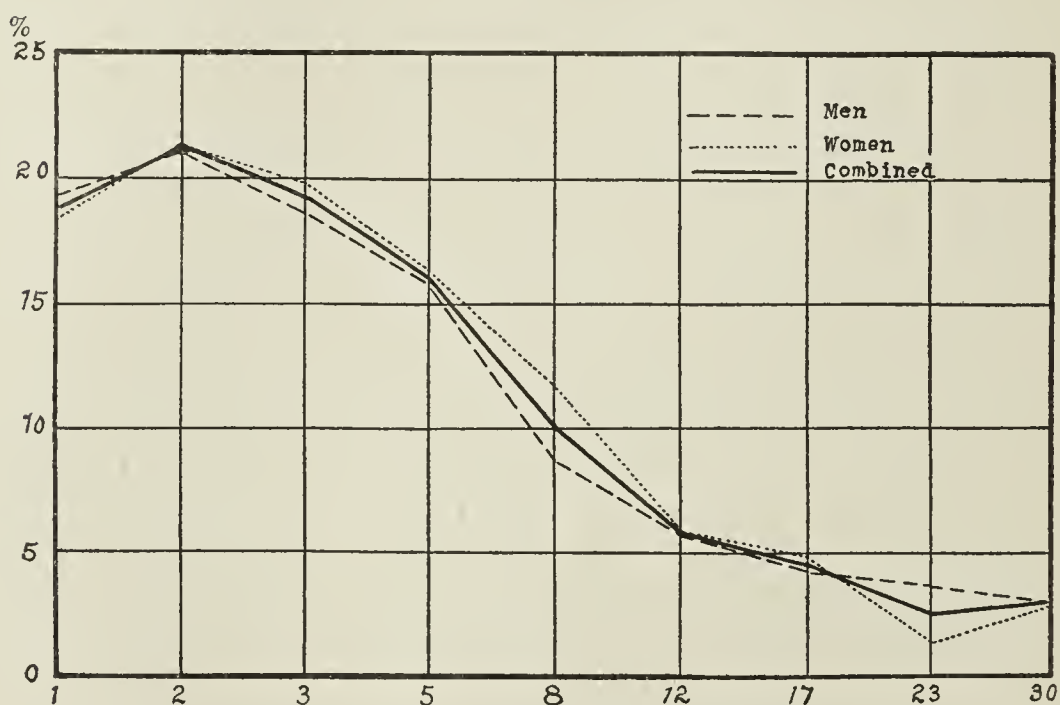This curve may be regarded as a norm. It may be expressed conveniently in a mathematical equation.



FIG. 1

3. *Comparison with Other Sensory Discrimination.* There are two characteristic features in the curve: the decided skewness, and the absence of a well defined mode. These together indicate that there is a wide range of difference in the capacity of individuals. There is no mode or unit which may be said to represent a norm. Instead of seeking an average, or norm, we must seek to establish some serviceable division into classes.

---

[1] Such corrections always lower the result. They may be found necessary in from 5 to 10 per cent of the cases.

In this respect this curve differs from curves representing sensory discrimination in other factors; *e.g.*, visual or tactual discrimination for linear magnitudes, form, or size; weight and pressure; and direction and intensity of sound.

4. *Extremes.* Individuals are found who can detect a small fraction of difference in pitch; these, if any in this group, are here included with 1. The marvelous keenness claimed by, and reported in the biography of, certain musicians has seldom been verified. A verified threshold of less than .3 vd. is rare. On the other hand defective tone discrimination gradually shades off into tone deafness. Such cases are included in 30 + in the above curve. Tone deafness probably prevails to the same extent as color blindness; but we have no satisfactory statistics of either.

## B. *Reliability.*

1. *The Mean Variation.* In the preliminary test we have the m. v. as a measure of validity in the record. Fig. 2 (from Franklin O. Smith) shows the distribution of the m.v. in a group of records as figured according to the directions given above. The figures at the base denote m.v.; those in the side of the curve, per cent of error. It ranges roughly, between .5 of a step and 2.5 of a step. A m.v. of 1.0 step or less indicates an entirely safe record; a m.v. above 2.0 steps proves the record worthless; between these extremes the reliability is roughly in inverse ratio to the magnitude of the m.v. The m.v. should therefore help us in determining the necessity of verifying a record by further tests. The standard deviation and the probable error if calculated, of course, serve the same purpose in a more exact way.

2. *Internal Evidence and Notes.* In the case of a large m.v. it is imperative to examine the distribution of the observations. Two extreme types are shown together in Fig. 3 (from F. Z. Wheeler). Each is based on 34 "sets" of trials. The average of the middle third is the same in the two records, namely 5.3 vd. The median is also approximately the same, but the m.v. for A is 1.1 and for B, 3.6. Thus the record for A

is fairly reliable, but for B utterly unreliable. But the record shows that B has 8 out of a possible 34 records right on $\frac{1}{2}$ vd. If he has no capacity in pitch discrimination the probability is 1 in 1024 that he should get a record on $\frac{1}{2}$ vd. in one set, *i.e.*, the whole column right, by pure chance. And from whatever point in the set pure chance operated, the chances of getting successive judgments right decreases in geometrical ratio. Thus, if his actual threshold were really above 5 vd. the chance that he would get 5 vd. (and all above) right is
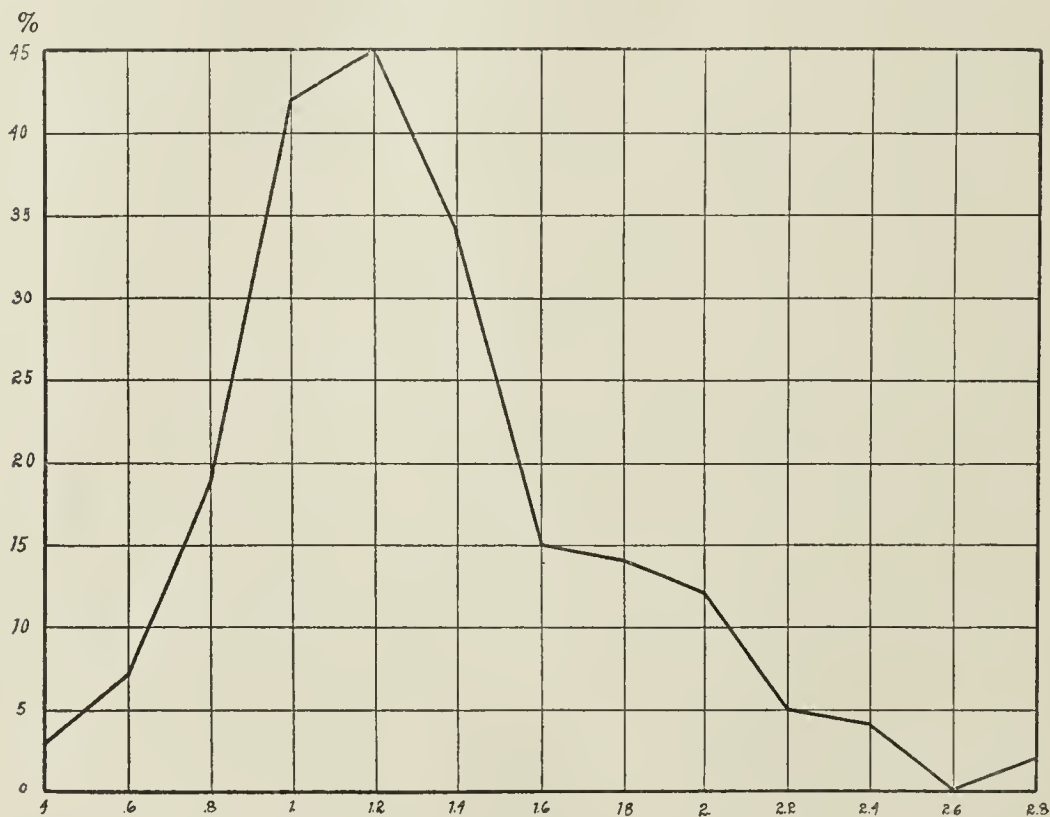


Fig. 2

1 : 2; 3 vd., 1 : 4; 2 vd., 1:8; 1 vd., 1:16; and $\frac{1}{2}$ vd., 1 :32. Eight records on $\frac{1}{2}$ vd. in 34 cases therefore almost amounts to certainty that B can hear a difference of $\frac{1}{2}$ vd. A correction must therefore be made, inserting the record $\frac{1}{2}$ vd. as one mode and the probable threshold.

But B's remarkable record may be due to one of at least two different causes. A glance at the original record shows that all the poor records were made in the first part of the test; toward the end he is almost perfect, which of course means that

in the first part of the test there was a lack of understanding of something. This adds even more assurance to our estimate of the $\frac{1}{2}$ vd. threshold. Another possible explanation of such a record, except for this last fact of progressive change, is that B had a good ear but was a "scatterbrain." It might also mean that there was some other disturbing element either subjective or objective in the test, but in no case would we have any doubt about the quality of the ear.

In accepting records we must always bear in mind the numerous sources of error—objective, in the apparatus or in the
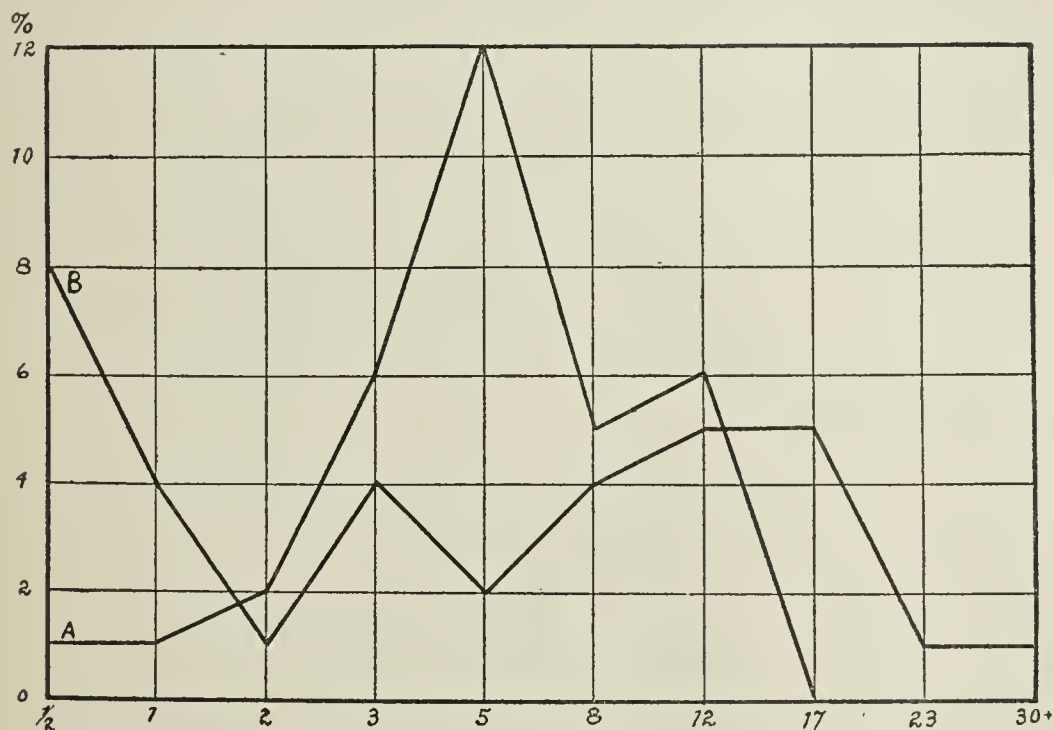


FIG. 3

manipulation by the experimenter; or subjective, due to ignorance, inattention, lack of effort, etc. The longer a careful experimenter works at it the more he learns of how conditions, apparently irrelevant, may play an important rôle. With patient practice we can learn to conduct the test on a group so that each individual record receives more than mechanical treatment. One must play over the full gamut of errors before he attributes a peculiarity in the record to the ear.

3. *"Final" Test.* In following the method proposed for the "final" test, with an individual or a homogeneous group, the reliability depends upon the number of trials, but first and last upon the favorableness and the accuracy of the control of the conditions under which the judgment is rendered.

4. *Proportionality of Increments.* It should not be assumed that the perceptibility of the difference in pitch varies directly with the difference in the pitch of the two tones. Tones that blend, such as the fundamental and the major third or fifth are readily confused. But, within the range of 30 vibrations, at this pitch, the variation is probably fairly direct.

5. *Difference vs. Direction.* The threshold for the perception of difference is of course different from the threshold for the perception of direction of the difference. While the former is the principle which operates in nearly all musical appreciation, the latter is usually taken as the measure of pitch discrimination, partly because it measures a capacity in musical expression but chiefly because it can be determined more accurately than the former.

6. *Relative Central Values.* In estimating the value of the threshold we must take into consideration (1) the form of judgment upon which it is based, (2) the step counted, (3) the kind of central value employed, and (4) the apparent plasticity of the threshold, illustrated in 7, 8, and 9, below.

The reasons for selecting the form of judgment here recommended and rejecting the threshold which depends upon subjective certainty have been stated in Sect. II A 9 above.

We have chosen to count the last judgment preceding the first mistake in a set because we have found empirically that it comes nearest to the actual threshold of just perceptible difference, when a sufficient number of sets of trials have been taken. Granting that this step is subliminal it would of course yield right judgments half of the time by pure chance, and the same is true of all steps below it, each considered by itself. On the other hand, it is well known that in this form of judgment errors will occur above the threshold for various reasons. The step we have adopted seems to strike a fair mean between

these two tendencies. Furthermore it represents a definitely definable procedure.

For central values we have at our command different kinds of averages, modes and medians. The ordinary average or arithmetical mean is out of question, since the large increments would count too heavily in proportion to the small. The average of the middle third largely obviates this difficulty and has the advantage of an average; it is therefore quite satisfactory. The mode is not always representative, but the mode, or modes, in the distribution of a set of observations must always be taken into consideration: we count it a secondary central value, often most significant. The calculated median (not the median of plain inspection) is the most representative value that we can use because it eliminates the extremes and gives a fair distribution of the most characteristic values.

The relation of the median as obtained in the preliminary test to the record by the method of right and wrong cases, as obtained in the "final" test has not yet been worked out to a satisfactory degree of certainty. A preliminary estimate of the relationship (by F. Z. Wheeler) makes the median to be not far from equal to 75 per cent of right cases. For 225 cases the median for the one hour preliminary test is 3.3; the threshold for 75 per cent correct cases is 3.1 in an individual test immediately following. Data are being collected for a direct comparison of the two central values in the same observations.

7. *The Cognitive vs. the Physiological Threshold.* In sensory discrimination of this sort we may speak of two thresholds: the physiological, which is set by the limits of capacity in the end organ; and the cognitive, which is set by cognitive limitations. Theoretically we always aim to reach the physiological threshold, but practically we often fall short of this and find a cognitive limit; *i.e.*, a higher threshold due to lack of information, best form of attention, interest, effort, etc.; or to disturbances of some sort. Usually inspection of a record or observations made in the test enable us to tell whether or not we have reached the physiological threshold. It cannot be judged by any single rule, although a small m.v. with one

well defined mode are pretty sure indications. This distinction is of greatest importance in classification, and in the theory of training.

8. *The Illusion of Difference.* One of the most bewildering experiences of a trained observer is to find that he is subject to very strong illusions of difference. Such illusions are common, in mild form, with reference to all liminal values, but here it is very pronounced and almost unbelievable.

Closely related to this illusion is the fact that the feeling of certainty in our judgment, near the threshold, is not as reliable here as in most other sensory discrimination. An extensive record of the degree of certainty by a trained observer, on a scale of five, for a long series of judgments shows that there is no close correlation between the degree of certainty felt and the actual correctness of the judgment.

Very much depends upon the direction of expectant attention. Listening to two tones physically of the same pitch one can make either tone seem higher or lower at will. This introduces a very serious difficulty in the experiment. It is exceedingly difficult to keep the observer from directing his expectant attention in favor of one tone or the other. Great ingenuity and watchfulness on the part of the experimenter is necessary in order to keep the observer as neutral as possible.

9. *Plasticity of the Threshold.* Except for the very keen mind with a keen ear, the threshold is not a sharply fixed limit. In the first place there are so many obstructions to the most favored hearing that we always expect fluctuations in judgments and do not assume that we have reached the absolute physiological limit, but rather a normal approximation which may be used as a norm. For most of our purposes it is quite as serviceable. On the other hand, it is not unwarranted to suppose that the surface of frequency for a given set of records actually, though in exaggeration, represents something of the flexibility of the physiological threshold.

## C. *The Tonal Register.*

1. *The Normal Curve.* Our tests have been made at one level, á, 435 vd. What knowledge have we of the relation of this level to other levels? Records have been published on individual musicians, notably Luft and Stumpf, whose discrimination is exceedingly fine, but a record of more nearly average observers is more directly applicable for us. Fig. 4



FIG. 4

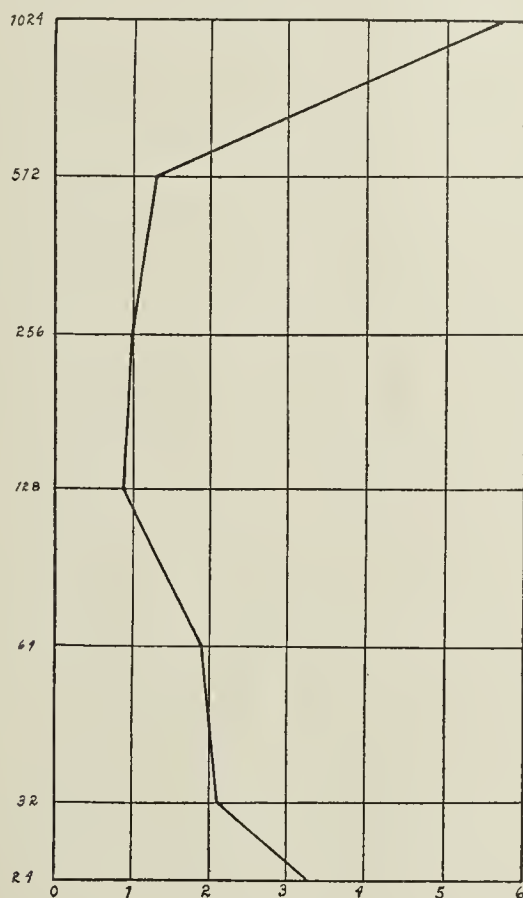(from unpublished MS by H. G. Schaefer) is a representation of the average for eleven good observers in the laboratory showing the relative sensitiveness to differences in pitch to different levels within the tonal register. From 100 vd. to about 500 vd., the least perceptible difference is almost a constant in terms of vibration frequency and a geometric ratio in terms of tone interval; but both ends of the curve

deviate from this rule.[1]   In general the individual curves are of the same type as the composite, regardless of whether the threshold is high or low.

For our present purpose the curve shows (1) that measurements made within the middle of the tonal register are comparable and convertible in terms of vibration frequency (2) that a moderate increment taken above the standard is practically equal to one taken below, in terms of vibration frequency; and (3) that 435 vd. represents approximately the most sensitive level in terms of the tone interval.

2. *Gaps.*   However, certain individuals have "gaps," *i.e.*, parts of the register for which they are less sensitive than the normal curve would indicate; or, they may even be tone deaf for a few notes, although keen in the hearing of other notes. In such islands of partial tone deafness the person hears the tone just as distinctly as elsewhere and feels that he can discern pitch difference, but, upon being tested, he finds that he is more or less helpless.   The writer has such a gap from 64 vd. to 45 vd., but it took a variety of crucial tests to convince him of it on account of the prevailing illusion of difference at this point.   Statistics in regard to the extent and frequency of this phenomenon are being collected.   Such gaps furnish most excellent opportunity for study of the traits of defective tone hearing by one who has a keen ear for other parts of the register.

### D.   Correlation of Pitch Discrimination with other Possible Factors in "Musical Ability."

Correlations have been worked out for musical education, environment, heredity, ability in performing, tonal fusion, appreciation, singing, rhythm, auditory imagery, auditory memory, etc., but it has not been found advisable to include results of that work in this preliminary report.   Reports on this subject will be published soon by G. H. Mount and others.

[1] Luft and Meyer (or Stumpf) did not find such deviation at the ends.

### E.  Practice, Age and Sex

1.  *Practice.*  The above correlation may be summed up in the statement that pitch discrimination is an inborn capacity which ordinarily reveals itself in full force without special training just as the psychophysic capacity in auditory acuity or visual acuity reveals itself spontaneously.

The problem of training is one of great practical significance.  It presents a fertile field for experiment in educational psychology.  Under our direction the problem has been approached through the following, among other, kinds of practice series in which intensive and exclusive training in pitch discrimination has been given by employing the "final" test above in a prolonged series of tests (*i.e.*, the training was gained from the continuous tests): with (1) persons with good ear and (a) musical education and (b) no musical education; (2) persons with average ear and (a) musical education and (b) no musical education; and (3) persons with inferior ear and (a) musical education and (b) no musical education. These tests are made both on adults and on children, and both by individual and group training.

At this preliminary stage, the following general conclusions may be stated:

(1)  When the proximate physiological threshold has been reached, practice is of no avail.

(2)  So long as a cognitive threshold prevails there is prospect of improvement by practice to the extent that the cognitive is above the physiological, limit.

(3)  This improvement is usually very rapid, often immediate and can usually be traced to the acquisition of knowledge, through experience or information about what pitch is, as distinguished from other attributes of tones.

(4)  In the majority of cases it is possible for the ingenious experimenter to discover the proximate physiological threshold to a fair degree of certainty in a well planned half-hour individual test, or in one heterogeneous and one homogeneous group test of one hour each; and, for most of the cases in which this fails, the same tests demonstrate positively that the threshold is only cognitive.

(5) The possibility of reaching the physiological limit in a single test depends to but very slight extent upon whether or not the person tested has had musical education; it is mainly a matter of expert skill and ingenuity on the part of the experimenter.

(6) Ordinarily musical education is not effective as a means of improving pitch discrimination. (From work of H. S. Buffum, F. Z. Wheeler and George H. Mount).

As an illustration of the degree of finality which may be obtained in a preliminary test we may turn to Figure 5 (from H. S. Buffum). Dr. Buffum experimented on twenty-five eighth grade pupils in a grammar school room. He first made a fifteen minute individual test of each pupil and classified them on this basis into three groups with modes at 3, 8, and 17 vd. respectively. The object was twofold: (1) to determine the effect of practice and (2) to determine the success of the preliminary examination. For this purpose he gave them twenty forty-minute periods of training.

The training consisted in taking records by the method of right and wrong cases. The figures at the bottom of the curve denote days, those on the side per cent of right cases; the fourth curve is a combination of the three. The results show (1) that for no group is there any evidence of improvement with this practice, and (2) that all except two children remained throughout the whole practice series within the group to which they had been assigned. Of these two, one who had been assigned to group III was immediately found to belong in group I as there had been a failure to understand the preliminary test; and the other, although retained in group II, proved really to be near the dividing line and could have been classified in group III. Evidently the physiological threshold had been reached in twenty-four of the twenty-five cases in the preliminary test.

The capacity for appreciation of music is partially inborn and partially the result of training. Thus, in judging the quality of an instrument or a voice, the expert hears and observes differences and peculiarities that entirely escape the untrained ear; and all differences in so-called quality or tim-
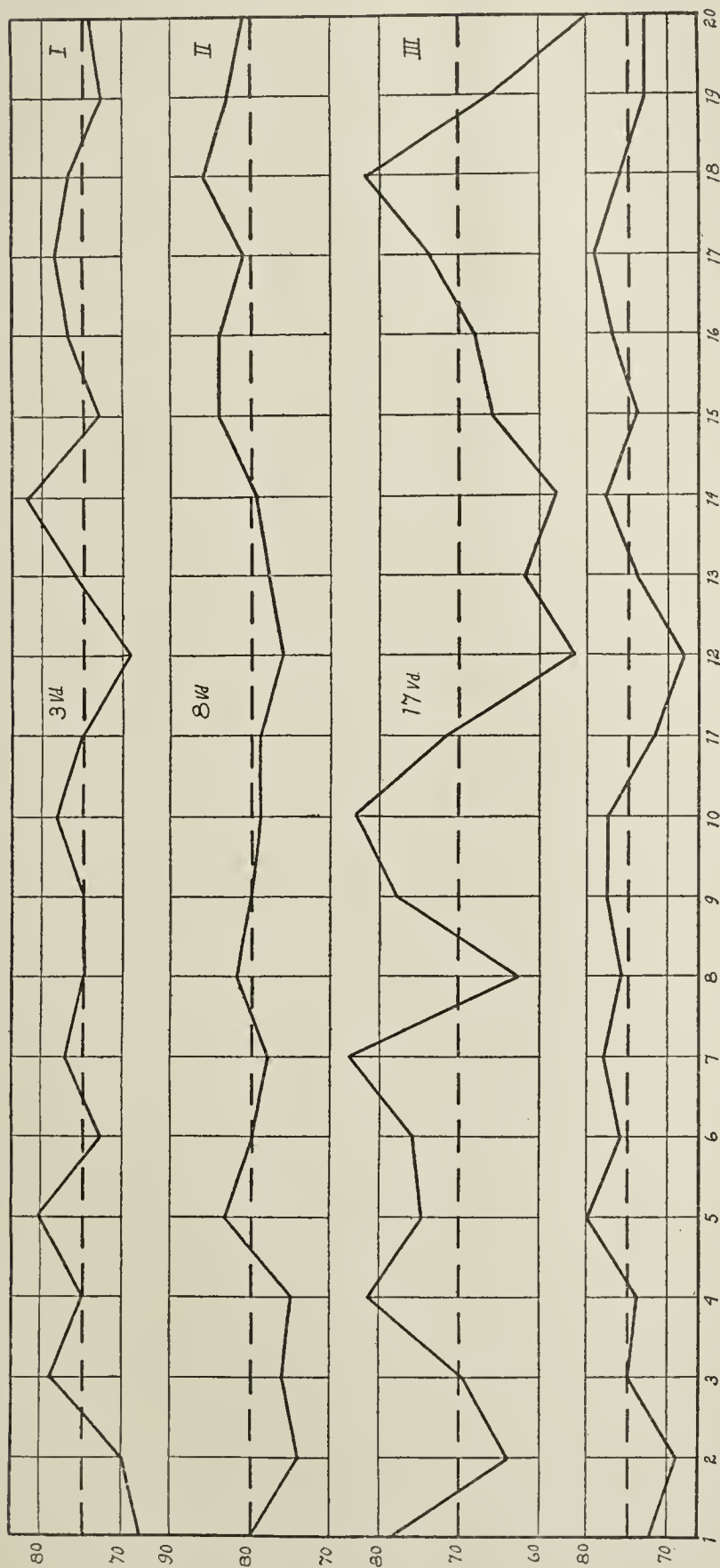
Fig. 5

bre of tones are reducible to pitch. But such hearing represents a complex process of interpretation, which can be mastered only after extensive training. The mere detection of pitch difference is, on the other hand, a simple process requiring only the slightest amount of training.

Now, to the extent that we control and simplify conditions so as to make our test a simple, sensory, direct comparison in regard to which there can be no misunderstanding, we have a valid test of pitch discrimination which approaches the physiological limit.

2. *Age.* A preliminary estimate of data on the variation of pitch discrimination with age may be summed up in the following general conclusions:

(1) The physiological limit is probably lowest in early childhood but remains fairly constant up to maturity.

(2) In a bright child with a good ear the physiological limit can be established for all practical purposes as early as the age of five.

(3) The slight inferiority of record which we find in a group of young children over the record of a group of adults is due to cognitive difficulties, which are not in the nature of a lack of skill in the sense of a slowly acquired ability but rather due to lack of knowledge. The university students have every advantage—musical education, maturity for reliability in observation, power of application, familiarity with the experimental conditions, etc., which is quite enough to account for the superiority of their group record over the record of a group of children. There is no such large difference of advantage between high-school pupils and grammar-school pupils, and their records practically agree.

3. *Sex.* Pitch discrimination does not vary in any constant manner with sex. The small oscillating differences shown in Fig. 1 are characteristic.

### F. Practical Use.

1. *In the Psychological Laboratory.* The main use of this measurement is of course in the psychological laboratory—

in teaching, and in research. But the discussion of that feature scarcely comes within the scope of this report.

2. *In the School Room.* While this is a "mental test," our report demonstrates that we are not justified in using it as a general mental test, or test of intelligence, as has often been done. It gives us a measure of a specific and peculiar capacity—probably dependent upon the structure of the sense organ; and that is a reason for employing it in the school room in this age of interest in individual capacity and adaptation.

Suppose that we find four children of equal age, advancement and general ability sitting together and one has a threshold, for pitch discrimination, of $\frac{1}{2}$ vd., another 3 vd., another 12 vd., and another 25 vd. They are to have singing lessons. How can we group them properly for this period? Nine years ago the writer proposed the following arbitrary classification as a tentative measure (*Educational Review*, June 1901):

Below 3 vd.: May become a musician;
3–8 vd.     : Should have a plain musical education (singing in school may be obligatory);
9–17 vd.   : Should have a plain musical education only if special inclination for some kind of music is shown (singing in school should be optional);
18 and above: Should have nothing to do with music.

This might serve as a basis for the adaptation and assignment of singing exercises, or substitutes for singing.

Exercises of this sort should be a part of the music course in school. They furnish the best kind of mental discipline; they direct attention to one feature in music in a concrete way; they stimulate interest and effort in observation and right execution of pitch; and, incidentally, they furnish instructive records which may be collected as a basis of classification. Children are entitled to know the facts thus learned. One period set aside for the interpretation of the records may be of untold value especially (1) to those children who are in danger of having a musical education forced upon them, although they are physically incapable of hearing music in

the true sense; and (2) to those children who first discover in this test their natural ability in this respect and are stimulated to take an interest in music.[1]

3. *In the Studio.* The first music lesson should be a pitch test. There are scores of other factors which go to make up musical ability, but unless a person has a good ear, he will fail to appreciate and will be unable to produce music which depends upon the ear. We often find music students who lack a general interest in music because they have no "ear."[2] A person may have a good ear and yet be lacking in other musical capacities, for example, rhythm, hearing, consonance and disonance, auditory imagery, affective tendencies, etc. Therefore, this pitch test should be followed up with other tests on musical capacity, but this is the first and fundamental.[3]

4. *General Cautions.* As this measurement constitutes a quantitative rating of a capacity of an individual in such a way that it may act as a powerful stimulus or an effective deterrent, we cannot be too emphatic in demanding that information about rating and advice based upon it should be given not only conservatively but with prudence and reserve. In one case the rating may blast a life ambition, in another, it may be the means of discovering the individual.[4]

[1] The place for the test is in the first six grades. It can readily be adapted to the ability of the children of the different grades. Thus the test should be short and answer may be made by simple means, such as showing of hands and counting the number of mistakes; but for the purpose of classification greater care must be taken. As a rule, the younger the child the more necessity there is for making the test individual.

[2] What a blessing to a girl of the age of eight, if the music teacher would examine her, and, if necessary say, "much as I regret it, I must say that you would find music dull and difficult, and I would advise you to take up some other art." What a blessing if that child could be started right; but current theory and practice is against her. There is too much faith in what music lessons can do for a person without native capacity. If we are to have musical ears, we must be born with them. That is the probable finding of current research.

[3] Among such tests in progress at the present time are the following: the sense of rhythm, and rhythmic action, tonal fusion (consonance and dissonance), auditory imagery, auditory memory, discrimination for intensity of sound, and vocal reproduction of a tone.

[4] The writer recalls the case of a university junior who had taken music lessons for twelve years, but took no real pleasure in it and made but little headway. When she found that her rating was 12 vd. she exclaimed, "That is it. That's what I have thought," and quit music.

And in such weighing of evidence we must ever bear in mind that there is a compensation for this lack of ability in hearing fine pitch differences. An inferior ear hears tones as tones, but fails to hear small differences in pitch and may fail to hear the true pitch of a clang; but this may be an advantage as, *e.g.*, in listening to instruments out of tune or inferior rendition. He may enjoy his own singing even though he cannot carry a tune. While such partial tone deafness prevents the critical and most varied enjoyment of music, it permits and even encourages idealizing and undisturbed, crude enjoyment of tone productions provided the individual has some "music in his soul."

We must also distinguish between advice to possible musicians and to others who may desire to have some music for "home consumption." There is an element of music in the bones, the drum, and the tambourine. The associations with music are often rich and valued in themselves. Even a bare intellectual familiarity with matters musical has its charm. The words of songs get a setting from music even in the unmusical ear. We must, therefore, bear in mind that pitch is not the only avenue to musical appreciation.

In conclusion, then, let us post the following warnings:

(1) A record should not be the basis for rating and advice until it represents the proximate physiological threshold.[1]

(2) No classifying tests should be employed with groups except by persons who are capable of taking proper experimental precautions and making a safe interpretation.

(3) Whenever possible, rating should be considered personal, confidential information.

(4) Persons with poor rating should not be discouraged in their enjoyment of music; there is much in music besides

---

On the other hand, it should not be supposed that the parties concerned always take advice. The writer was called in to tell why the oldest of three children in a musical family was not making satisfactory progress with her music; the younger children clearly outclassed her. Pitch test showed: boy, age 11, 1 vd.; girl, age 15, 1.2 vd.; girl, age 19, 13 vd. Explanation and advice was given on the basis of this rating, but the parents sent the oldest girl to a better conservatory.

[1] The cases left in suspense are no worse off than they have been heretofore. With further patience they may be reduced.

pitch of tone, and much enjoyment of music does not depend upon the hearing of fine distinctions of pitch.

(5) Although perhaps the most fundamental, pitch discrimination is only one of a score or more of technical tests which, taken together, constitute a measure of "musical ability." Conclusions must be limited to the element tested.

# METHODS FOR THE DETERMINATION OF MENTAL IMAGERY.

## By James R. Angell.

### Part I—Examination of Methods in Use.

In few fields of psychological investigation has the situation changed so completely in the last few years as in that covered by the title of this report. When the writer in 1906 undertook to make a comparison of the tests then in use, the literature relating to such work was relatively circumscribed and the devices then in vogue relatively simple and supposedly fairly accurate.

The author imagined, as did others, that his task would consist in making a careful comparison of the recognized tests, eliminating untrustworthy features, possibly adding a few new ones, but at least standardizing the tests in terms of one another. Although this seemed a fairly large contract, it appeared to be quite definite and was entered upon with an optimistic spirit.

Since that date a large amount of excellent work bearing upon the subject has been produced and complexities have been revealed previously little suspected. This development was foreshadowed by the first investigations made by the author in 1907 and indicated in the preliminary report made to the association at the annual meeting that year. In attempting to compare the relative merits of several familiar forms of procedure, he became convinced that the problem was much more baffling than had ordinarily been recognized and much broader in its outlook than had been admitted. The present situation is far from satisfactory, as will be made clear below, but the author herewith offers certain recommendations for tests in accordance with the promise made at the meeting of the association in 1909.

It is not without interest to remark in passing that the period during which we have heard most about imageless thought and about the fallacious importance attributed to imagery in general, should coincide with the period in which many of the best and most exhaustive studies of imagery have been made.

Before proceeding to present the recommendations, it will be desirable to pass in critical review the several forms of procedure at present in use in connection with the study of imagery. This survey is not offered as exhaustive of all variations in detail. But it includes the main forms of which the author knows, so far as he has found them productive.

The methods employed to diagnose imagery fall into two main divisions (as regards their purpose, if not as regards their results[1]); (1) objective, in which the subject is set to perform a certain task so devised that the results he achieves will disclose forthwith the kind of imagery he uses. (2) Subjective, in which by various devices the subject is given tasks to perform which will presumably facilitate his ability to discover and report the imagery he uses. Evidently the first form of test is the ideal thing, for, if successful, it avoids the unreliability of merely introspective analysis and description. Objective methods may be, and ordinarily are, supplemented with introspective analyses, so that combination subjective-objective methods, rather than those of purely objective character are in general to be distinguished from the purely subjective modes of procedure.

DIVISION I.  OBJECTIVE METHODS.

*A.* One of the earliest and still most frequently used of objective tests is that in which material is given to be memor-

---

[1] Several other classifications have been proposed and a different meaning is attached to the classification here used by Lay, in his monograph on Imagery (*Psychological Review Monographs*, vol. ii, no. 3), who, as I understand him, makes the term subjective apply to methods in which the subject exercises his own introspective analyses. The term objective he applies to all methods in which there is any effort at experimental control. Dr. M. R. Fernald in a forthcoming monograph recognizes three groups, (1) subjective, (2) objective and (3) a combination subjective-objective class.

ized visually, or aurally, or with the assistance of motor activities as in spelling and writing. The form in which the memorizing is most effective is assumed to indicate the preferred and dominant type of imagery.

This procedure has been modified in various ways, *e.g.*, by offering to memory objects, sounds, and words, the latter designating sense experiences of the eye, the ear, the muscles, etc. The terms best remembered are again assumed to indicate the ideational type, visual characters being best remembered by a visualizer, auditory characters by those who chiefly employ auditory images and so on.

Tests of this type may have an undoubted practical significance in connection with studies in economy of learning and permanency of retention. They may certainly be so administered as to show over what sensory arcs the best results may be achieved in assimilating information of various kinds. They may convey information as to dominant interests. They also possess an unquestioned value as a stimulus to intropsection. But as objective tests of imagery, apart from introspection, they have few virtues and no reliability. For instance, a subject may learn a series of objects visually as rapidly as he learns them from hearing their names pronounced, and yet his recall of the object may be in terms of auditory-motor words. He translates the visual objects into words and at times recalls better such a series than one learned originally from words themselves. The writer is himself an illustration of this type. Were the test employed to give objective information in such a case, it would clearly result in a false conclusion. Numerous other difficulties of this same variety are met with in the effort to use the results objectively.

*B. Methods of (1) Distraction or disturbance, and (2) of 'helps.'* (1) The theory on which these tests proceed is old in experimental psychology, but relatively new in its application to the study of imagery. The distraction test as used, for example, by Cohn,[1] has many variant forms and

---

[1] *Zeitschrift f. Psychol. u. Physiol.*, 1898, vol. iv, p. 161. Cf. also Meumann, *Vorlesungen z, Einführung in d. Exper. Pädagogik*, Leipzig, 1907.

has been employed in combination with a number of other methods here mentioned. Its application may be illustrated in the realm of sound.

If a person who relies mainly upon auditory imagery is memorizing a series of words presented visually, it is assumed that a stimulation of the ear will increase the difficulty of his learning much more than it does for a learner of visual type. Similarly, distracting visual stimuli will react most disadvantageously upon the memorizing of one of visual type, and motor activities of the vocal mechanism or the hand, will most hamper one of motor type.

The assumption is broadly true on the positive side; that is to say, users of the corresponding imagery are often seriously embarrassed by such distractions, but in practice it is found quite impossible to infer, *e.g.*, that one who is disturbed by an intruding noise is therefore a person who uses auditory imagery predominantly, and the amount of the disturbance caused is in no wise regularly related to the amount of auditory imagery employed. The general distraction of attention produces anomalies in the results which render the method highly ambiguous except when conjoined with adequate introspection. Moreover, it is found quite impossible in the case of some individuals to devise any distraction which will measurably decrease the use of their preferred forms of images, short of such a distraction as will preclude any learning at all. Furthermore, cases have been found by Dr. M. R. Fernald in the Chicago Laboratory where a vocal distraction has emphasized rather than impeded the auditory-vocal imagery used in learning and recall. The vocal distraction has on the other hand proved very disturbing to the visual processes with certain subjects. In my own case such a distraction appears to be effective not so much by its stoppage of the articulatory accompaniment of the learning (it stopped this very imperfectly), but because it breaks up the associative processes of perception, so that I do not apprehend the meaning of the word easily or correctly.

It may be added in passing that although relatively sat-

isfactory auditory and motor distractions have been determined, equally satisfactory visual distractions have not been discovered.

(2) The method of helps is even more ambiguous in practice, although equally simple and obvious in theory. The commonest form of the test is in connection with such tests as were mentioned under I *A*. Suppose several series of digits are to be learned. A person of motor type would presumably be most helped by being allowed to pronounce the digits, aloud or under the breath, or possibly by being allowed to make the movements of writing them. The auditory imagery as judged by many memory tests would presumably be helped by a rhythmic arrangement of material. Visualizers might be helped by arranging material in certain groups or patterns.

As in the method of disturbance or distraction, the general assumption underlying this program is probably correct, *i.e.*, that there are variations in procedure which will assist particular individuals because of their peculiar imagery. But the assumption as applied is much too simple and schematic to meet the complications of actual mental life. For example, our experiences in the Chicago Laboratory seem to show that persons who use a good deal of visual imagery may be quite as much helped by rhythmic arrangement of material as audiles, and an occasional audile is disturbed rather than helped by rhythm, at least by those imposed by the experimenter.[1]

Speaking well within the mark, then, one may say that the methods of helps and distractions when taken apart from introspection, are quite untrustworthy as thus far developed. Both are, however, extremely useful in stimulating introspection by throwing into strong perspective the processes actually employed in learning a given material.[2] It must

[1] The terms, visualizer, audile, etc., are simply used to indicate persons who in the test under discussion belong to these groupings. The ambiguity of such terms for general classification is dealt with at a later point.

[2] Both may be used either with children or with adults in a way to throw valuable light on the conditions practically desirable for learning, but these merits must not be confused with the objective adequacy of the method for revealing imagery.

also be clearly and emphatically stated that the objective results of these tests often afford most striking confirmation of the introspections.

*C. Kraepelin's test.* A test of Kraepelin's to which certain writers have assigned objective value consists in asking subjects to write for five minutes lists of words designating objects primarily characterized by color, in another five minutes objects primarily referring to sound, etc. The lists when compared are supposed to indicate the scale of distribution for different kinds of imagery, the dominant type being represented by the most extended list. I do not find that Kraepelin himself ever laid claim on behalf of the test to objective validity of imagery analysis. He speaks of it as giving interesting and suggestive information about mental organization.[1]

It undoubtedly has some value as indicative of verbal imagination, for the test can be, and often is, carried out by subjects purely on the basis of verbal imagery. In my own experiments it has appeared to possess, as nearly as possible, no value whatever for the objective determination of type, or for the objective ascertainment of the presence or absence of any special form of imagery.

To mention only one of several ambiguous cases illustrative of the difficulties the method presents, we may notice that such a word as 'bird' is suggestive of visual and of auditory qualities in essentially like degree, and the presence of such a word in a list of reactions would leave one quite unable to decide whether it indicated auditory or visual imagery in the reagent. It is surprising to find how many terms suffer from similar kinds of ambiguity.

*D. Method of 'Style.'* The same approbation and the same criticism may be accorded the so-called method of 'style,' which is supposedly available for the determination of types through an examination of the vocabulary employed in a writer's works, especially when these are of a descriptive character. The method *may* often give correct results within

---

[1] *Psychol. Arbeiten,* 1895, vol. i, p. 73.

the narrow field to which it applies, but it is never possible to be sure of these and the outcome is certainly misleading in its indications in many instances.[1]

E. *Tests on Imagination Rather than Imagery.* Paradoxical as it may sound, there are in use several forms of test sometimes used in a purely objective way, which are known as tests of imagination, although little or no imagery need be actually involved in performing them.

The 'ink-blot' test is a case in point. Ink-blots are shown to the subject and he is asked whether the blot suggests any object to him. If he replies promptly and perhaps sees several things in the blot, he is credited with an active visual imagination. If he fails to do so, he is alleged to be deficient in visual imagery. The data are too few to say in what degree the actual possession of vivid and profuse visual imagery is correlated with expertness in this class of test. But one thing is entirely certain, that persons may make a very respectable showing in the test who employ a relatively slight amount of visual imagery.

These tests and others like them are really tests of perceptual fertility rather than tests of imagery in any proper sense.

### Summary.

The preceding paragraphs are offered as showing conclusively that no purely objective methods of imagery analysis are at present reliable. There may be other tests of which the writer has no knowledge that satisfactorily fulfill the necessary conditions. In any event full recognition must be accorded the valuable confirmation of introspection frequently afforded by tests of this type.

DIVISION II. COMPOUND SUBJECTIVE-OBJECTIVE METHODS.

Many of the following methods are at times used purely subjectively, though they all contain elements capable of use for objective interpretation. But, as in the preceding group,

[1] Cf. A. Fraser, *Amer, Jour. Psychol.*, 1891, vol. iv, p. 230.

none of them possesses any serious significance apart from the introspective evidence gained by it.

*A. The Questionary.* The method of Galton, modified and elaborated by many subsequent psychologists, is too familiar to require extended description. Specific questions are set which invite the production of images from each sense department, and the subject's introspections are recorded. Quality, number, vividness, objectivity, persistency, controlability, etc., are reported upon.

Dream imagery may be treated in the same way and many interesting experiments have been made on the control of dreams, the relation of dreaming to depth of sleep and so on; but the subject seems to be a trifle aside from the main purpose of this report and will not be further considered at present.

In the hands of competent introspectionists a good questionary gives quickly a fair view of the capacities of the individual for the voluntary arousal of imagery. The method is too difficult for occasional persons untrained in introspection. It is probably rarely successful in informing us with accuracy or certainty of the forms of imagery actually most employed and their manner of employment in common mental tasks. Moreover, we cannot judge that specific images are entirely lacking because they are not reported in replies to a questionary. Other forms of test frequently bring out images which escape notice in the questionary method.

*B. Association from a Stimulus Word.* A series of words is shown or read one at a time and the subject at once reports the ideas called up and the imagery involved. If the stimulus words are skillfully collected, the results are often very satisfactory in the calling out of various forms of imagery and in giving indications of the type prevailing for this kind of task. The reactions themselves, as given in words, are often suggestive, though by no means adequate for an objective test. The test is less significant for persons who use chiefly word imagery, and it may lose value still further in the case of the 'automatic' reagents for whom a reaction word comes solely and immediately as a suppressed enunciation.

This is the main limitation of the method, that it tends to magnify word imagery. Of course it gives no adequate indication of the distribution of imagery in ordinary activities.[1]

C. *Method of Letter Squares.* This method has been employed with sundry modifications by several psychologists and in its most frequent form is known by Binet's name. Letters are presented in a frame-work of squares so arranged as to bring the squares into vertical and horizontal lines like a chess-board. In the simplest form of the test the subject reads the lists a certain number of times visually and then attempts to repeat them. Full introspections are invited.

It is commonly assumed as a matter of objective import that a visual-minded person can reproduce the series in orders other than that in which the list was presented, more easily than one using no visual imagery. It is further assumed that the nature of the errors made will reveal the character of the imagery. If confusions are made where the appearance of the stimulus letters is similar, as E and F, while the sound is different, visual imagery is assumed to have been used. If the confusions are of letters of similar sound, such as Q and U, it is presumed that auditory imagery was involved.

This method, apart from the many complications which have been introduced into it—in my own work, for instance, where it has been combined with features such as distractions and helps taken from other tests—shares with the questionary and the word methods the advantages of stimulating introspection in a very useful manner. Like the word method it is somewhat superior to the questionary in simulating more nearly a normal mental task and its results are accordingly more likely to be indicative of ordinary imagery processes.

The objective features of the test are quite untrustworthy. For instance, persons who use visual imagery, but use it poorly, may be able to visualize the separate letters in memory, but be quite as embarrassed to reproduce them in an order other than that in which they were learned, as persons

---

[1] The test in this form is known by the name of its inventor, Secor. *Amer. Jour. of Psych.*, 1899, vol. xi, p. 225. The literature dealing with the general problems of association reactions contains much available material.

using auditory imagery. Moreover, persons using auditory-motor imagery are by no means unable to vary the order in reproduction. The method as generally used affords no adequate means of distinguishing auditory from kinaesthetic ideational material, leave alone the distinction of sensational from ideational kinaesthetic elements. Again, the confusions made are frequently ambiguous, so far as concerns indicating the imagery employed. When the letter B is confused with P, and M with N, it is evident that either auditory or visual similarity would afford an adequate clue. On the other hand the confusions are often extremely suggestive.[1] Owing to the remarkable facility with which some individuals 'shunt' from one form of imagery to another, or use two or more forms conjointly in a given test, as will presently be further explained, and especially in view of the ease with which they may use words as substitutes for objective forms of imagery, it is impossible to infer with confidence the domination of a given kind of imagery merely from the objective results attained.

A case peculiar to my own learning processes in this test may illustrate another possible ambiguity. If I am obliged to repeat the words before they are completely learned, my recall is likely to be dominated by visual processes. On the other hand if I am allowed to proceed until the learning is quite perfect, the recall is likely to be mainly in auditory-motor terms, and the more perfectly automatized the act becomes, the more I lose the visual element. Judged at one stage of the process, I should then be set down as a visualizer; judged at another stage, I should be auditory-vocal-motor.

*D. General Tests on Memory and Sense-Discrimination.* As these tests just mentioned concern primarily visual and auditory imagery with a subordinate recognition of motor images, it may be remarked in passing that all tests on memory can be employed to throw light on imagery analysis,

[1] On the general subject of linguistic lapses as it bears on this subject, see, Bawden *Psy. Rev. Monog.*, 1900, vol. iii, No. 14.

and most of the recent memory investigations have been characterized by detailed description of the imagery discovered. Nonsense syllables, numbers, words, letters, colors, tones, noises, every kind of material has been studied in this way.[1] The differences between immediate and mediate recall, between mere recognition and independent recall, between memory and imagination, have been pointed out for many cases. Tests on sensory discrimination have been similarly employed. Wolfe's tests on tonal memory, and Lehmann's tests on the discrimination of grays, may serve to illustrate the point.[2]

The comments made on the merits and defects of the immediately preceding test apply here too, and need not be repeated.

*E. Tests Based on Description.* Tests of this character have been used for various purposes and are clearly applicable to the investigation of certain phases of imagery. The tests can be complicated indefinitely especially by combination with distraction and help devices. The observer is shown an object or a picture, preferably one in colors, for a definite period, *e.g.,* ten or fifteen seconds, and then at once, or at a later time, is asked to describe what was seen. Introspections are then asked concerning the process during the original stimulation and that involved in recall. The test proves in practice to elicit a good deal of non-verbal imagery not easily stimulated in other tests.

To illustrate: a series of words memorized by vision, may be recalled almost wholly through vocal-kinaesthetic imagery, thus leading to the inference that a subject uses no visual memory material. The same person if given a picture to observe and subsequently describe, may find abundant vis-

---

[1] Several experimenters have commented upon their observations that visualizers tend to learn more slowly than audiles. The experiments carried on in the Chicago Laboratory have not confirmed this generalization. Several of the quickest learners have been visual, and no general results have been reached which substantiate this view.

[2] Wolfe, *Phil. Studien,* 1886, vol. iii, p. 534. Lehmann, *Phil. Studien,* 1889, vol v, p. 96.

ual imagery present. It will often be found extremely illum-
inating to allow the subject to see the object again after he
has given his description.[1] The discrepancies between the
stimulus and the memory imagery are often most suggestive
as to the influences under which imagery is built up and the
second inspection is remarkably helpful in inciting the subject
to greater accuracy.

In this connection an interesting confirmation of the
extremely subtle shifts in imagery resulting from slightly
altered conditions has been observed by Dr. M. R. Fernald.
Observers have been found whose imagery alters to a con-
siderable extent when they know before the experiment that
a description of the picture or object is to be asked of them,
as compared with their procedure when they are ignorant
that such a request is to be made. In the former case words
are more likely to be used during the inspection of the picture
or object and the recall may be largely in word form, whereas
otherwise the whole process—inspection and recall alike—
may be dominantly visual with little or no evidence of verbal
material.

Skillfully conducted, the test is an excellent one which
can be made to simulate with practical fidelity the condi-
tions of common mental processes. Its limitations are sim-
ply those incident to all subjective tests, *i.e.*, difficulties in
introspection and lack of generality in the results. The imag-
ery employed in describing one set of visual experiences, for
instance, may differ in many particulars from that used in
recalling another set.

*F. Tests on Reading.* We may mention next tests involv-
ing reading, (*a*) aloud by the experimenter, the subject lis-
tening, (*b*) by the subject either silently or aloud with refer-
ence either to (1) immediate introspection on the process of
apprehension, or (2) also and primarily with regard to the
mode of recall and its accuracy, whether tested at once or
after an interval, whether for verbatim transcript, or for the
giving of the general meaning. It is ordinarily advised that

---

[1] Except in vision the test is not significantly different from memory tests elsewhere
discussed and we therefore mention only the visual case at this point.

the materials for the purpose should cover a considerable range of intellectual interest and should include both prose and verse.

An interesting variant on the test which has often been used as a test on verbal imagination, and as a test on general intelligence, consists in the submitting to the subject prepared sentences in which occasional words or parts of words, or even whole phrases have been left out, and the subject is called upon to supply relevant material for the gaps. Sometimes nouns and verbs are given and the subject must make a sentence. This is similar to the old word-building game where the prize goes to the competitor who can make the most words out of combinations of the letters in the stimulus word.

Obviously the same limitation belongs to the tests of the immediately preceding paragraph as that which we have already pointed out in connection with the test I E, namely, that certain subjects can perform them using little or no discernible imagery and that with many persons introspection is extremely difficult. On the other hand for persons of strong auditory-motor verbal tendency the tests may prove very useful.

This general group of tests in any of its forms is essentially subjective, so far as concerns its contribution to the analysis of imagery; but many secondary objective indications are found in the actual procedure. It is a distinctly useful group for practiced introspectionists, though in general unduly difficult for untrained persons. Its greatest merit is that it employs common and fundamental forms of mental activity for investigation. When successfully controlled, it gives information bearing directly on actual norms of mental conduct. It is easily complicated by distraction and help methods, often with excellent results in the facilitation of introspection.

*G. Tests on Sensory Similarity in Recall.* This test has been devised by Dr. M. R. Fernald and, so far as I remember, has not been previously used for this purpose. The test has various forms. In one form the subject is given a word,

say 'tone,' and asked to write all the words he can which rhyme with it. His introspections are asked and especially as concerns any words which suggested themselves and were rejected. Again, a subject is asked to write all the words he can recall which end with certain letters, say 'ine' regardless of the pronunciation of the word. In other forms of the test lists are prepared either with or without a letter-square frame, in which words pronounced alike, but spelled differently are to be memorized. In other lists words spelled alike, but pronounced differently are dictated. Here, as in the distraction tests, one might assume that persons using visual imagery would fare best in writing lists of words ending in the same letters, and those with auditory imagery would do best in writing lists which rhyme. Again one might anticipate that audiles would be more confused in memorizing the lists containing words pronounced alike and spelled differently, whereas the visualizer would fare worse on words spelled alike and pronounced differently.[1] But like all attempts at objectively valid tests, the results here are untrustworthy. Nevertheless, the 'rhyming' brought out auditory or vocal-motor verbal imagery in practically every case examined, and persons who use such materials normally found it present here in great distinctness and profusion.

If persons were rigidly confined to the use of a single kind of imagery, objective tests of this variety might be practicable. In point of fact, however, as we have repeatedly noted, individuals are rarely or never so circumscribed as regards their available imagery that they cannot at need fall back upon some variety not regularly used, or by calling in supplementary aids manage to master the situation. For example, an individual who commonly uses visual imagery for purposes of recall of objects may, in the recall of words, spelled alike but pronounced differently, resort either to the

---

[1] Such words as the following have been used: cite, site, dew, due, scent, cent, sent; lead, as a verb and as a noun, dove as a verb and as a noun; similarly the words wind, row, tear, wound, and the following words, which also have two pronunciations, close, bass, live, vase, gill. It should be said that Dr. Fernald undertook no objective application of the test.

use of auditory and kinaesthetic imagery, or to the making of supplementary associations of a helpful kind based on meaning. The objective result can consequently never be used apart from introspective confirmation.

As a stimulus and aid to introspection the tests are extremely useful; especially as regards the exhibition of shunt systems in imagery, by means of which commonly unused forms or devices may be successfully enlisted when occasion requires. The work is sufficiently akin to normal mental tasks to give the introspections genuine significance. Like all the tests thus far noticed, with the possible exception of the questionary, disproportionate emphasis is probably given to visual and auditory materials at the expense of motor imagery other than the linguistic type, whereas the remaining forms of imagery are practically neglected entirely.

Dr. Fernald has used the test in connection with letter-squares employing in pairs words of the kind described and adding distractions, but no further comment need be made on the case.

*H. Tests Involving Spelling.* These can be used in a number of ways, the words being spelled orally, or in writing, with eyes shut or open, spelling forward or backward. Tests on backward spelling seem to have been used less than one might anticipate. Like the previous test, this type may be complicated with distraction and help devices.

It is an admirable form of test to use in conjunction with tests on description, in as much as it brings into the foreground peculiarities of verbal imagery as compared with the imagery of objects. This virtue constitutes also its main limitation. Other forms of imagery are notably infrequent with most subjects.

*I. Tests Involving Simple Problems.* Simple arithmetical problems may illustrate this method. The addition of numbers seen or dictated, the multiplication of others, serve to give material for introspection. The problems can of course be complicated indefinitely and can be connected with practical situations after the manner of the text-books on arithmetic. Geometry and Algebra can be similarly drawn upon.

Tests on geometrical forms are peculiarly successful in throwing visual imagery into the foreground and algebraic problems often emphasize motor elements. Mental manipulations of the cube have been frequently employed: *e.g.*, think of a cube; cut it through from top to bottom on a line joining two diagonally opposite corners; what shapes have the resulting pieces? Problems in ethics, in physics, or in deportment, anything one chooses may be employed, and when the conclusion has been reached, the subject is asked to describe as fully as possible the mental material used in reaching a solution. Or he may be unexpectedly interrupted during the process and asked to report his introspection.

For persons unpracticed in introspection the problems must be very simple. (The 'imageless thought' advocates must be reckoned with in interpreting the results of any such tests). There is always danger that the imagery recorded may not have been essential to the solution, but rather a fringe of more or less irrelevant suggestion. If this danger can be safeguarded, and only by accurate introspections can it be, the test can be made very useful, and it has as not the least of its merits, the normality of the tasks set.

*J. Tests with Writing.* A great deal of experimentation on the processes of writing is more or less directly available for purposes of analyzing imagery. When so applied, the aim is of course to devise situations in which the imagery employed may be thrown into the foreground of consciousness so as to be introspectively available. Writing from dictation, from printed or written copy, from memory; writing prose or poetry, writing not only forward but backward, and also upside down, left-handed, and in looking-glass script, writing amid various distractions, all have been tried and with most interesting results. These tests lead naturally to the mention of a large group of tests devised primarily, as were the foregoing, for the study of voluntary control.

*K. General Tests on Voluntary Control.* Here may be mentioned in addition to the tests on reading and writing, tests on the use of habitual coördinations such as are concerned in speech, in walking, in grasping objects, in muscular reaction

to various kinds of stimuli, and especially, tests on the securing of control over new coördinations. The latter may be illustrated by the experiments on learning to move the ear, learning to move the final joint of the fingers, independently of the lower joints. The learning of acts of skill such as typewriting, skating, bicycle riding and the like, represent another important group.

All of these tests, except those on learning to control new muscles, enjoy one merit in common, *i.e.*, that they represent in the most concrete and tangible form ordinary kinds of occupations which are largely free of the merely laboratory flavor. They all suffer too, under a common disability, *i.e.*, that for unpracticed observers, the introspection is often extremely difficult, and sometimes essentially impossible. Many sensory elements are inevitably interwoven with the ideational ones, and this adds a very distracting complication which destroys confident introspection on the imagery for not a few observers, even of the practiced sort.

*L. Tests for Kinaesthetic Imagery.* The experiments just referred to on voluntary control frequently bring into marked prominence the presence of motor imagery so-called, and most of the tests previously mentioned serve to disclose the part played in mental life by the articulatory imagery. It might seem that any further mention of motor imagery would be superfluous. (We waive here, as previously, the question whether such alleged imagery is or is not invariably connected with muscular excitation, with the sensory results of which it is merged.) But it is worthy of mention that in tests specially devoted to the analysis of motor memory we sometimes secure clearer evidence of this type of imagery than in such tests as we have heretofore described, in which attention is in the nature of the case less immediately directed at the motor act, and more immediately directed at the objective results of the act. Tests on the estimation of the extent of successive arm-movements may illustrate the point. Many subjects to be sure, make judgments of this kind by means of secondary criteria, such as the time and rate of movement, or by immediate discriminative reactions, in

which no kinaesthetic imagery appears. But others resort
very definitely to the kinaesthetic image, or to something
which they regard as this image, and compare with it the
second reaction.

Obviously any muscle group may be submitted to tests of
this character, and any exploration of motor imagery which is
to be complete must make some such general survey before
it can be accepted.

A further form of motor test which has been studied in
the Chicago Laboratory on several occasions consists in
allowing a blind-folded subject to trace with a short pencil
patterns of various designs cut out of wood or brass. For
one form of test he traces the pattern with the intent to
determine its shape, which he is then asked to draw on a
sheet of paper, his eyes still being closed. In other forms he
is asked to find and draw perfectly the true pathway of a
maze. The maze may then be rotated through 45,° 90° or
270° and the effects noticed. Or the patterns may be traced
over the head or with the hands behind the back. Intro-
spections are required in every case.

These tracery tests are frequently much more successful
in eliciting visual imagery than kinaesthetic, but with cer-
tain persons they bring out the kinaesthetic element with
great distinctness.

Another form of motor imagery, or at least another way
of using it, must be recognized and provided for in any ade-
quate tests although it is introspectively very difficult to
identify and only practiced observers are likely to detect it.
I refer to what Professor Colvin, in a recent very interesting
paper[1] has termed 'mimetic imagery,' not altogether hap-
pily, as it seems to me, for although the imagery is sometimes
mimetic, and possibly it was all originally so, it is often purely
symbolic. I should prefer the term 'symbolic motor imag-
ery.' I have had one marked case under observation more or
less directly for upwards of twenty years. The matter is
referred to in my *Psychology*, but in a manner which might
well escape attention.

[1] *Psych. Bull.*, vol. vi., 1910, p. 223.

Individuals using this form of imagery carry on their reflective thinking in terms of schematic and often faint motor imagery which they use in as purely symbolic a way as that in which verbal thinkers use their words.

No special tests have been devised so far as I know to bring out this type of imagery, and I have none to suggest. If the subject of the test is introspectively acute, almost any of the tests heretofore described will serve the purpose, particularly perhaps those dealing with the solution of simple problems. If this introspective alertness is wanting, I know of no device which will materially assist. None of the ordinary distraction and help tests promises to be of any substantial assistance, because the imagery employed carries a symbolic import often entirely remote from any peripheral factor that can be introduced. It is accordingly all but impossible to impede the flow of the imagery except by such distractions as obliterate the attention process itself; and aids are equally impracticable because one cannot anticipate and so 'hitch on to' the reigning symbolism.

*M. Cutaneous Imagery.* The main fields for the collection of evidence dealing with this group of images, apart from the questionary, have been the tests on sensory discrimination, as in the experiments on touch and temperature, those on localization and a few on memory. They seem to indicate in general, that imagery of the several cutaneous varieties can be detected by most persons under favorable conditions, that with not a few individuals it is very vivid, that it is subject to confusion of a very baffling kind with the corresponding sensations, and finally, that many people make little or no use of it if other imagery can function for it, as is generally the case, particularly where, as in localization processes, visual images are available.

*N. Organic Imagery.* Introspection, which is practically the only mode of approach to organic images, is peculiarly difficult for two reasons: (1) We have little practical necessity to attend to imagery of this variety and we may well suppose that lack of practice is as significant here as it is elsewhere. (2) Much more important, however, is the well

recognized fact that when we attempt to call up an organic image we are likely to stimulate reflexly motor and circulatory changes in the part of the body thought of, and forthwith, we have sensations blending with or displacing our images. So much is this the case, that not a few trustworthy psychologists deny the reality of organic images just as they have denied the existence of kinaesthetic images.

*O. Tests on Gustatory and Olfactory Imagery.* Little or nothing has been successfully achieved here to work out a special technique, and our information rests again mainly on (1) ordinary introspection, (2) sensory discrimination tests, and (3) memory tests. The practical difficulties are, if possible, even more serious than in the case of the previous group. In the gustatory process, introspection is handicapped because the thought of tastes almost inevitably excites reflexly the glands and muscles of the mouth, so that sensations are at once aroused to fuse with or displace the image, assuming that the image is really available. Similarly, in smell, there are reflex excitations of the inspiratory muscles leading to illusory sensations.

The most trustworthy experimental work on the memory of these sensations indicates what theoretically, and on the basis of introspection has generally been held, *i.e.*, that very few persons can readily get genuine olfactory images. This fact is a trifle paradoxical in view of the extreme ease with which olfactory illusions can be produced, a circumstance which would apparently indicate a high degree of central olfactory excitability. Taste-imagery is perhaps more frequent, but the images are probably fusions of several distinct properties as are the corresponding sensations, and they rarely occur free of combination with the sensory activities for the reasons indicated above.

## PART II. RECOMMENDATIONS.

After this brief resumé of available methods we may proceed to present certain recommendations for those who propose to use imagery tests. But first it must be made clear

what conditions such recommendations are designed to meet.

There are several distinct aims which have been more or less well differentiated in the previous investigations of imagery.

Tests have been sought by psychologists which might be used with groups of persons in an objective way and especially has this been true of the search for methods to use with children in the schools. Other investigators have been mainly interested in the intensive analysis of individuals and especially in the search for 'types.' They have sought (1) to determine all the forms of imagery which an individual can command at will. They have sought (2) to ascertain the forms actually employed in the common mental processes of daily life. This latter undertaking may be so executed as to determine not merely *what* imagery is thus used, but also *how* it is used as regards the distribution of the several varieties of it among different kinds of mental activities. In more detailed execution of both these two main aims, attempts have generally been made to grade the imagery discovered with reference to one or more of its ordinarily recognized characteristics like intensity, stability, profusion, accuracy, ease of attainment and the like.

Investigations have also been made with a view to determining the function of imagery. In this connection we meet the experimental literature of imageless thought.

No one of these aims can be realized by any single quickly executed test and one must squarely face the alternative of abandoning any pretense of accurate information concerning the imagery of a given individual, or else of undertaking a rather extended group of tests, whose outcome may be finally problematic if the subject proves too deficient in introspective powers.

One will be well advised who enters upon such analyses without expecting to come at once upon any of the conventionally accepted 'types.' Certainly the visual and auditory types of Charcot and Galton will rarely be recognized in any adequate tests, and even the more recent distinction of word-types and object-types will be found extremely elusive in

many cases. The shifting and substituting of one form of imagery for another under slight changes of conditions, where no one could possibly have suspected the occurrence of such a transformation, speedily render the unprejudiced observer skeptical of all rigid divisions into types of the familiar kind. This is not to deny the reality of types, but simple to urge that they do not follow with any great regularity the lines heretofore laid down. They represent problems we still have to solve, rather than solid foundations on which we can build. A subject who uses visual-object imagery almost wholly in one class of tests may be equally wedded to auditory-motor word imagery in another type. My own observations have abundantly convinced me of this, and confirmation exists in plenty in the extant literature of the subject. No doubt it often occurs that imagery of some one sensory mode dominates markedly over others, but the predominance is probably in many cases much less than has been generally assumed since the work of Charcot and Galton, and it is certainly much more difficult to establish the facts than has been commonly recognized.

As far as possible it seems desirable to make a considerable number of the tests used comparable in character with ordinary mental occupations, so that the results may have a closer bearing on normal conditions. This may be urged while still recognizing that such tests need to be supplemented by others designed to sound possibilities rather than merely to determine norms. We must admit the possibility that even our tests intended to simulate ordinary tasks and familiar situations may be abortive simply because they are known to be experiments on which introspection will subsequently be demanded. The failure of all purely objective tests robs us of any power wholly to escape this possible difficulty. No doubt this element enters in to disturb certain results. We have mentioned above cases where the knowledge that introspections were to be asked has radically transformed the character of the imagery. However, it has not been because the subjects were attempting to take at one and the same time two contradictory attitudes, one of absorbed learning and one

of watching the learning, but simply because the attitude of memorizing with the expectation of being subsequently tested proves to be different from the attitude of concentrated attention without thought of future recall. The difficulty must be recognized and as far as possible overcome. It is in no way fatal and is one familiar to every experimentalist.

We shall recommend two groups of tests containing in part identical members. One of these will be designed to furnish a *brief* survey of the imagery capacities of the subject and the other will offer a more *intensive* and more *accurate* analysis. The tests selected are believed to possess points of superiority in each case justifying their selection in preference to other available ones.

On the negative side it is to be clearly understood that in this report we are not primarily engaged upon the problem of the function of imagery. It is also to be understood that the defects of all exclusively objective methods debar us from recommending any series of tests as conspicuously and unambiguously suited to testing groups of adults or children.

For the briefer series we recommend as described below the questionary, the test on melodies, the 'description' test, the spelling test, the reasoning test and the test on writing. This provides for a cursory examination at least of voluntary imagination, of auditory and visual objective memory, of auditory, motor and visual verbal memory, of reasoning and of one form of voluntary control. Save for the questionary, such a group of tests skillfully conducted need not occupy over two hours, and the questionary can be written out at leisure. The order in which the tests is given is not essential, but the order above represented is advantageous as it places the reasoning test and the writing test last. These are apt to be rather more difficult than the others to secure introspections from.

The recommended tests are numbered as follows: Part II— I-$A$, II-$D$ ($1$), II-$E$, II-$G$, III-$M$, IV-$N$.

For the more extended series we recommend, in the form described below, the questionary as a test on voluntary imagination; an association word test and a test on verbal

similarities as testing both spontaneous and voluntary verbal imagery. If one wishes for the sake of completeness to examine spontaneous imagery of objects, this can best be done by taking care to include in the questionary used a series of inquiries about the train of images in revery and when falling asleep. We regard this as unnecessary, because the essential facts are almost certain to appear in connection with other tests to be mentioned, especially the descriptive tests.

In the range of memory we recommend for the survey of auditory imagery a test on single tones, one on melody and harmony, and a test on noise. The tests on words occur in connection with a later group. For the analysis primarily of visual memories the test on the description of objects and the description of pictures, colored and in black and white, are recommended. These will care for objective-visual imagery. The letter-square test and the spelling will bring out the verbal material and the first may be used for forms and numerals also. The test on reading should be added. It belongs in part to the analysis of perception, in part to volition and in part to memory. It sometimes produces little or no material other than verbal; sometimes it succeeds in calling forth profuse objective imagery. In any case it represents a highly important and familiar mental operation which should not be omitted from examination.

Kinaesthetic memory imagery (?) is certain to be elicited from many subjects by the tests on melody, on words and on the specific voluntary acts like writing. It can be further examined in the tests on estimating extent of movement, and on following tracery patterns.

Cutaneous memory imagery may be tested by experiments on localization.

No special tests on taste are recommended, because the results of all those available are too ambiguous. The questionary replies represent about as high an order of accuracy as we can secure at this point.

A single test on olfactory memory is recommended more by way of completeness than because its outcome is likely

to be very important. This is in substance Miss Gamble's test on reconstruction.

On reasoning processes we recommend problems in arithmetic and geometry as described and at least one problem of a social or ethical character, and one of a purely practical character.

Tests on voluntary control may be carried out to great advantage in any learning process and an extensive literature is at hand for guidance in such tests. We recommend only the tests on writing, (1) because the consumption of time in the other tests is prohibitive for most persons and (2) because, with the exception of the cases where a new and unused muscle group is coming under control, we believe the writing control is essentially typical of all the ordinary motor activities under conscious direction.

Such a group of tests as this intelligently applied, with the minor variations which may be required by the peculiarities of a given subject, will result not only in affording a highly satisfactory inventory of the individual's equipment of imagery, it will also convey a very significant impression of the actual distribution of his imagery and of the manner in which he uses it.

We pass now to a more explicit and detailed account of the tests selected, together with a brief commentary upon their characteristics.

In order to make sure that the tests chosen cover the ground fairly each of the main forms of mental activity involving imagery is herewith examined. For practical purposes this may be done by examining the play of imagination under various conditions, (this includes the process of apprehending meaning as in listening to reading or talking); the use of memory, both mediate and immediate; the process of reasoning and the execution of acts of voluntary control. The tests recommended are grouped on this basis, although other principles of classification might be equally useful. Apology may be made once and for all for the inevitable repetition in the next section of points touched on in the first part of this report.

## DIVISION I.

### *Imagination.*

*A. Questionary.* Although as commonly employed, the questionary involves a measure of distinctively memory material, as a whole the method may be regarded as one appealing primarily to the voluntary use of imagination. It is, moreover, the best available general test on this mental process and may with rare exceptions be advantageously used.[1]

Several questionaries are available of which Galton's, Titchener's, Seashore's and Betts' may be mentioned.[2] These may serve as patterns. The last three possess advantages over Galton's, but his has been so frequently used that there is a good deal of interest in comparing one's results gained by means of it with those previously reported. The main considerations to be met by such a document are clarity of questions, recognition of every genus of images and each important species, and care not to over-emphasize the commoner varieties such as visual and auditory; suggestions for evaluating on some arbitrary scale differences in such characteristics as vividness, accuracy, detail, permanency, and ease of control of the various images. The length of the document and the refinement of questions into which it goes are matters to be decided by the amount of time available.

*B. Association Test.* Certain of the tests on association are fairly to be regarded as tests of imagination and one of these we recommend because of its simplicity and ease of application and also because, as contrasted with the questionary, it brings out spontaneous in distinction from voluntary imagery. It involves on the whole a more natural type of mental act than the questionary. With some subjects the test is likely to prove merely a test of verbal imagination, but even so it is worth making.

Two lists of words should be prepared, the number to be

---

[1] It will be remembered perhaps that we have rejected such tests as the ink-blot test, which undoubtedly have an interesting bearing on certain forms of constructive capacity, because they do not bring out imagery with sufficient certainty.

[2] Galton, *Inquiries into Human Faculty*, Appendix E; Titchener, *Experimental Psychology;* Student's *Manual*, Qualitative, 198; Seashore, *Elementary Experiments in Psychology*, 108; Betts, *Distribution and Function of Mental Imagery*, p. 20.

determined by the time available (twenty has proved a convenient number in my own tests and those of others), but not less in number than will allow some appeal to each important type of imagery.[1] One of these lists is to be read visually, a word at a time, by the subject, the other is to be read to him orally. After each word he is to write down the idea or word suggested by the stimulus word, describing the imagery in which the associates are presented.

Sometimes the test is made by confining the reaction to this associated idea, sometimes the associative train is allowed to run on for a period—five or ten seconds or more. The series is at once recorded and introspections are given. Other complications such as distraction can be introduced. With unpracticed subjects the briefer procedure is preferable. With practiced introspectionists the results of the longer trains are more suggestive, although difficult to tabulate or quantify satisfactorily.

*C. Sensory Similarity.* We recommend a test which lies on the border between imagination and memory involving in part spontaneous and in part voluntary reactions. This is the test on sensory similarity. The subject is given a word *e.g.*, 'cat' and told to write all the words he can which rhyme with it. He is then instructed to write all the words he can think of ending in a given series of letters, *e.g.*, 'ous' without regard to their pronunciation. A limited time, five minutes perhaps, is set, and introspections are asked. Several experiments may profitably be made with different word-stimuli. The test is even more purely verbal than the previous one, but in practice the rhyming form has been found to be peculiarly useful in throwing into the foreground of consciousness auditory and motor speech imagery, not so readily observed in the preceding tests. The other form appeals strongly to visual imagery. Like all these tests it can be complicated indefinitely by distractions and other devices.

---

[1] The experiences in the Chicago Laboratory would not indicate that any particular words chosen possessed permanent merit over others. The list published by Secor (*Amer. Journ. of Psych.*, vol. xi, p. 225 ff.) may serve as an example, although it tends to slight possible taste and smell imagery. Such words as sky, whistle, violets, sugar, ice may illustrate the materials.

DIVISION II.

*Memory.*

To be in any way complete as an account of the use made of imagery in memory processes, involves tests covering the various sense departments recognizing the difference between mediate and immediate memory, and the differences dependent on the condition during learning, *e.g.*, whether the subject knows he is to be examined on his knowledge. The tests will be described without invariable reference to these qualifications, but it must be clearly understood that they concern factors which distinctly affect imagery. They must be taken into account in any complete series of tests.

*D. Auditory Tests,—Tone.* Two types of tests on tone imagery are available, one involving single tones, or groups of tones experienced simultaneously, the other involving melody, thus bringing in rhythm and if desired, harmony. The second can be complicated indefinitely by adding harmonization in terms of a single instrument like the piano, or of many instruments as in an orchestra. The phonograph can be used for this purpose. With persons richly endowed with auditory imagery the more complex tests are sure to be worth while in disclosing intricacies in the organization of imagery not otherwise revealed.

(1) The test can be made with any musical instruments but a familiar instrument like the piano is preferable, as strangeness of tone quality is a distraction to untrained observers, and even with trained observers, the familiar tones are more easily and confidently reproduced. If brevity of time must be consulted, the test on single tones may better be omitted than that on melody. The writer has never chanced to find a case where auditory imagery was gotten with one test and not with the other; but the relative vividness of the imagery occasioned by the two tests varies with different individuals, so that on a scale to show this feature the tests would give contradictory results. The melody test simulates common experience more closely than the other and is easier for untrained subjects to perform.

The subject is told that the first few notes of a familiar melody will be played, enough so that he can recognize it. He is then to finish the melody mentally—'in his head.' Then he is asked to describe the experience. Tone quality, pitch and tempo should be noted and the melody may then be actually played through again to see how it compares with the subject's recollection of it. If possible, the melodies should be played at three points on the tone scale, namely in the deep bass, in the high treble and in the middle register.

The test has two disadvantages, (1) with persons possessing weak auditory, but strong vocal kinaesthetic imagery, it tends to the submerging of the auditory in the kinaesthetic. (2) If the melody has familiar words set to it, persons of strong verbal imagery may again tend to swamp the auditory factors with enunciatory elements. The chief merit of the test is that it 'works itself' almost without instructions, provided the subject has any musical imagery at all.

(2) The test of single tones is conducted as follows: A series of at least five tones is chosen forming a chromatic series, no successive notes to be more than a half tone apart. The piano is again the best instrument, but the tuning fork or organ pipes of the ordinary laboratory equipment will serve.

A tone is struck by the experimenter, the subject then counts silently 'by threes' for five seconds (to break up the carrying over of the after-sensations, or the suppressed humming of the tone). Then he is asked if he can hear the tone mentally. If he replies 'yes,' the operator strikes one of the closely related tones and asks, 'Is it this;' striking several notes, one after the other, including the original tone. This procedure not only stimulates a subject to try to form a correct image, but it also assists him to observe peculiarities in the image, whether it is relatively faint or vivid, whether it is flat or sharp as compared with the original, whether the timbre is correct, or is, as often happens, the tone quality of the subject's own voice.

(3) Noise. Tests on the auditory imagery for noise have less practical importance and for diagnostic purposes

are only essential provided the questionary has given nega-
tive results, although if one desires to ascertain with some
approach to completeness the distribution of imagery forms,
such a test as follows would be needed.

A test may be made following the lines of that on tones.
Homely devices may be pressed into service.   A large empty
box, a small cigar box, a table top, a chair back, and a wooden
rod, may serve as equipment.   Other similar utensils will
suggest themselves.   The first four articles are to be smartly
struck one at a time with the rod; the subject is to count
again as in the tone test and then try whether he can get
an image of the noise.   The experimenter may also crush in
his hand a stiff piece of paper using a single quick motion
for the purpose.   He may also make a sharp hissing sound
and test the subject's imagery as above for the two kinds of
stimuli.

(4) Numerals and Words.   Special tests on the memory of
numerals and words from auditory stimulation are not indis-
pensable in view of the tests to follow, provided one is seeking
simply for the existence of imagery rather than its distribu-
tion.   But if for any reason one feels the need of tests iso-
lating these factors more completely, the following is recom-
mended:

The old dictation test used so often for investigating the
memory span.   Lists of digits are prepared and read dis-
tinctly to the subject in groups of five, six, seven or more at
a time, thus—7, 4, 3, 6, 9, 8,—5, 2, 7, 4, 9, 1, 3.   The subject
is at once to write down or pronounce the numbers as remem-
bered.   The writing is likely to emphasize visual imagery in
persons who can use it freely, especially if done with closed
eyes.   The pronouncing is likely to emphasize the auditory
vocal elements.   The lists may be read partly without rhythm,
and partly with it.   The recall will generally reflect the
effects of this variation. An interruption of a few seconds
between the stimulus and the reproduction, as in the test on
tone, will naturally decrease accuracy of memory, but it is
quite likely to assist the introspective identification of the
imagery.

Nonsense syllables and lists of disconnected words may be similarly dictated. Mediate memory for such materials may obviously be tested by learning the lists by heart and reproducing after an interval of time.

*E. Visual Description Tests.* Of the visual tests we recommend the following which in practice elicit many forms of imagery other than visual, but which are particularly successful in stimulating the latter. They belong to the general class of description tests.

A test for objective visual memory may profitably be used which is based on an old game. In a tray or on a table are placed at least ten or fifteen small and familiar objects—coin, thimble penknife, ring, and the like. Some of these should be distinctly colored, *e.g.*, a red pen-wiper, a blue ribbon and a yellow pencil. The subject is allowed to inspect the articles for a specified time—about one second of time for each article has been found judicious—and he is thereupon required to write down or name all the articles he can remember. His introspections are then taken, and questions are asked about the color, size and relative location of the objects. An excellent test proceeding in a similar manner and resembling the well-known 'Aussage' tests, involves the use of drawings, part of which should be colored and part in black and white.[1] A set of ten, five of each, will serve the purpose. These may be used either as a test for immediate recall or for recall after a few minutes, hours or days.[2] In either case the subject is asked to describe the picture with detail, as in the preceding tests and then tell how he executes the task, grading his imagery if intensive classification is at stake.

[1] Dr. Fernald has used successfully prints of the picture postal card size. She has found the small size which enables the eye easily to get the whole scene at a glance preferable to larger pictures where attention is scattered more widely over the field.

[2] Although the imagery changes in the most amazing fashion as is well known, the author has found no reason to think that commonly there is a change in memory depending primarily on time, from one sense quality such as vision to another such as audition. One may, under pressure of experimental direction, use in immediate memory a form of imagery not ordinarily employed, *e.g.*, tactual; but if the immediate recall would normally be in visual terms, the recall at a later period is not likely to be other than visual. Other investigators have confirmed this. Cf. Segal, *Arch. f. d. ges. Psychol.* 1908, vol., xii, p. 124.

Questions containing erroneous suggestions as to what he has seen bring out most interesting results and should be followed by showing the picture a second time.

Where the test is used merely to ascertain the presence of imagery, no further precautions are necessary, and it may be safely laid down that persons who secure no visual imagery from such a test are, practically speaking, lacking in it altogether. If the tests are used as a method of ascertaining the distribution of imagery in normal processes, a sharp distinction must be drawn between the cases in which the subjects know that they are to be held for a report and those in which they are unaware of this. A relatively naïf inspection of the card has been attained by Dr. Fernald through the use of 'puzzle pictures,' which the subject inspects merely to find the concealed objects. This gains a good degree of attention without throwing the subject into the learning attitude. But the same result can be gained in many other ways. The pictures may, for instance be shown with a view to passing on their relative aesthetic excellence.

*F. Letter Square Test.* Blanks are prepared with nine squares arranged to form a larger square. (My own have been 1 inch on the side, making a large square 3 inches on a side.) In these squares can be placed numbers, pictures, forms, letters, syllables, or words. The subject reads through the list three times and then attempts to reproduce it by writing in the spaces of a blank form, or he may reproduce the materials verbally. Introspections are then asked. The test may be usefully and easily complicated in several ways; by reading (when the material permits) the lists aloud to the subject after which he is to put the remembered elements in appropriate order on a blank square. Distractions, auditory, vocal-motor, or visual, may be introduced as heretofore described. Rhythm may be introduced, the subject may be instructed to think of the material in groups and so on.

This test is one of the best in the list for bringing out verbal imagery, both visual and auditory-motor. It also brings out certain forms of the objective-visual, but it is on the whole

much less satisfactory for producing objective imagery than the description tests. If the auditory dictation is employed, it would rarely be worth while to use the previous test on numerals and words, number II-D4. Both overlap in some degree the following test on spelling, a test which ought in no event to be omitted.

G. *Spelling Tests.* In part of the tests the subject should write the word to be spelled, in part pronounce the letters. In spelling forward, words should be selected which will put some tax on the subject's capacity—shorter and more familiar words can be used with children, longer and more uncommon words being selected for adults. With the common and well known words the spelling will often be found so highly automatized as to involve little or no imagery.

Spelling backwards is even more useful than spelling forward. The test can be made in two ways: (1) The subject is given a word, *e.g.,* 'locality,' and then pronounces or writes— better the first—the letters of the word backward. (2) The experimenter pronounces the letters backward and the subject attempts to determine the word and pronounce it himself. The control of the rate at which the letters are read will materially effect the success of this form of the test. Different subjects vary widely in their response to the test and the rate must be adjusted to their individual capacity. It should be rapid enough to keep the attention strained, but not so rapid as to produce a serious confusion. For most subjects the rate of one letter per second is a good one to begin with. The subject gives his introspection at the end of each test. The test can be complicated with the methods of disturbance and help.

H. *Reading Tests.* Reading affords an opportunity for various forms of tests. They should be included in any series designed to afford a survey of the imagery in normal types of mental action; but they present introspective difficulties which often render them of secondary value for use with untrained observers and especially children.

The passages selected may be determined in part by local conditions. As a rule they should not exceed ten or fifteen

lines in length and may well be shorter. Both prose and poetry should be represented, and among the poems should be representatives of well-known songs (to stimulate auditory-vocal imagery). They may advantageously include (1) passages descriptive of nature in her appeals to the various senses, (2) passages describing some concrete practical operation, *e.g.*, the planting of trees, the making of brick, and the like, (3) passages discussing some distinctly abstract ideas, whether of a philosophical or a scientific character.

The material should be presented (1) visually and (2) orally. It should be given (a) under conditions where the subject knows that he will *not* be examined on the contents of the passages. He is simply asked to read or listen as he would in the case of any matter momentarily attracting his interest: (b) contrasted with this procedure, the test should be made when the subject knows he is to be asked to give the import of the passage. This he may either write or give orally, the former generally proving in practice the more useful. He may be asked for the general import of the passages or for as much as he can recall verbatim. He may react at once or after an interval of time. His introspections are asked both on the process of apprehension and of reproduction. The passages read for introspection on the process of apprehension are not to be used in the tests on retention. Various complications will suggest themselves.

*I. Cutaneous Tests.* For contact stimulations the subject (his eyes being closed or averted) may be touched lightly on the forearm with a blunt cork point. He then counts aloud by 'threes,' for ten seconds, to allow any after-sensations to subside. He then attempts to recall the sensation and points at the spot stimulated. His introspections are then requested.

With blunt metal rods a similar test may be made for warmth and cold, the rods being kept in hot or cold water and carefully wiped before being applied to the skin. The points should feel to the subject's skin distinctly cool and distinctly warm, but not smarting nor painful when left on a few seconds. The after-sensation must be carefully avoided.

The stimulations should in no case exceed a second in duration. The counting should go on for at least ten seconds, and twenty is better. The subject is then invited to recall the sensation if possible and point to the spot where it was located.

Pain may be tested by a light needle prick or by snapping a rubber band on the back of the hand. A considerable interval of distraction must be used, the length depending on the violence of the stimulus. The experimenter must determine himself by control tests how long to allow for the subsidence of the after-sensation. Otherwise the procedure is as with temperature sensations.

These tests fall out very differently with different persons, some subjects failing to get the images with any certainty, others getting them easily, and still others being flooded with visual imagery of the parts stimulated. But it may be fairly assumed that a subject who under these conditions gets no trace of cutaneous imagery, never makes any important use of it.

Interesting experiments involving imagery from fusions of touch and motor elements may be made by using the raised letters of the alphabet for the blind. The subject is aside from our present main purpose and need not be discussed.

*J. Kinaesthetic Tests.* Cut out of wood, metal or cardboard designs into which a pencil point will just easily fit. The designs may be geometrical figures, *e.g.*, triangles, squares, oblong figures, circles, ellipses, etc., or they may be continuous line letters like 'S,' 'Z,' 'O,' 'M,' or 'V.' We have used in the Chicago Laboratory a figure like the maze of the comparative psychologists, but other purposes were primarily in view and for untrained observers simpler figures are better.

With eyes closed or screened the subject traces the figures with a pencil in the slot, the operator putting the pencil at the start at one or other extremity of the figure. The tracing may be done one or more times depending on the complexity of the figure. Then the subject is given a blank sheet of paper in place of the pattern and on this he is to report the design as he remembers it. His introspections are then asked. Half a dozen tests with different designs are sufficient.

An interesting variation of the test, often resulting, as does the original, in throwing more light on visual than on kinaesthetic imagery, consists in having the subject attempt to reproduce the figure as it would appear if turned about through 45°, 90°, or 270°. We have gotten many striking results, too, from attempting either to interpret or to reproduce figures traced in unusual positions, *e.g.*, behind the back or above the head.

A simpler test, but on the whole less successful, is the familiar estimation of linear magnitudes. The subject sits facing a table on which is a horizontal wooden or metal rod supported by clamps attached to upright rods. Two stops are attached to the horizontal bar at a convenient distance apart, say 25 centimeters. The index finger of either hand measures off the distance between the stops by sliding along the bar. The second stop is then removed and the subject attempts to measure off the original distance. . The eyes must be closed or averted. If short distances are used, the finger may move freely in the air between stops which must be higher for this purpose. The object of this procedure is to eliminate the contact sensations. Introspections are then recorded.

*K. Taste and Smell.* In the case of olfactory and gustatory imagery the author is disposed to recommend going beyond the questionary only when the results from that are negative or highly ambiguous, or when a complete census on distribution is desired. The difficulty of distinguishing between sensations and images is similar to the difficulty with organic imagery and perhaps quite as serious. We have however in the case of smell, at least, available tests which may assist by stimulating introspection.

One of Miss Gamble's tests may serve to meet the olfactory case. A series of ten odorous substances is prepared in shallow wide-mouthed bottles, with glass stoppers.[1] The subject sitting blindfolded takes the bottles in the order

---

[1] The following list is easily secured and will serve the purpose. The material in any psychological laboratory will afford other equally good groups. Anise, benzoin, bergamot, cassia, cloves, camphor, creosote, musk, peppermint, sandalwood.

in which they are presented, removes the cork, inhales once and replacing the stopper, turns to the next bottle. After the series has been twice thus presented, the subject is given the bottles in disarranged order and asked to arrange them as first experienced. The test can be repeated several times with the same materials by changing the order of the stimuli.

If olfactory imagery is available, it is reasonably certain to appear under conditions such as these. The best statistics make it seem probable that such imagery is relatively rare. This conclusion is at variance with the outcome of many questionary investigations. I have frequently, when using the questionary, received affirmative answers, as have other investigators, to the inquiries about such images. I am disposed to attribute the discrepancy to a failure on the part of more naïf observers to discriminate between pure olfactory images and mental states which are either illusory sensations or fusions of peripheral with centrally aroused factors. The more careful experimentation tends to render subjects more sensitive to this distinction and more capable of avoiding the possible confusion involved.

No experimental technique seems to the author successful in so controlling the glandular and muscular activities of the mouth as to make certain that sensations and images of taste are kept apart. The ordinary laboratory forms of stimulation in connection with sensory discrimination, for example, aggravate rather than help this difficulty. At present, therefore, the writer is not prepared to urge any tests applicable in a general way as being materially better than the results from the questionary, uncertain and unsatisfactory as these are.

*L. Organic Tests.* Owing to the serious difficulty of distinguishing organic sensation from organic imagery it does not seem profitable at the present time to go beyond the inquiry of the questionary, or the possibilities of ordinary introspections.

DIVISION III.

*M. Reasoning Tests.* A few tests on reasoning processes
ought to be included in any diagnostic series, not only for
completeness in a program aimed at determining the dis-
tribution of imagery, but also because they serve to bring
out under quite natural conditions forms of imagery essen-
tial to the thought processes of the subject.

Among the most practicable are simple problems in arith-
metical computation. For instance, the experimenter may
dictate two-place numbers, to be added or subtracted men-
tally. Three, four and five place numbers may be written
down and the arithmetical operations carried on while look-
ing at the paper, but without the assistance of pen or pencil.
Introspections are asked for after the subject completes and
announces the result of his work.[1]

To test the geometrical processes the problem of the cube
mentioned on page 76 may be used or one of this kind: The
subject is told to think of a square; if a line is made joining
a pair of diagonally opposite corners, what shapes have the
resulting figures? If the other two corners are similarly
joined, what shapes have the resulting figures and how many
are there? Are they alike in shape and size? 'How do you
know?'

Simple problems may be selected from any range of prac-
tical, political, or scientific interest, the only essential con-
ditions are that the question shall be lucid and the answer
within the capacity of the subject to reach. For this purpose
a problem so difficult as simply to baffle attack is futile.

Something of this kind may serve as a practical problem:
What route would you take and how would you go about
making arrangements, if you were to travel from your home
to Siberia? Give the steps in as much detail as possible.
Many other more immediately personal problems will sug-
gest themselves and may profitably be used.

Problems dealing with concrete ethical situations may

---

[1] Dr. Fernald has tried having the computation carried on in terms of the Roman
alphabet numerals, but without eliciting any forms of imagery not otherwise accessible.

bring out interesting results, although the introspection required is rather too difficult for untrained subjects. Such questions may be illustrated by the old ethical problem whether it is ever right to lie. In the case of a dying man who cannot possibly recover, is it not justifiable to lessen the pain of his final hours by concealing from him the seriousness of his condition? Questions of this type and problems dealing with relatively abstract relations, may bring out interesting evidence concerning the conceptual imagery employed. It may be mentioned in this connection that the recent studies of the process of judging, afford a large range of tests in which the analysis of imagery may be carried out.[1]

The experimenter ought to be especially alive in tests of this general variety to the presence of symbolic motor-imagery, and enough non-mathematical problems should be given to make sure whether or not one has this kind of imagery in use. We bar mathematical problems because vocal-motor, or hand-motor imagery referring directly to numbers, is likely to be used in such cases.

## DIVISION IV.

*N. Voluntary Control Tests.* Writing affords a very easy and effective means of access to imagery as employed in the control of voluntary movements. Like spelling it is in part a memory test, but in the forms to be mentioned here it is primarily a test on imagery in its relations to voluntary action, and it furnishes, together with the tests on spelling and reading, a transition from experiments on mere memory to those concerned with the analysis of voluntary action. [Spelling and reading have been treated under memory.]

The subject may be asked to write either from oral dictation or from the printed copy. His introspection may then be obtained and the results compared with a repetition of the test in which fresh material is used. A paragraph containing a hundred words will serve the purpose. It is well to select one in which the vocabulary is not too familiar and common-

[1] Cf. Titchener, *Experimental Psychology of Thought Process*, New York, 1909.

place. Passages from text-books on philosophy and psychology fill the bill admirably for ordinary subjects. In at least one set of tests the material written should be something known by heart, a verse perhaps. The writing begins at a given signal and on completion introspections are at once asked.

Writing the letters of the words backwards [the words being spelled either forward or backward] from oral dictation, writing every other letter of the word, writing left-handed, writing upside-down, and looking-glass writing, all are excellent devices for throwing imagery into the foreground, and all but the two last are done with sufficient readiness by most subjects. Full introspections are in every case required.

The entire series selected may then be repeated with the eyes closed. This is often a strikingly useful variation of the test.

We have found it possible to introduce the distraction methods quite successfully into the test. For auditory distraction one may listen to a metronome, for motor distraction one may repeat some well-known verse, either aloud or under the breath, or one may count rapidly. For visual distraction one may let the eyes follow the contour of some conventional pattern, or trace the lines of overlapping spirals. (The difficulties in securing satisfactory distractions have already been mentioned.) The procedure in the case of auditory-motor distraction may suffice to make clear the method. The subject with eyes closed is told upon receiving a signal to begin writing the first verse of 'America;' while he writes, he is to sound the vowel 'a.'

A great many experiments have been conducted in recent years, to determine the psychological processes controlling voluntary action, and especially with a view to determining the part played by imagery. If the tests on writing and spelling are used, especially in the more infrequent forms like writing with the left hand, and like spelling backward, there need be very few additions, at least so far as concerns ascertaining the kinds of imagery which can be commanded. If exhaustive tests on normal distribution are desired, we must

add to the list tests involving new coördinations, or at all events, radically new combinations of old coördinations. Learning (1) to move the ears, to move the scalp over the crown, to move the toes independently of one another, and (2) learning to finger a musical instrument may illustrate the two main types of cases.[1]

Tests of this character require so extended a period of work as to render them rather inappropriate for detailed mention in context with the other tests of this group. Reference is made below to one or two studies on the learning of such acts. An investigation of one form of the process is now nearing completion in the Chicago Laboratory.[2]

## PART III. GRADATION OF IMAGES.

Most of the investigations hitherto conducted have undertaken to offer some quantifying statements with reference to the vividness, stability, ease of attaining the images, and so on. In the opinion of the present writer the gradations which have often been employed in such scaling of the qualities of imagery, are too refined to be successfully employed by the ordinary observer. For instance, writers on imagery who report upon grades of intensity or grades of vividness[3] have often employed six or more distinctions in accordance with which they have asked their subjects to make their reports.

The effort to subject this matter to a satisfactory series of control tests proved impracticable for the following reasons.

[1] The case of learning to move the ears and learning to move the toes independently of one another differ in that one involves gaining control over a wholly unused muscle, the other involves separating muscular innervations which have always occurred together.

[2] Bair, Acquirement of Voluntary Control, *Psychol. Rev.*, vol. viii, 1901,. p. 474. Book, Psychology of Skill, *University of Montana Studies*, vol. i, 1908.

[3] The author does not feel under obligation to enter upon the controversy concerning the genuineness of the intensive attribute as applied to imagery, nor the differentiation of intensity from vividness and of both from clearness. If vividness or intensity are either of them genuine characteristics of the image process, gradation is at least theoretically possible and it becomes proper for a report of this kind to make recommendations concerning it. That ordinary observers are able to make distinctions which seem to them natural and easy, when asked to grade their imagery in this way, affords a strong *prima facie* case for the genuineness of the distinction reported.

If a series of images could be secured on a number of successive occasions, with precisely the same degree of vividness, it would be possible by tabulating the classifications made by the subjects, to ascertain with what degree of nicety they could consistently grade the experiences. But evidently we can secure no such guaranty of the actual similarity in vividness of the imagery on these successive occasions. Any consistency which the subject may evince becomes therefore ambiguous and inconclusive. On the other hand a failure to classify the images in the same way on different occasions is no guaranty at all that the classifications are untrustworthy.

The author's observations extending over a number of years and including in their range a large number of individuals leads him to believe that occasionally subjects are able to grade their images on the scale of eight with entire confidence and with practical consistency. The difficulty which is most likely to be troublesome in the case of such observers is the reduction to a common scale of imagery belonging to two different sense departments like vision and audition. The maximal degree of vividness in the one case may seem very different from the maximal degree in the other, and still the subject may feel indisposed to call the weaker group of images in any sense faint or indistinct. Our statistics do not at present in the author's judgment afford us adequate information as to whether the scale of intensities for the different forms of imagery is of equal extent, much less whether the threshold of noticeable differences in it is any wise homogeneous. Under these circumstances one must regard any estimates of gradation as being merely rough approximations.

For the reasons just discussed, as well as for another highly important reason next to be mentioned, the author has come to believe that three gradations are as much as can be profitably and safely recognized in the presentation of statistics on imagery. The additional reason which has had great weight with the author has been the testimony of several psychologists of distinction and wide experience that they themselves found it entirely impossible to make so many gradations as the questionary commonly demands. The procedure here

recommended would in the case of vividness, for example, recognize three divisions which may be called very vivid, moderately vivid, and faint. Subjects who feel the necessity for intermediate grades can be supplied with these by the use of plus and minus signs. Thus, the gradation faint might be modified to provide two more classes by adding a plus sign for images which are slightly above this level and a minus sign for those which are below. A similar treatment of each of the other main divisions would give nine differenti-ations, which is fully adequate to care for any case that the author has ever encountered.

Even after we have agreed upon the number of gradations to be recognized, the problem still remains in dealing with individual subjects, how we are to instruct them in the use of such standards, which are in the nature of the case rather arbitrary and altogether subjective. On this point the author has no dogmatic advice to offer. In his own experience two types of procedure have seemed practicable. In one, the subject is given no definite advice or suggestions, but is sim-ply advised to wait until he has inspected twenty or thirty images before beginning his gradations. He is, moreover, allowed to revise his gradations if they appear to him incon-sistent with one another. For example, a subject may start out classifying a given image as very intense, only to find toward the end of his observations that another image is noticeably more intense. In this case he naturally wishes to scale down all the other statements which he had made on the basis of his first relatively vivid image. The second pro-cedure involves attempting to give the subject suggestions in which his attention is called to the possibility, for example, that the most vivid image might approach a perceptual exper-ience in its character. It is also suggested that the faintest image would be almost impossible to detect. Suggestions of this sort are sure in the case of some subjects to be helpful, but others are not materially assisted and in a few cases the suggestions have proved confusing rather than otherwise. In the view of the writer these difficulties indicate simply that gradations of the kind mentioned are of very problem-

atic significance when gathered from untrained observers without the possibility of some measure of verification.

Gradations of stability would likewise be made on the basis of three distinctions; first, very stable, approaching perception in fixity; second, moderately stable; third, very fleeting. Here again by a system of plus and minus symbols the demand for a finer classification may be met if it arises.

It need not be argued that stability as a category of mental imagery is primarily applicable to the space mediating sensations. Images of sound are often in the nature of the case momentary and fleeting. In their case revivability would serve the same function in large part which stability serves in the case of the visual imagery. In a similar way olfactory images, when they can be gotten, are likely to be transitory, even though they may be recalled repeatedly.

Almost all subjects are able to report with confidence on the ease or difficulty with which they can command imagery. The gradations used may be 'very easy,' 'moderately easy,' 'difficult.' The plus and minus system affords adequate recognition of the differences met with in practice.

In the case of memory tests a grading of accuracy is instructive and the matter of detail may always be advantageously graded if gradation of any kind is to be attempted.

In the judgment of the author as was intimated a few lines above, the entire problem of gradation may well wait upon a more complete investigation of the qualitative peculiarities of imagery before attempting to define a definite line of procedure. At the present time it seems more than doubtful whether we can secure standards for this purpose so intelligible and so readily applicable as to enable us to secure uniformity among different investigators.

The author undertakes to offer no exhaustive bibliography of the subject. The references which have been cited in the body of the text are here brought together for convenience, and to them are added a number of other titles. The list contains the works which the author has found most directly useful, and which he believes present a just and essentially adequate record of the progress in the study of imagery since Galton's work.

BIBLIOGRAPHY.

ABBOTT, E. A. Memory Consciousness in Orthography, *Psychol. Rev. Monog.*, 1909, vol. xi, no. 44.

BAIR, J. H. Acquirement of Voluntary Control, *Psy. Rev.*, 1901, vol. viii, p. 474.

BALDWIN, J. M. Internal Speech and Song. *Phil. Rev.*, 1893, vol. ii, p. 385.

BALLET, G. *Le Langage Interieure*, Paris, 1886.

BAWDEN, H. H. Study of Lapses, *Psy. Rev. Monog.*, 1900, vol. iii, no. 14.

BELL and MUCKENHOUPT, Methods of Determination of Ideational Types, *Amer. Jour. Psy.*, 1906, vol. xvii, p. 121.

BETTS, G. H. Distribution and Function of Mental Imagery, *Columbia University Contributions to Education*, 1909, no. 26.

BENTLEY, I. M. Memory Image and its Qualitative Fidelity, *Amer. Jour. Psy.*, 1899, vol. xi, p. 1.

BINET, A. *Psychologie des grands Calculateurs et joueurs d'échecs.* Paris, 1894. Description d'un Objet, *L'Année Psych.*, 1896, vol. iii, p. 296. et Henri V.—Memoirs des Mots, *L'Année Psychol.*, 1894, vol. i, p. 1. Memoire des Phrases, *ibid.*, p. 24.

BOOK, Psychology of Skill, *University of Montana Studies*, 1908, vol. i.

BROWN, C. E. Psychology of Arithmetical Processes, *Amer. Jour. Psy.*, 1906, vol. xvii, p. 1.

COHN, J. Exper. Untersuch. u. d. Zusammenwirken d. akust.—mot. und d. visuellen Gedächtnisses, *Zeitsch. f. Psychol.*, 1897, vol. xv, p. 161.

COLVIN, S. S. Methods of Determining Ideational Types, *Psychol. Bull.*, 1909, vol. vi, p. 223.

COLVIN, S. S. and MYERS, E. J. Develop. of Imagination in School Children and Relation between Ideational Type and Retentivity of Material Appealing to Various Sense Departments. *Psychol. Rev. Monog.*, 1909, vol. xi, no. 44.

DEARBORN, W. F. Psychology of Reading, *Columbia Univ. Cont. to Phil.*, 1906, no. 14.

DODGE, R. *Die Motorische Wortvorstellungen*, Halle, 1896.

DUGAS, L. Recherches experiment, sur les differents types d'images, *Rev. Phil.*, 1895, vol. xxxix, p. 285.

EGGER, V. *La parole interieure*, Paris, 1881.

FRACKER, G. C. Transference of Training in Memory, *Psy. Rev. Monog.*, 1908, vol. ix, no. 38.

FRANKL, E. *Ueber Vorstellungselemente und Aufmerksamkeit*, Ausgburg, 1905.

FRASER, A.  Visualization as a Chief Source of the Philosophy of Hobbes, Locke, Berkeley and Hume, *Amer. Jour.*, 1891, vol. iv, p. 230.

FRENCH, F. C.  Mental Imagery of Students, *Psy. Rev.*, 1902, vol. ix, p.

GALTON, F.  *Inquiries into The Human Faculty*, London, 1883.  Statistics of Mental Imagery, *Mind*, 1880, vol. v, p. 301.

GAMBLE.  Memorising—Materials by the Reconstruction Method, *Psy. Rev. Monog.*, 1909, vol. x, no. 43.

GAMBLE AND CALKINS, M. W.  Die Reproducierte Vorstellung beim Wiederkennen und beim Vergleichen, *Zeitsch. f. Psychol.*, 1903, vol. xxxii, p. 177.

HUEY, E. B.  Psychology and Physiology of Reading, *Amer. Jour. Psy.*, 1900, vol. xi, p. 283, *ibid.*, 1901, vol. xii, p. 292.

JASTROW, J.  Eye-Mindedness and Ear-Mindedness, *Pop. Sc. Mo.*, 1888, vol. xxxiii, p. 597.

KELLER, HELEN.  *The World I Live In*.  New York, 1908.

KÉRAVAL, P.  *Le langage écrit*, Paris, 1897.

KRAEPELIN, E.  Der Psychol. Versuch in d. Psychiatrie, *Psychol. Arbeiten*, 1895 vol. i, p. 73.

KRIES, J. V.  Ueber das absolute Gehör, *Zeit. f. Psychol.*, 1892, vol. iii, p. 257.

KUHLMANN, F.  Mental Imagery and Memory of Meaningless Forms. *Psychol. Rev.*, 1906, vol. xiii, p. 316.  Memory Consciousness for Pictures of Familiar Objects, *Amer. Jour. Psy.*, 1907, vol. xviii, p. 389.  Auditory Memory Consciousness, *Amer. Jour. Psy.*, 1909, vol. xx, p. 194.

LAY, W. A.  *Führer durch d. Rechtschreib-Unterricht*, 1904. *Experimentelle Pädagogik*, 1908.

LAY, WILFRID.  Mental Imagery, *Psy. Rev. Monog.*, 1898, vol. ii, no. 7.

LEHMANN, A.  Ueber Wiederkennen, *Phil. Studien*, 1889, vol. v, p. 96.

MAC DOUGAL, R.  Music Imagery, *Psy. Rev.*, 1898, vol. v, p. 463.

MEUMANN, E.  Methoden z. Feststellung d. Vorstellungstypen. *Zeitsch. f. Exper. Pädag.*, vol. iv, p. 23.  *Oekonomie und Technik d. Gedächtnisses*, Leipzig, 1908.  *Vorlesungen z. Einführung in d., Exper. Pädagogik*, Leipzig, 1907, 2 vols.

MOORE, C. S.  Control of the Memory Image, *Psy. Rev. Monog.*, 1903, vol. iv, p. 277.

MURRAY, ELSIE.  Peripheral and Central Factors in Memory Images of Visual Form and Color.  *Amer. Jour. Psy.*, 1906, vol. xvii, p. 227.

NETSCHAJEFF, A. Experim. Untersuchungen u. d. Gedächtniss-Entwickelung bei Schulkindern, *Zeits. f. Psychol.*, 1900, vol. xxiv, p. 321.

PERKY, C. W. Exper. Study of Imagination, *Amer. Jour. Psy.*, 1910, vol. xxi, p. 422.

PFEIFFER, L. *Ueber Vorstellungstypen. Päd. Monographien*, Leipzig, 1907.

PHILIPPE, J. Les transformations d'images mentales, *Rev. Phil.*, 1897, vol. 43, p. 481.

PICK, A. *Sprachverstandniss*, Leipzig, 1909.

SEASHORE, C. E. *Elementary Experiments in Psychology*, New York, 1908.

SECOR, W. B. Visual Reading, *Amer. Jour. Psy.*, 1899, vol. xi, p. 225.

SEGAL, J. Ueber d. Reproducktionstypus und d. Reproduzieren v. Vorstellungen. *Arch. f. d. ges. Psychol.*, 1908, vol. xii, p. 124.

SLAUGHTER, J. W. Behavior of Mental Images, *Amer. Jour. Psy.*, 1902, vol. xiii, p. 526.

STERN, L. W. *Psychologie d. Individuellen Differenzen*, Leipzig, 1900.

STETSON, R. H. Types of Imagination, *Psy. Rev.*, 1896, vol. iii, p. 398.

STRICKER. *Sprachvorstellungen*, Vienna, 1880.

SYBEL, A. VON. Ueber d. Zusammenwirken versch. Sinnesgebiete bei Gedächtnissleistungen, *Zeits. f. Psychol.*, 1909, vol. 53, p. 257.

TITCHENER, E. B. *Exper. Psychology of Thought Process*, New York, 1909.

WHIPPLE, G. M. Mental Image and Judgment in Discrim. of Clangs and Tones, *Amer. Jour. Psy.*, 1901 vol. xii, p. 409. *Ibid.*, vol. xiii, p. 219.

WHITEHEAD, L. Visual and Aural Memory Processes, *Psy. Rev.*, 1896, vol. iii, p. 258.

WOLFE, H. K. Unters. u. d. Tongedächtniss, *Phil. Stud.*, 1886, vol. iii. p. 534.

# APPENDIX

After the foregoing report was in type, Professor L. J. Martin presented before the American Psychological Association, at its annual meeting held at Minneapolis during the last week of December 1910, an interesting account of some experiments on the projection of visual images.

Professor Martin put before her subjects common articles such as books, ink bottles and water glasses. They were then instructed to attempt to form and project a visual image of the object which was to appear just beside the perceptual original. The image and the percept were then compared. A considerable number of her subjects were reported as able to do this satisfactorily.

The tests in the present paper [Pt. II, Div. II E] dealing with the description of objects and pictures resemble Professor Martin's tests in some particulars and proved extremely illuminating, as has already been stated, when a comparison was made at the end of the test between the image and its perceptual original. Such comparison is obviously analogous, though by no means identical, with her procedure in requiring a direct comparison of the image with the object, the two being ostensibly present simultaneously. Her method certainly deserves further study. But it is clearly exposed to certain introspective inaccuracies which would need to be carefully safe-guarded, if the test were employed upon inexperienced observers. Our own results indicated far less capacity of projection in our subjects than in hers.