



UNIVERSITY OF
ILLINOIS PRESS

"General Intelligence," Objectively Determined and Measured

Author(s): C. Spearman

Source: *The American Journal of Psychology*, Vol. 15, No. 2 (Apr., 1904), pp. 201-292

Published by: University of Illinois Press

Stable URL: <http://www.jstor.org/stable/1412107>

Accessed: 26-04-2018 05:54 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://about.jstor.org/terms>



University of Illinois Press is collaborating with JSTOR to digitize, preserve and extend access to *The American Journal of Psychology*

"GENERAL INTELLIGENCE," OBJECTIVELY DETERMINED AND MEASURED.

By C. SPEARMAN.

TABLE OF CONTENTS.

	PAGE
Chap. I. Introductory.	
1. Signs of Weakness in Experimental Psychology	202
2. The Cause of this Weakness	203
3. The Identities of Science	204
4. Scope of the Present Experiments	205
Chap. II. Historical and Critical	
1. History of Previous Researches	206
2. Conclusions to be drawn from these Previous Researches	219
3. Criticism of Prevalent Working Methods	222
Chap. III. Preliminary Investigation	
1. Obviation of the Four Faults Quoted	225
2. Definition of the Correspondence Sought	226
3. Irrelevancies from Practice	227
(a) Pitch	228
(b) Sight	232
(c) Weight	233
(d) Intelligence	233
4. Irrelevancies from Age	233
5. Irrelevancies from Sex	235
6. The Elimination of these Irrelevancies	236
7. Alternations and Equivocalities	238
Chap. IV. Description of the Present Experiments	
1. Choice of Laboratory Psychics	241
2. Instruments	242
(a) Sound	243
(b) Light	244
(c) Weight	245
3. Modes of Procedure	246
(a) Experimental Series I	246
(b) " " II	247
(c) " " III	248
(d) " " IV	249
(e) " " V	249
4. The Estimation of Intelligence	249
5. Procedure in Deducing Results	252
(a) Method of Correlation	252
(b) Elimination of Observational Errors	253
(c) Elimination of Irrelevant Factors	255
Chap. V. The Present Results	
1. Method and Meaning of Demonstration	256
2. Correspondence between the Discrimination and the Intelligence	259

	PAGE
(a) Experimental Series I	259
(b) " " II	263
(c) " " III	264
(d) " " IV	265
(e) Conclusions	268
3. Correspondence between General Discrimination and General Intelligence	268
(a) Village School	269
(b) The High Class School	269
(c) Practical Verification of the Argument	271
(d) Conclusion	272
4. Universal Unity of the Intellectual Function	272
5. The Hierarchy of the Intelligences	274
6. Outer Factors Determining the Amount of Correlation	277
7. Previous Researches Conflicting with the Present Results	279
8. Summary of Conclusions	283
Appendix. Tabular Statements of Original Data and Calculated Correlations	286

CHAPTER I.

INTRODUCTORY.

I. Signs of Weakness in Experimental Psychology.

To-day, it is difficult to realize that only as recently as 1879 Wundt first obtained from the authorities of Leipsic University one little room for the then novel purpose of a "psychological laboratory."

In twenty-four years, not only has this modest beginning expanded into a suite of apartments admirably equipped with elaborate apparatus and thronged with students from the most distant quarters of the globe, but all over Germany and in almost every other civilized country have sprung up a host of similar institutions, each endeavoring to outbid the rest in perfection. The brief space of time has sufficed for Experimental Psychology to become a firmly established science, everywhere drawing to itself the most vigorous energies and keenest intellects.

But in spite of such a brilliant career, strangely enough this new branch of investigation still meets with resolute, wide spread, and even increasing opposition. Nor are its enemies at all confined to belated conservatives or crotchety reactionaries; they are rather to be found among the most youthful schools of thought; their strength may be in some measure estimated from the very elaborate apology which one of the best known experimental psychologists has lately found himself called upon to utter on behalf of his profession.¹

¹ Münsterberg: "Die Grundzüge der Psychologie."

And, indeed, when we without bias consider the whole actual fruit so far gathered from this science—which at the outset seemed to promise an almost unlimited harvest—we can scarcely avoid a feeling of great disappointment. Take for an example Education. This is the line of practical inquiry that more than all others has absorbed the energy and talent of the younger workers and that appears to offer a peculiarly favorable field for such methods. Yet at this moment, notwithstanding all the laborious experiments and profuse literature on the subject, few competent and unprejudiced judges will venture to assert that much unequivocal information of capital importance has hitherto thus come to light. Nor have the results been more tangible in Psychiatry or in any other department of applied psychology.

Those, then, who have the highest opinion concerning the potentialities of this new science, will feel most bound to critically examine it for any points of structural incompleteness.

2. The Cause of this Weakness.

Most of those hostile to Experimental Psychology are in the habit of reproaching its methods with insignificance, and even with triviality. They regard it as an infatuation to pass life in measuring the exact average time required to press a button or in ascertaining the precise distance apart where two simultaneous pinpricks cannot any more be distinguished from one another; they protest that such means can never shed any real light upon the human soul, unlock the eternal antinomy of Free Will, or reveal the inward nature of Time and Space.

Such blame, however, would appear ill founded—at any rate, in principle. This same apparent triviality lies at the base of every successful science. The three laws of Newton on first inspection are by no means remarkably significant; yet by a large number of instructed persons they have been found implicitly to contain the supreme key to every event on the earth below and in the heavens above. When starting any new branch of mathematics, again, most people have had occasion to be astonished at the curious suddenness with which the seemingly shallow beginnings have shelved down into drowning deep water. The general fact is that our limited intellects can only hope to deal with the infinite complexity of Nature after analyzing it down into its bare unæsthetic elements.

On the other hand, it must frankly be admitted that such a procedure is, after all, only indirect; that it does not immediately handle the things which really interest us, but other things which are believed to accurately enough betoken the former; that the results arrived at concerning the simpler terms are therefore always worthless, except in proportion as their

elements have been proved beyond dispute to be identical with those of the more complex terms. Now, even in physical sciences this proof is not such an infallible operation that we can afford to neglect the possibility of lurking errors which may vitiate all our conclusions; and in psychical research such dangers are enormously magnified. When we pass an electric current through water until it vaporizes away into bubbles of hydrogen and oxygen, we can with reasonable precautions be tolerably certain that we have still got in our jars almost the whole of the same material substance, only reduced to simpler forms. But when we assert that the decision of Regulus to vote against making peace with Carthage was no more than a conglomeration of visual, auditory, and tactful sensations in various stages of intensity and association, then there is an undeniable risk that some precious psychical elements may have slipped through our fingers.

On this vital matter, it must reluctantly be confessed that most of Wundt's disciples have failed to carry forward the work in at all the positive spirit of their master. For while the simpler psychoses of the Laboratory have been investigated with great zeal and success, their identification with the more complex psychoses of Life has still continued to be almost exclusively ascertained by the older method of introspection. This pouring of new wine into old bottles has not been to the benefit of either, but rather has created a yawning gulf between the Science and the Reality. The results of all good experimental work will live, but as yet most of them are like hieroglyphics awaiting their deciphering Rosetta stone.

3. *The "Identities" of Science.*

Here, we naturally arrive at the important question as to what actually constitutes "identity" for scientific purposes.

As regards the material atoms of the physical sciences, this relation is of two orders. There is the Identity in the looser use of the word, which really means no more than uniformity of potential function, or the fact of having like reactions under like conditions; this alone constitutes the proper topic of the science. And then there is the true Identity involved in the metaphysical idea of persistence of substance, which in science is only a convenient working hypothesis to aid in establishing uniformities of the former order.

For psychology, also, the identification is of two orders. First, there is once more Uniformity of Function, and again this appears to be the proper topic of the positive science. But the second order is quite disparate from anything in physics, being that of inward resemblance as ascertained by introspection; such a "Conceptual Uniformity," though in metaphysics per-

haps of primary importance, in psychology is but an indispensable substructure—and one of lamentable fallibility. It cannot even be forthwith assumed necessarily to imply complete Functional Uniformity; and it is peculiarly insusceptible of scientific precision, propositions scarcely ever admitting of either decisive confirmation or refutation.

Now, it is one of the great merits of experimental psychology to have largely introduced the direct investigation of these Functional Uniformities, which have the infinite advantage of being eventually susceptible of conclusive proof, and on being securely established are in their turn capable of throwing back a valuable corrective light upon the Conceptual ones also. So far, however, this matter of research seems to have been almost entirely confined to such correspondences as are approximately complete (these, indeed, being the only ones attainable without a new development of methodics). But the vast majority of the functional relations are not thus complete; they are more or less thwarted by other factors; they outwardly present themselves only in the form of stronger or weaker *tendencies*. And precisely of this incomplete nature are most of the Functional Uniformities which connect the psychics of the Laboratory with those of real Life.

4. *Scope of the Present Experiments.*

The present article, therefore, advocates a "Correlational Psychology," for the purpose of positively determining all psychical tendencies, and in particular those which connect together the so-called "mental tests" with psychical activities of greater generality and interest. These will usually belong to that important class of tendencies produced by community of organism, whereby sufficiently similar acts are almost always performed by any one person in much the same manner; if, for example, he once proves good at discriminating two musical tones, he may be expected to manifest this talent on any subsequent occasion, and even in another portion of the scale.

For finding out the classes and limits of these individual functions, modern psychology seems to have mainly contented itself with borrowing statements from the discredited "faculties" of the older school, and then correcting and expanding such data by inward illumination. The following work is an attempt at the more fatiguing procedure of eliciting verifiable facts; the good intention and the difficulty of such an enterprise may, perhaps, be allowed to palliate the shortcomings in its execution. Our particular topic will be that cardinal function which we can provisionally term "General Intelligence;" first, there will be an inquiry into its exact relation to the Sensory Discrimination of which we hear so much in

laboratory work; and then—by the aid of information thus coming to light—it is hoped to determine this Intelligence in a definite objective manner, and to discover means of precisely measuring it. Should this ambitious programme be achieved even in small degree, Experimental Psychology would thereby appear to be supplied with the missing link in its theoretical justification, and at the same time to have produced a practical fruit of almost illimitable promise.

CHAPTER II.

HISTORICAL AND CRITICAL.

1. *History of Previous Researches.*

Though, as above stated, mental correlation has in general met with great neglect, yet a certain number of psychologists, including several of the best known, have from time to time turned their attention that way also. It therefore seems advisable, before entering into the present work, first briefly to survey the results of these previous researches; they will be found on the whole to indicate some very remarkable conclusions.

Only those correspondences will be taken into account in which both terms compared are of a psychical nature; many investigators, after determining the chief measurements of their subject's mind, proceed to make their record still more complete by also noting his most prominent bodily characteristics and external relations, such as his height and weight, the shape of his head, the color of his eyes and hair, the birthplace of his mother, etc. Such considerations, however interesting, do not quite fall within the scope of the present inquiry.

Galton. The first hint appears to have come from that suggestive writer, Francis Galton. As early as 1883, the latter stated that he had found men of marked ability to possess on the whole an unusually fine discrimination of minute differences in weight.¹ The pregnancy of this idea is unmistakable. But Galton appears to have been diverted from the point by other interests, and to have contented himself with the above general impression, without clinching the matter in systematic investigation. In 1890, however, on Cattell publishing an article about "Mental Tests and Measurements,"² a remark was appended by Galton suggesting the desirability of comparing such laboratory values with "an independent estimate of the man's powers. . . . The sort I would suggest is some-

¹ "Inquiries into the Human Faculty."

² Mind, 1890, p. 380.

thing of this kind,—‘mobile, eager, energetic; well shaped; successful at games requiring good eye and hand; sensitive; good at music and drawing.’” It will be seen that subsequent investigators have unanimously preferred a much less lively programme.

Oehrн. The earliest actual experiments in mental correlation seem to have been those of Oehrн,¹ in 1889, which at the same time furnished the starting point for that special branch termed by him, and now popular as, “Individual Psychology.” The latter must, however, be fundamentally distinguished from the “Correlational Psychology” here advocated. For the former deliberately bases itself upon introspectively determined faculties and upon mental tests; whereas the latter *begins* by empirically ascertaining both the faculties and the precise value of the tests. The former endeavors to discover those small deviations from general law which constitute “individuality,”² while the latter, on the contrary, proposes methodically to eliminate individualities as an obstacle to further progress, being itself, no less than General Psychology, in search of laws and uniformities.

Oehrн tested ten subjects in Perception (“Wahrnehmungsvorgang”),³ Memory, Association, and Motor Functions. In accordance with his standpoint of a priori assumed faculties, he does not correlate the results with any independent estimate of his subjects’ intellectual powers, but only the tests with one another. He eventually comes to the conclusion that Perception, Memory, and the Motor Functions are “proportional to one another,” but that Association is rather inverse to all the others!

Boas. The comparison desired by Galton between these laboratory tests and on the other hand the psychics of practical life was, as far as I am aware, first undertaken seriously by Boas. In 1891, the latter examined no less than 1,500 school children as to their Sight, Hearing, and Memory; and then—following the example of the semi-anthropometrical correlations of Porter and others—he proceeded to compare their performances in the above respects with their “Intellectual Acuteness” (as estimated by their teachers). On the first two heads, un-

¹“Experimentelle Studien zur Individual-Psychologie,” Dorpaten Dissertation, 1889.

²Oehrн and Krapelin propose to study “the fine deviations from the great fundamental features of psychical conformity to law;” or again, “to determine the essential differences between minds.”

³Even the term “individual” does not seem very happy, since it chiefly awakens the impression of dealing with individuals as contrasted with masses. In this latter and much more appropriate sense, Wundt uses “Individual Psychology” in opposition to his “Folk Psychology” (*Grundriss der Psychologie*, p. 28).

fortunately, the results have never been published. But as regards Memory, wherein his method of procedure in the main resembled that of Oehrn, the facts elicited were elaborated by Bolton,¹ who comes to the following conclusions:

"The Memory Span increases with Age rather than with the growth of Intelligence."

"The Memory Span measures the power of concentrated and prolonged Attention."

"Intellectual Acuteness, while more often connected with concentrated Attention, does not require it, and it cannot be said that those pupils who are bright intellectually are more distinguished on account of their good memories."²

It will be observed that these results are in sharp antagonism with the view of many modern psychologists, notably Wundt, who would make Attention the very essence of intellectual power.

Gilbert. In 1893, at New Haven, another series of experiments was carried out upon an almost equally extensive scale, and is still among the most important contributions to the subject. J. Gilbert applied several mental tests to about 1,200 children of both sexes, and then compared the results with their "general ability" (again as estimated by their respective schoolmasters).³

On this occasion, the original assertion of Galton was to some extent practically corroborated. For Gilbert believes himself to find a real correspondence of Intelligence with Sensory Discrimination both of weights and of shades. He also, like Bolton,⁴ discovers a slight correspondence with Memory; in Gilbert's experiments the child, instead of learning by heart a row of figures, had to give his judgment as to when a musical tone had lasted just as long as a previously sounded standard one.

But the correspondence deemed most positive and conspicuous was that between Intelligence and "Reaction-time." This is particularly suggestive, on reflecting how especially this Reaction-time depends upon concentration of the Attention. The indication would therefore accord rather with Wundt's view than with that of Boas. Curiously enough, when the Reaction-time is made more obviously intellectual by further complications (Discrimination and Choice), then the above correspondence becomes reduced in amount.

Scripture. In the same little volume, appears an account of an interesting experiment by Scripture, as to the correspondence

¹ *Am. Jour. of Psych.*, Vol. IV, p. 362.

² *Am. Jour. of Psych.*, Vol. IV, pp. 379, 365, and 366.

³ *Stud. Yale Psych. Lab.*, Vol. II, p. 40.

between shortness of Reaction-time and swiftness in lunging with foils. Unfortunately, his subjects are only seven in number. He feels himself, however, "fully justified" in coming to the conclusion that "the average fencer is not quicker in simple reaction than a trained scientist, and neither class shows an excessive rapidity."¹

The first part of the above sentence would well harmonize with the intellectuality found also by Gilbert to be connected with speed in pressing a button; but the latter part is difficult to reconcile therewith, at any rate without painfully lowering the credit of "trained scientists."

Dresslar. Also in 1893, a quite new kind of correlational factor is investigated, that of natural illusions. As is well known, if we pick up two things of the same weight but of different size, we are almost irresistibly inclined to estimate the larger one as being the lighter of the two; strangely enough, the illusion still persists even after we know that the weights are really equal. Some 173 boys and girls were tested in this respect by Dresslar, and at the same time were classed by their teachers into "bright," "good," and "dull."² Dresslar found that the phenomenon was perfectly constant throughout the ages tested, 7 to 14 years, and that instead of the fallacy chiefly affecting the stupid children, as might have been expected, it on the contrary showed itself the more powerful in proportion as the child was "brighter;" hence he concluded: "The more intelligent the children, other things being equal, the stronger are the associations between the ideas of size and weight."

Griffing. In the following year, H. Griffing examined the other chief aspects of Attention.³ The two former workers had dealt with its concentration or *intensity*; Griffing now inquires about its amplitude or *extensity*. He does this by the well-known method of "tachistoscopy:" a number of syllables are exposed to view for a very brief moment, and the subject has to try how many he can read in this practically simultaneous manner.

The result of two independent sets of such experiments is in both cases that "the brighter students tend to excel."

Bourdon. About now we come upon a new and significant phenomenon in the course of these researches. The latter, though originally prompted in England, had forthwith been transplanted to America, in which country alone up to the present date they had been cultivated so as to bear fruit. But at

¹ Stud. Yale Psych. Lab., Vol. II, p. 122.

² Amer. Jour. Psych., Vol. VI, p. 343.

³ Amer. Jour. Psych., Vol. VII, p. 227.

length the Old World also woke to the necessity of answering the questions. France led the van, soon bestowing upon the problem an original and characteristic impress.

In August, 1895, appeared an article by Bourdon entitled "Recognition, Discrimination, and Association."¹ For this investigation, about a dozen subjects were tested in

(1) the power of recognizing words previously shown to them;

(2) the power of quickly and accurately erasing from a printed page certain given letters of the alphabet (this was one of Oehrns tests for "den Wahrnehmungsvorgang");

(3) the number of ideas arising in the mind within one minute on a given suggestion.

The conclusion arrived at² is that all three faculties present some correspondence with one another, but that this is much most marked between Recognition and Association.

Thus upon this occasion, as in the work of Oehrn, the mental tests were only compared with one another, and not with any independently obtained estimate of Ability. Like the German, Bourdon appears to consider that these three faculties, Recognition, Discrimination, and Association, are so satisfactorily represented by the tests,³ that any otherwise gained values of Intelligence would merely be vaguer and less trustworthy versions of the same data.⁴

Binet and Henri. Again in France, towards the end of the year, there appeared an important article of similar tendency, bearing the well-known signatures of Binet and Henri⁵ and setting forth the urgent need of "studying the relations that exist between different psychical processes." They propose the following ten tests: Memory, Mental Images, Imagination, Attention, Faculty of Comprehending, Suggestibility, Ästhetic Sentiment, Moral Sentiments, Muscular Force and Force of Will, Cleverness and "Coup d'oeil." By these means, they hope to measure off "a personality" in a fairly exhaustive manner within 1 to 1½ hours.

In the tests themselves, there is a new feature to be noticed. Hitherto, these had been of the most elementary and unequivocal nature possible, as befits the rigor of scientific work. But

¹ *Revue philosophique*, Vol. XL, p. 153.

² Based only upon nine of the subjects, for some had not undergone all three tests.

³ Same article, p. 185.

⁴ "La méthode à laquelle j'ai eu recours pour étudier ce phénomène (de la Reconnaissance) . . . pourrait servir à déterminer assez promptement et avec une rigueur satisfaisante quelle est l'aptitude d'une personne à reconnaître." And similarly for the other two faculties.

⁵ *L'année psychologique*, Vol. II, p. 411.

this very simplicity had much increased the difficulty of making the test truly representative of any more complex psychosis. Binet and Henri appear now to seek tests of a more intermediate character, sacrificing much of the elementariness, but gaining greatly in approximation to the events of ordinary life. The result would seem likely to have more practical than theoretical value.

Next year Binet begins to put his interesting programme into execution.¹ He examines about 80 children and 6 adults as to powers of describing a picture shown to them, and by this means discovers the existence of five fundamental types of character, the "describer," "the observer," "the erudite," "the emotional," and "the idealist." "It is perhaps the first result," Binet remarks, "that has hitherto been produced by the experimental study of the higher intellectual faculties."

Binet then compares these new types with "the notes and comments which the professors wrote about their pupils and which the Director of the school has carefully checked." But as to the result of this comparison, unfortunately, only the following brief remark is made public: "Of five pupils whom I had put into the 'emotional' group, four had a cold temperament, a dry nature, and a little sensitiveness; the fourth alone seemed sensitive."

Sharp and Titchener. The above work of Binet and Henri found a speedy re-echo from the other side of the Atlantic. Some experiments with the avowed object of examining this new class of test are now recorded as taking place at Cornell University under the direction of Dr. Sharp and with the aid of Prof. Titchener.² These were expressly intended to depart from the older "German procedure" of dealing solely with the "elementary mental processes," and instead were to subject to trial the "French procedure" of directly handling the "complex" ones.

The following classification was adopted: Memory, Mental Images, Imagination, Attention, Observation, Discrimination, and Taste. The subjects consisted of three male and four female advanced students. No independent information was obtained concerning the subjects' respective mental powers, it being only attempted to ascertain whether the tests were consistent among themselves.

The results are not very encouraging:

"The lack of correspondences in the individual differences observed in the various tests was quite as noticeable as their presence."

¹ *L'année psychologique*, Vol. III, p. 296.

² *American Jour. Psych.*, Vol. X, p. 348.

"But little result for morphological psychology can be obtained from studies of the nature of the above investigation."

"In the present investigation the positive results have been wholly incommensurate with the labor required for the devising of tests and evaluation of results."

In conclusion, Sharp suggests the advisability of judiciously combining the characteristics of both the French and German procedures with one another.

Wagner. Almost simultaneously, the idea of collating mental tests with more practical methods of appraisement begins to take root in Germany also. In 1896, a series of experiments for the purpose of inquiring into the question of fatigue of school children was carried out at Darmstadt under the direction of Dr. Wagner.¹ The children were from the new Gymnasium there and seem to have amounted in all to 44 (though the information on this point is not very definite.) The test investigated was the old one of Weber which had recently again been brought to the notice of pedagogical circles by Griesbach. As is well known, it consists in ascertaining how near together two points can still be distinguished from one another by touch. On this occasion, care was taken to obtain an estimate of every child's Natural Talent (*Begabung*), Industry, Attentiveness, Nervousness, and sometimes Temperament.

Unfortunately for our present purpose, the intention of the experiment was not so much to correlate these psychical qualities with the children's absolute sensitivity, as with the reduction in such sensitivity produced by the fatigue of lessons. This reduction is stated to correspond closely with the amount of Attention paid by the child, but to be almost independent of his Natural Talent. Once more, therefore, Attention and Ability are contrasted instead of being identified.

As far as concerns the children's unfatigued condition, our real present topic, we only learn that the nervous and indisposed have a less fine tactful sensitivity than the others.

Ebbinghaus. About the same time, another and much more extensive investigation was officially instituted in Silesia for the same purpose. Two entire upper schools, a boys' Gymnasium and a girls' High School, were before and after work subjected to three tests: the two old ones of Oehrni for Memory and Association (memorizing and adding numbers respectively), and the new "Combination Method"² of Ebbinghaus. The latter observer in discussing the results devotes no less than one entire section out of four to considering the rela-

¹ Sammlung von Abhandlungen a. d. Gebiete der päd. Psych., Vol. I.

² In this, a purposely defective text is given to the subject, and the latter has conjecturally to fill in the missing parts.

tions shown between these tests and the children's general intellect.¹

He comes to the conclusion that the school order shows an appreciable correspondence with all three tests, but least so with Memory and most with his own new Combination Method. He particularly points out that in the last mentioned this correspondence applies, not only to difference of class, but also to position within each class; whereas in the case of Memory, he thinks that if anything the least intelligent succeed the best!

The Combination Method would appear to resemble the new type of test recommended by Binet and Henri, to the extent of presenting a rather intermediate character between the elementariness of normal laboratory work and the complexity of practical activities.

Wiersma. To depart for a moment from the chronological to the logical order of events, this favorable verdict of Ebbinghaus concerning his own new method was in 1902 strongly corroborated by some experiments of Wiersma.² This time, three schools were brought into service. Two of them were special training establishments for male and female teachers respectively, from fourteen to nineteen years old. The third was a "Nachbildung" School, namely, one for those of both sexes who had already gone through the six classes of the elementary school; they consequently aged from eleven to fourteen. The total number came to about three hundred.

Following closely in the steps of Ebbinghaus himself, Wiersma finds his average results to improve regularly with the higher classes and with the higher sections of each class. He takes great pains to analyze the factors upon which such school position depends, and arrives at distinguishing Age, Educational Development (*Entwickelung*), and Natural Talent (*Begabung*). In many complicated tables and graphs, he marshals evidence that the observed correspondence is most of all due to the last named factor.

Binet and Vaschide. In 1897, the question is again attacked by Binet, now in partnership with Vaschide.³ But there is a remarkable return, as far as psychics are concerned, to the old less aspiring forms of tests. For he once more examines children in Reaction-time, Reaction-time with Choice, and Memory of Numbers. In addition thereto, he devises the ingenious test of motor ability called Dots ("petits points"); this consists in seeing how often the subject can tap with a pen on a piece of paper in 5 seconds. The intellectual order of the

¹ Zeitschrift für Psychologie und Physiologie, Vol. XIII, p. 401.

² Zeitschrift für Psychologie und Physiologie, Vol. XXX, p. 196.

³ L'année psychologique, Vol. IV.

children was again obtained from their respective ranks in class. The subjects numbered 45 and averaged about 12 years of age. The results are exactly opposite to those of Gilbert, for Binet sums up as follows:

"The Intellectual Order harmonizes badly with Reaction-times and harmonizes well with the Memory of Numbers." But better than either appears the correspondence with his own "Dots."

This work was quickly followed by similar tests upon older subjects. For such purpose, Binet and Vaschide turned to the Normal School of Teachers at Versailles and there examined 43 youths ranging from 16 to 20 years of age.¹ This time, the scanty positive results of the former experiments are still further reduced; for even the correlation with Memory is somewhat less in evidence. The relation with the "Dots" again presents an unbroken regularity, but this time it seems to have become *inverse*, the stupidest tapping with the greatest speed!

Seashore. We next come upon an interesting series of carefully conducted experiments, to which we shall frequently have occasion to refer. It took place from 1897 to 1899 at the University of Iowa, under the direction of Dr. Seashore, the subjects being nearly 200 children varying from 6 to 15 years inclusive.²

Here, the negatory note that we have first heard from Binet is reiterated, and now in much fuller tones. As regards General Intelligence (again as estimated by the teachers) and Memory of Time, between which Gilbert had found a very marked correspondence, Seashore on the contrary disposes of the question in the following brief words: "There appears to be no functional relation between the two processes."

So, too, between Intelligence and Discrimination of Pitch. For while Gilbert believed himself to have discovered some such correlation, Seashore again says curtly: "There is no functional relation; the distribution of the results practically coincides with the most probable distribution according to chance."

He further compares General Intelligence with the faculty of discriminating Loudness and with various Illusions of Form, Color and Weight. In each case, he finds himself forced to the same conclusion, that there is no indication of the bright children doing differently from the dull ones.

Pearce. Again temporarily deserting the chronological order, some other illusional experiments were carried out in 1903 by H. Pearce.³ These dealt with the subjective localization of

¹ *L'année psychologique*, Vol. IV. ² *Stud. Psych. Iowa*, Vol. II.

³ *Archiv für die gesammte Psychologie*, Vol. I, Part I.

touch sensations and indicated that judgment tends to be warped by other immediately preceding touch sensations in the same neighborhood of the skin. Pearce tested 32 children in this way and came to the conclusion that the warp is directly proportional to the child's intelligence. While thus corroborating Dresslar rather than Seashore, he differs even from the former in that he declares the fallacy to diminish continually with increase of age.

Bagley. The negations of Binet and Seashore were soon carried to a still further extreme. W. Bagley, experimenting at Madison upon 160 children,¹ corroborates Binet to the effect that Reaction-time shows no correspondence with School Intelligence, and also supports the latter's second and inverse rather than first and direct result with "Dots;" for Bagley not only denies all correspondence between any motor abilities and mental ones, but believes his work to demonstrate positively that there is even a marked antagonism between the two, so that excellence in either direction is apt to be accompanied by deficiency in the other.²

Carman. In 1899, at Saginaw, there was another investigation which, if we are to go by the number of children tested, must be judged the most important that has taken place up to this day.³ A. Carman examined 1,507 of them as to their sensibility to pain (and also their strength of hand), noting in each case, as had now become usual, whether the teacher pronounced them to be "bright," "average," or "dull." Not much detail is given, but the following general conclusions are arrived at:

"Boys reported by their teachers as bright were more sensitive than those reported as dull."

"Girls reported as bright were more sensitive and stronger than those reported as dull."

"Those reported as being especially dull in mathematics were more sensitive on the right temple than on the left."

It is further discovered that

"Girls with light hair and blue or gray eyes are less sensitive to pain on left temple," but "on right temple they are more sensitive than the dark."

This information is very curious.

Kirkpatrick. In 1900, a slight rally against the emphatic denials of Binet, Seashore, and Bagley is attempted by E. Kirkpatrick.⁴ About 500 children were tested in three "sim-

¹ *Am. Jour. of Psych.*, Vol. XII, p. 193.

² The relation between motor and mental ability appears to have also been treated by Handcock, Peckham, and Porter, but I have not been able to obtain access to their works.

³ *Am. Jour. of Psych.*, Vol. X, p. 392.

⁴ *Psych. Review*, Vol. VII, p. 274.

ple motor activities," including Binet's Dots, Counting Aloud, and Sorting Cards; their respective performances were then compared with their degrees of intelligence as estimated in the usual way by their teachers.

The result is in every case a decided correspondence.

Thorndike and Woodworth. Hitherto, we have only seen attempts to ascertain what I may, perhaps, be allowed to call "statical correlation." But in 1901, Messrs. Thorndike and Woodworth make a vigorous onslaught upon the still more important and difficult "dynamical correlation." It is useful enough to know whether any child that "taps," etc. with unusual slowness may thereupon straightway be considered as "dull;" but it would be even more to the point to learn that daily practice with the tapping machine could make him any brighter.

Various previous researches had been distinctly encouraging in this matter. Stumpf declares: "The power of mental concentration upon certain points, in whatever region acquired, will show itself effectual in all others also."¹ Gilbert and Fracker had found that practice in one form of discrimination or reaction-time brought with it improvement in the other forms.² Scripture writes, intending apparently to include intellective activities: "Development of will power in connection with any activity is accompanied by a development of will power as a whole."³ And again, Davis comes to the conclusion that "practice in any special act" develops ability "for all other acts."⁴

The experiments of Thorndike and Woodworth, however, give once more a flat negation. The indications are rather that the effect of training in any one mental achievement is of little or no use for other intellectual performances, even very closely akin. The persons tested were carefully exercised until they had acquired considerable proficiency in judging the relative sizes of some pieces of paper of a particular shape. But this so obtained talent seemed completely to depart as soon as new tests were made with papers of a different shape, or even of a somewhat different size. Similar experiments in other sorts of feats led to the same result.

Binet. About the same time, we have another interesting and long contribution from Binet.⁵ His subjects numbered eleven and were specially selected as being the five cleverest and the six most stupid out of a class of thirty-two. These

¹ *Tonpsychologie*, 1883, Part I, p. 83.

² *Univ. Iowa Stud. Psych.*, Vol. I.

³ *Psych. Rev.*, Vol. VI, p. 165.

⁴ *Stud. Yale Psych. Lab.*, Vol. VI.

⁵ *L'année psych.*, Vol. VI, p. 248.

two groups, the "intelligent" and the "unintelligent," were in all the tests opposed and compared.

Binet again confirms, and more positively than ever, that Reaction-time, either with or without the complication of "Choice," has no correspondence with Intelligence. He also contradicts the correlation found by Griffing with the extensive dimension of Attention, in the form of simultaneously reading a large number of letters exposed to view for a small fraction of a second; though, curiously enough, Binet finds a certain amount of correspondence when he quite similarly exposes some arabesque designs. And finally, he finds no correlation with a new test of his own devising, namely: a trial how small a change in the rate of the beats of a metronome can be accurately detected.

But, on the other hand, his formerly successful method of Memory of Numbers now once more showed a marked correspondence with Intelligence. So also, and to a similar amount, is a correlation shown by Erasure of Letters (like that of Oehrn and Bourdon) and by Arithmetical Addition (more complicated than that of Ebbinghaus). So, again, do his new tests of Accuracy in Counting Metronome Taps and in Counting Dots. And so does his other new test, that of Copying: the subject is to copy a certain amount of writing, and then note is taken as to how many syllables he writes from each glance at the original; the more intelligent, the more words per glance.

But the fullest correspondence of all was presented by the very old test of Tactile Discrimination, which we have already seen successfully assayed by Wagner in 1896.

Binet is further strongly of opinion that all these correlations with Intelligence are most marked upon first trial, and that they continually diminish in proportion as the intelligent and unintelligent are both alike given more and more practice in the tests.

Simon. Directly inspired, apparently, by the last research, the correspondence there discovered between intelligence and the Copying test was now corroborated under new conditions. M. Simon conceived the idea that any such correlation should be manifested in especially prominent relief at the Vaucluse colony for backward children. He therefore tries seventeen of them, and finds in fact that, with one exception, all those classed medically as "Idiot" or "Imbécile" can copy fewer syllables at a time than do those merely termed "Dégénéré" or "Débile."¹ He concludes enthusiastically as follows:

"Convenient, short, and exact, this copying of phrases at once constitutes a good method of diagnosing a child's intellectual development at the very moment of the experiment."

¹ *L'année psych.*, Vol. VII, p. 490.

Kraepelin, Cron. Other observers, however, would appear to have been less fortunate in this region. Their application of experimental tests, even to such trenchant opposition as intellectual health and disease, has not led them to results that they have felt able to pronounce entirely unequivocal. The careful work of Kraepelin and Cron¹ comes to the following close: "At the end of these considerations, we will not hide from ourselves that the obtained results have fallen far short of what one is accustomed to expect from collective experiments with the simplest 'mental tests.'"

Reis. When the above investigation was renewed on a more extensive scale by Reis, the latter finds indeed that these tests perfectly well admit of being executed upon the patients in the asylum; but the success would appear almost *too* great to fulfill the desired purpose, for often the patients prove the better performers of the two; a man, for instance, medically diagnosed as suffering from Dementia Paralytica with marked mental incapacity (*deutliche geistige Schwäche*) more than once comes out top of all fifteen subjects sane and insane alike.

Cattell, Farrand, Wissler. Now we come to about the latest and in many respects far the most important of all these attempts to correlate laboratory work with the psychics of real life. For amplitude of design, special experience of the directors, and lucid collation of the results, nothing up to the present has approached the researches which for about the last ten years have been progressing at Columbia University under the guidance of Cattell.

In 1896, the latter, together with Farrand, allowed a brief insight into the nature and extent of the proceeding being carried on. But not till 1901 was the total upshot of all this labor carefully put together and published by Wissler.² By that time, 250 freshmen and some 35 seniors of the University, besides about 40 young women in Barnard College, had undergone the following elaborate series of tests (in addition to others not belonging to the present topic, such as anthropometrical, etc.):

Perception of Size,	Reaction-time,
Strength of Hand,	Rate of Perception,
Fatigue,	Naming Colors,
Eyesight,	Rate of Movement,
Color-vision,	Rhythm and Perception of Time,
Hearing,	Association,
Perception of Pitch,	Imagery,
Perception of Weight,	Auditory Memory,
Sensation Areas,	Visual Memory,
Sensitiveness to Pain,	Logical Memory,
Color-preference,	and Retrospective Memory.

¹ Psych. Arbeiten, Vol. II, p. 324.

² Psych. Rev. Monograph Supplement, June, 1901.

The general intelligence of each student was settled by his average grading in all the different University courses; an amalgamation of these separate gradings resulted in forming eleven classes.

This class standing and all the above laboratory tests are now, for the first time in the history of the problem, correlated together with some mathematical precision. The final conclusions are about as blankly negative as could well be imagined. We are summarily informed that

"The laboratory mental tests show little inter-correlation."

"The markings of students in college classes correlate with themselves to a considerable degree, but not with the tests made in the laboratory."

And on inspecting the actual figures representing the faint correlations in question, it is mathematically evident that not one of them is more than would be expected to occur by mere accidental coincidence.

Aiken, Thorndike and Hubbell. Finally, in 1902, there appears an interesting contribution to the subject from Aiken, Thorndike, and Hubbell. Here "the functions in question were much more alike than were those examined by Wissler.¹ We have examined the relationships between functions in an extremely favorable case." Nevertheless, on the whole the previous negative results are once more strongly corroborated; when some mental functions usually regarded as most purely typical of the associative process are compared together, their correlation turns out to be "none or slight."

2. *Conclusions to be Drawn from these Previous Researches.*

Thus far, it must be confessed, the outlook is anything but cheerful for our project contemplated at the end of the first part, or, indeed, for Experimental Psychology in general. There is scarcely one positive conclusion concerning the correlation between mental tests and independent practical estimates that has not been with equal force flatly contradicted; and amid this discordance, there is a continually waxing inclination—especially noticeable among the most capable workers and exact results—absolutely to deny any such correlation at all.

Here, then, is a strange enough answer to our question. When Laboratory and Life, the Token and the Betokened, are at last objectively and positively compared as regards one of the most important Functional Uniformities, they would seem to present no correspondence whatever with one another. Either we must conclude that there is no such thing as general intelligence, but only a number of mental activities perfectly

¹ Psych. Rev., Vol. IX, p. 374.

independent of one another except for this common word to designate them, or else our scientific "tests" would appear to have been all so unhappily invented as to lie outside the widest limits of those very faculties of which they are supposed to form a concentrated essence.

It is true that Functional Uniformities might conceivably exist of other kinds; but for any such there is even less evidence; nor would they appear at all *a priori* probable, in view of the complete and surprising absence of that important one constituted by community of organism. Failing all Functional Uniformities, any connection between the experimental procedure and practical intelligence can then be no more than "Conceptual." But this is a position scarcely tenable for those whose chief claim is finally to have escaped from the endless tangle of purely introspective argument; moreover, such an admission would shear every experimental research of almost its whole worth and deprive the systems built thereon of their essential base.

Further, if thus the only correspondences hitherto positively tested, those between Intelligence and its variously supposed Quintessesences, have totally failed to reveal any real existence, what shall we say of all the other by no means so apparently self-evident correspondences postulated throughout experimental psychology and forming its present backbone? To take one of the most extensive and painstaking of them, Dr. Schuyten, from 1893 to 1897, continuously amassed evidence to prove a close relation of the middle European temperature with the faculty of "voluntary attention" and even more generally with "the intensity of cerebral activity;" he seems to have repeated his observations on about five hundred different days, upon each occasion indefatigably proceeding round Antwerp from one school to another, visiting most of the time as many as eight. Now, his actual test of "voluntary attention" and "cerebral activity" consisted entirely in noting how many children kept their eyes on their lesson books for five consecutive minutes; but, as far as I am aware, there has not yet been any positive proof that this posture sufficiently coincides with all the other activities coming under this general term of "voluntary attention;" and in view of the universal breakdown of evidence for much more plausible correlations, Schuyten's *a priori* assumption can hardly be admitted as an adequate basis for his wide reaching theoretical and practical conclusions. To try another example, we have seen that a favorite test, successively adopted by Oehrн, Bourdon, and Binet,¹ is that of erasing from a printed page certain given letters of the alpha-

¹The last named seems to have been the first to inquire into, rather than assume, the range of function involved in this test.

bet; but sceptics are still able to contend that because any person can dash a stroke through a's and i's with unusual speed, he need not therefore be summarily assumed to possess an abnormally large capacity for discrimination generally speaking, say, for telling a fresh from an over-night deer's trail, or distinguishing sound financial investments from unsound. Precisely similar criticism may be extended to almost the whole mass of laborious attempts to establish practical applications of Experimental Psychology, whether for pedagogical, medical, or other purposes.

Nor is the case much otherwise even with those stricter and more theoretical researchers who are rather inclined to regard as superficial any experiments involving large numbers of subjects. For however modest and precise may seem the conduct of their own actual investigation, it nearly always terminates with and justifies itself by a number of sweeping conclusions; and these latter will be found to essentially imply some assumed general function or process, such as "memory," "association," "attention," "fatigue," "practice," "will," etc., and at the same time that this function is adequately represented by the laboratory test. To take for instance the speed of mental association, there is hardly a psychologist of note who has not at some time or other made wide reaching assertions on this point, often indeed finding herein one of the pillar stones of his philosophy; the more practically minded, as Kraepelin and his school, content themselves with demonstrating the details of its actual conduct, showing us how the rate will rise with practice or on imbibing tea, how it sinks in proportion to fatigue or mental disorder, how under the influence of alcohol it for a brief moment slightly ascends and then becomes permanently and profoundly depressed. But all these conclusions are derived from observation of one or two supposed typical forms of this "association;" while the extensive experiments of Aiken, Thorndike, and Hubbell reveal that every form of association, however closely similar on introspection, must, nevertheless, always be considered separately on its own merits, and that "quickness of association as an ability determining the speed of all one's associations is a myth."¹ The most curious part of the general failure to find any correspondence between the psychics of the Laboratory and those of Life is that experimental psychologists on the whole do not seem in any way disturbed by it. But sooner than impute to them—the avowed champions of positive evidence—such a logical crime as to prefer their own *a priori* convictions to this mass of testifying facts, it is perhaps pardonable to suspect that many of them do not realize the full significance of the situation!

¹ Psych. Rev., Vol. IX, p. 375.

3. *Criticism of Prevalent Working Methods.*

There is, however, an intermediate way between ignoring all this serious testimony and submissively accepting it; this consists in subjecting it to the most searching criticism of which one is capable. But such a procedure quickly leads to questions of greater generality; if we would deal with the matter at all adequately, we are compelled to enter into a general discussion of the methods universally prevalent for demonstrating association between two events or attributes. To this important topic a special work has been devoted.¹ For the present, we must limit ourselves to the following brief exposition of the chief deficiencies appearing especially to characterize the long series of experiments just reviewed.

In the first place, only one out of them all (Wissler at Columbia) attains to the first fundamental requisite of correlation, namely, a *precise quantitative expression*. Many writers, indeed, have been at great trouble and have compiled elaborate numerical tables, even bewilderingly so; but nowhere do we find this mass of data focused to a single exact result. In consequence, not only has comparison always been impossible between one experiment and another, but the experimenters themselves have proved quite unable to correctly estimate even their own results; some have conceived their work to prove that correspondence was absent when it really existed to a very considerable amount; whereas others have held up as a large correlation what in reality is insignificantly small. Later on, we shall come upon examples of both kinds of bias. With this requisite is closely bound up another one no less fundamental, namely, that the ultimate result should not be presented in some form specially devised to demonstrate the compiler's theory, but rather should be a perfectly impartial representation of the *whole* of the relations elicited by the experiments.

Next, with the same exception as before, not one has calculated the "probable error;" hence, they have had no means whatever of judging how much of their results was merely due to *accidental coincidence*. This applies not only to the experiments executed with comparatively few subjects, but even to those upon the most extensive scales recorded. The danger of being misled by combinations due to pure chance does, indeed, depend greatly upon the number of cases observed, but in still larger degree upon the manner in which the data are calculated and presented.

Thirdly, in no case has there been any clear explicit definition of the problem to be resolved. A correspondence is ordi-

¹ "The Proof and Measurement of Association between two Things."
Am. Jour. Psych., XV., 72.

narily expressed in such a general way as neither admits of practical ascertainment nor even possesses any great theoretical significance; for a scientific investigation to be either possible or desirable, we must needs restrict it by a large number of qualifications. Having done so, any influence included (or excluded) in contravention of our definition must be considered as an *irrelevant and falsifying factor*. Now, in many of the experiments that we have been discussing, even in those upon quite a small scale, the authors have tried to kill as many birds as possible with one stone and have sought after the greatest—instead of the least—diversity; they have purposely thrown together subjects of all sorts and ages, and thus have gone out of their way to invite fallacious elements into their work. But in any case, even with the best of intentions, these irrelevant factors could not possibly be adequately obviated, until some method had been discovered for *exactly measuring* them and their effect upon the correlation; this, to the best of my knowledge, has never been done. As will presently be seen, the disturbance is frequently sufficient to so entirely transform the apparent correlation, that the latter becomes little or no evidence as to the quantity or even direction of the real correspondence.

Lastly, no investigator seems to have taken into any consideration another very large source of fallacy and one that is inevitably present in every work, namely, the *errors of observation*. For having executed our experiment and calculated the correlation, we must then remember that the latter does not represent the mathematical relation between the two sets of real objects compared, but only between the two sets of measurements which we have derived from the former by more or less fallible processes. The result actually obtained in any laboratory test must necessarily have in every case been perturbed by various contingencies which have nothing to do with the subject's real general capacity; a simple proof is the fact that the repetition of an experiment will always produce a value somewhat different from before. The same is no less true as regards more practical appraisements, for the lad confidently pronounced by his teacher to be "dull" may eventually turn out to have quite the average share of brains. These unavoidable discrepancies have always been ignored, apparently on some tacit assumption that they will act impartially, half of them tending to enhance the apparent correlation and half to reduce it; in this way, it is supposed, the result must in the long run become more and more nearly true. Such is, however, not at all the case; these errors of observation do not tend to wholly compensate one another, but only partially so; every time, they leave a certain balance *against the correlation*, which

is in no way affected by the number of cases assembled, but solely by the size of the mean error of observation. The amount of consequent falsification is in physical inquiry often unimportant, but in psychology it is usually large enough to completely vitiate the conclusion. This falsifying influence has in many of the above experiments, especially the more extensive ones, occurred in exaggerated form; for even those experimenters who are most careful in the ordinary routine of the laboratory have yet allowed themselves to be seduced by the special difficulties attending this sort of work; urged on the one hand by the craving after an imposing array of cases—somewhat ad captandum vulgus—and sternly restricted on the other side by various personal considerations (such as restiveness and fatigability of the youthful subjects, fear of deranging school hours, etc.), they have too often fallen into almost incredibly hurried and inadequate methods of testing. Here, again, mere goodness of intention will not avail beyond a very limited extent, for the most painstaking work is far from entitling us to assume that the observational fallacy has been reduced to insignificant dimensions; we can have no satisfactory guarantee, until some method has been devised of precisely measuring the disturbance, and this does not seem to have ever been attempted.

The above criticism, of a perfectly general nature, must suffice for the bulk of the researches cited in this chapter; later on will be found a more detailed examination of those three particular ones which have dealt with precisely the same topic as the present article. If here methodological imperfections have admitted of formulation with unusual sharpness, the fact must by no means be taken as an especial condemnation of these and kindred experiments. Certain faults have, indeed, been especially prominent, as, for instance, the large errors of observation; but, on the whole, the majority of them would appear to contain at least as much good solid work as most of those more strictly confined to the laboratory and to a very small number of "trained" subjects; the former have only afforded a firmer foothold for criticism, because they have confined the question to a more simple unequivocal issue—though not yet nearly *enough* so—and because they have assailed their problem in a square positive manner. The final inconclusiveness of all their labor is not so much due to individual shortcomings of the investigators, or even of the whole branch of investigation, as to the general non-existence of any adequate system for proving and measuring associative tendencies.

Under all these circumstances, in spite of the many previous inconclusive and negatory verdicts, the question of correspondence between the Tests of the Laboratory and the Intelligence

of Life cannot yet be regarded as definitely closed. The only thing so far demonstrated is that the old means of investigation are entirely inadequate. The present undertaking, therefore, has only ventured once more to approach the problem, because believing to have elaborated a new and reasonably complete methodological procedure, such as appears capable of at last bringing light upon this and innumerable other important regions hitherto inexplorable.

CHAPTER III.

PRELIMINARY INVESTIGATION.

I. Obviation of the Four Faults Quoted.

In the last chapter, four grave faults have been charged against the entire antecedent literature on the present topic. The first thing insisted upon was a *precise quantitative expression derived impartially from the entire available data*; we must renounce adroit manipulation of tables and graphs, still further rounded into the desired shape by ingenious argument; the whole of our experimentally gained figures must without any selective treatment simply of themselves issue into one plain numerical value (varying conveniently from 1 for perfect correspondence down to 0 for perfect absence of correspondence); this will here be done in the method that has been successively elaborated by Bravais,¹ Galton,² and Pearson,³ and whose formula will be found on page 252; in addition, lists of individual amounts will be given in full as originally obtained, and therefore can be freely used either for checking the present results or for other inquiries. The second requisite was the probable error; concerning this it may at once be remarked that indispensable as is some evaluation, yet here less than anywhere else have we need, or even possibility, of rigorous exactness; approximate estimates will be appended to each correlation in tables. The third fault indicated was the one deriving from the errors of observation; little can be said in this place concerning the best means of reducing these to a minimum, for on such head our requirements are scarcely different from those already prevailing in all serious psychological research; only some matters of special interest will be briefly touched upon while describing the procedure of the present experiments in the next chapter. Much more important for us is the fact that

¹ *Mémoires par divers Savans*, T. IX, Paris, 1846, pp. 255-332.

² *Proceedings Royal Society of London*, Vols. XL and XLV.

³ *Philosophical Transactions*, R. S. London, Vol. CLXXXVII A, p. 164.

the total effect of all such errors can be measured *en masse* and mathematically eliminated, and that until this has been done, no correlational value can be assumed as even approximately accurate. The formulae for this purpose are given and explained in the already mentioned article;¹ sufficient for the present practical purposes will be briefly recapitulated at the end of the next chapter.

The remaining point is that of irrelevant factors; this also is more fully explained in the said article; it involves a thorough preliminary investigation of all the terms concerned, without which the most skillful experimentation and lucid exposition will only be wasted labor; the produce of such preparatory exploration will form the remainder of the present chapter, while the processes of calculation will be given at the close of the next one.

2. *Definition of the Correspondence Sought.*

The first step towards eliminating irrelevancies is to clearly lay down how much we are to consider relevant, or in other words to properly define the problem at issue. As already stated, universal correspondences can never be the subject of investigation; in practice we are forced to introduce a large number of conventional restrictions, and for profitable work these must be explicit and unequivocal.

Let us first take that of *Kinship*. Putting other animals out of the question, we clearly cannot pretend adequately to sample even the whole existing human species. In order to obtain the simplest and least ambiguous results, it might seem desirable to reduce this source of variation to the least possible dimensions; an ideal experiment would then be wholly confined to a number of sets of brothers or sisters and to determining how far the more intelligent brother is also the better discriminator. But these narrow limits are most inconvenient on practical grounds, and even theoretically there appears no great objection against extending the kinship to any range that does not introduce inconvenient complications; we may eventually find it necessary to exclude differences of social stratum, of sex, etc.; or it may, on the contrary, be found allowable to admit all these and even some amount of internationality.

Next, we must bear in mind that the action of correctly distinguishing between two sounds or weights is a matter depending on many factors, and we must decide how many of these should be rejected as foreign to our purpose. In the present case, it seems best to limit the object of our research to that portion of the discriminating act which appears to constitute its

¹ "The Proof and Measurement of Association between Two Things."

specific core, excluding as far as possible such more outward influences as Zeal, Endurance, Manual Dexterity, Memory, etc. The last named requires especial attention, since it necessarily enters into all these mental tests. Its influence will greatly differ according to the method of procedure: the interval intervening between the two compared sensations may vary from a small fraction of a second to several minutes; it may leave the reagent's attention undisturbed, or it may make distracting calls upon it (such as causing him accurately to adjust an instrument). One of the investigations cited in the last chapter has gone so far in this direction, that a test called by the experimenters themselves "Perception of Pitch" is by the compiler preferably termed "Pitch Memory."¹ Now, the correlations of Memory — as far at any rate as my own researches have hitherto gone — would indicate laws entirely different from those of Discrimination; if this be confirmed, any interference of the former factor may gravely perturb an investigation into the latter.

Thirdly, the correspondence selected for inquiry in the present case is that between *natural innate faculties*. By this definition, we explicitly declare that all such individual circumstances as after birth materially modify the investigated function are irrelevant and must be adequately eliminated; our results might, therefore, be wholly vitiated if we threw together people in disparate ages, those in full vigor and those tired or ill, those who have already practised the test in question and those to whom it is new, etc. To obviate this, we are obliged to search through the records of previous work, so as to ascertain all the influences that have been found seriously to affect any of the variants now in question. The chief results of such a preparatory investigation will now be briefly detailed, and for the sake of conciseness and unequivocality we will at once as far as possible explain and correct these antecedent data by collating them with the information subsequently derived from the present experiments themselves. The most prominent turn out to be Practice, Age, and Sex, which will be discussed in this order.

3. *Irrelevancies from Practice.*

The importance of having defined our correlation will here at once be evident. For if we had wished to inquire whether

¹Some experimenters (Lehmann, Angell), it is true, have found that "correctness of sensory judgment is practically independent of time interval" up to at least a minute, and even of deliberate distraction (Angell, Wundt's Phil. Stud., Vol. XVII, p. 11). I have, however, been unable to convince myself that this conclusion holds at all good for the ordinary conditions of experiment.

sensory acuteness and general intellect are correlated *dynamically*, if (as stated by many persons on the strength apparently of a priori reasoning) we had assumed that the discrimination of minute differences of sensation "is to be cultivated as the foundation of all intelligence,"¹ then we should have had to admit the variations due to Practice as perfectly relevant and we should have looked for a continual expansion in people's general ability in proportion to the labor they had expended on distinguishing tones, shades, and weights from one another. In the present experiments, however, it has been preferred to commence by investigating the *statical* relation; it has not been asked whether Intelligence is produced by development of sensory acuteness, but whether the former original endowment is on the whole accompanied by a corresponding amount of the latter also. From this it follows that we are bound to carefully eliminate differences of previous exercise.

(a) *Pitch.* In this branch of Discrimination the effects of Practice are especially conspicuous, but nevertheless they are not very easy to trace out with the quantitative precision required for our purpose. Experiment has, indeed, unanimously demonstrated that the threshold of discrimination to be reached by trained and competent acousticians, even when using quite different apparatus, is in the immediate neighborhood of $1/3$ v. d.;² but there is no such general agreement concerning the average threshold of "unpractised" persons. Delezenne, for instance, finds that an interval equivalent in the centre octaves to about 1 v. d. "becomes sensible to the least practised ears, as I have assured myself on several people,"³ But Preyer, on the other hand (though quite agreeing with Delezenne as regards the powers of practised acousticians), tried a few unpractised persons and found that they could not all decide with certainty unless the two tones differed by as much as 8 v. d.⁴ And

³ Die Grenze der Tonwahrnehmung.

¹ Pulling: Teachers' Text book of Practical Psychology, p. 16.

² In the present work, it is taken that a threshold obtained in v. d. (=vibrations difference) for any tone within the two octaves above the middle C may be assumed also to hold approximately good for the remainder of these two octaves and both above and below gradually to augment. So much appears sufficiently demonstrated by the works of Preyer (Die Grenze der Tonwahrnehmung), Luft (Wundt's Phil. Stud., Vol. IV, pp. 511 ff.), and Meyer (Zeit. f. Psych. u. Phys., Vol. XVI, p. 352). Hence the results of all experiments conducted anywhere near the centre of the ordinary musical scale admit of being easily collated with one another. For procedure in comparing any values obtained by the method of "Minimal Changes" with those by "Right and Wrong Cases," see Lorenz (Wundt's Phil. Stud., Vol. II), Merkel (*Ibid.*, Vols. IV and VII), Kämpfe (*Ibid.*, Vol. VIII), and Mosch (*Ibid.*, Vol. XX).

⁴ Recueil des travaux de la Soc. de Lille, 1827, p. 4.

when we come to the more extensive experiments of Cattell and Farrand upon male and female University students, we find that the average error for the F below the middle C was over a whole tone, which (after making all possible allowances for diversity in procedure; both experiments were conducted on monochords) must be taken as at least twenty times worse than the results of Delezenne.¹

This and similar apparent discrepancies, however, seem to proceed from a too vague and also too narrow conception of "Practice." Many writers freely make assertions concerning "highly trained reagents" (*geübte Versuchspersonen*) without stating — often, it would appear, without clearly realizing — whether the latter are merely trained in music generally, or have had special practice in distinguishing minute differences of Pitch. Other authors, on the contrary, term all their subjects "unpractised," even though accomplished musicians, if they happen to be new to this particular experiment; in such sense must be taken Delezenne's designation "*les oreilles les moins exercées*," for his reagents seem from the context to have been distinctly above the average as regards general musicality. Now, though the hyperacute sensitivity characterizing the acousticians has *not* been found (by the present writer at least) in most musical performers, even professional, yet the latter are at any rate considerably superior to the average person. Thus we so far have three very distinct grades of training: acousticians, musicians, and the rest.

There is also another kind of practice that must always be duly considered, namely that gained during even brief experiments. The Columbia undergraduates, for instance, who had a monochord put into their hands and were told to forthwith adjust it so as to produce unison with a previously given note, were, as regards training, almost as far removed from subjects whose examination lasts a quarter of an hour, as these again are from the long exercised reagents of Stumpf, Luft, etc.

Lastly, there is a factor which though analogous to Practice yet must not be too summarily identified therewith; this is the influence of General Culture. It will require discussion in some detail, as it does not seem to have hitherto gained much attention, since it is peculiarly liable to engender fallacies in comparative tests of children's discrimination, and may eventually prove of profound theoretical importance. A glance at Table V will present a very marked inferiority on the part of the villagers as compared with the better social classes, for the former have a threshold more than twice as great as the latter. Small as is the number of cases involved, the discrepancy is so

¹ Psych. Rev. Monograph Supplement, June, 1901.

large as practically to exclude mere accident.¹ The next readiest explanation is that villagers have far less opportunities of hearing music. But nearer investigation does not very much confirm this view; for the village in question rejoiced in the possession of an unusually fine set of church bells, and whoever has listened to a band of sturdy yeomen ringing for eight consecutive hours through all the possible "changes" will realize that those in the neighborhood have ample opportunity of hearing musical intervals. (The present writer had the misfortune to reside within a few yards of the said church.) Moreover, with one exception, all the villagers examined were in the habit of singing at home; while two of them were no mean performers on the violin and organ respectively. The violinist, it is true, exhibited the same discrimination as the average good musician; but the organist formed no exception to the general obtuseness, other than a far quicker capacity for improvement.

The same phenomenon is equally conspicuous among the young; for the village children, in spite of regular instruction and practice in singing, had a threshold twice as great as that found in the high class school. Also the relation between the children and the adults tells a similar tale; for in those social strata which continue after full growth to fully exercise their higher mental capacities, there is no appreciable alteration of ability to distinguish tones; the high class school has precisely the same average as the cultivated grown-up person. Among those, on the other hand, who at an early age are compelled to turn all their energy to muscular activities, we find a correspondingly rapid falling away of discrimination; bad as the child villagers were, those adult proved themselves nearly twice as inapt.

The only other experiments explicitly in this grade of society, with which I am acquainted, are the interesting ones of C. Myers;² and to these, without much straining, a similar interpretation may be given. His European village children (many of them possibly very young; unfortunately the ages are not shown) present an even lower average than those in Table I; that his adults slightly improve, while mine, on the contrary, go down, may well be attributable to the known habits of self-cultivation characterizing the Scotch peasantry. Myers' Papuan children, with only a smearing of alien education, are accordingly worse than the Europeans of any age or class; and on returning to the more congenial occupations of

¹ The improbability of mere chance works out mathematically to about 100 to 1 for this series of observations alone, not to mention the corroborative evidence of the children and other experimenters.

² Archives of Otology, XXXI, 4, 1902.

pearl-diving and cannibalism, their discriminative powers appear to sink lower still.

There is thus considerable evidence that the ability to discriminate pitch is largely affected by deprivation of General Culture. At the same time, this influence is only conspicuous in tests of an unfamiliar character; the villagers were incompetent at the—for them useless—task of distinguishing tones, but they show no inferiority whatever in the more practical faculties of telling one shade and weight from another. It would therefore seem that General Culture is not an independent factor that may be superadded to special sensory training, but no more than a possible partial substitute, just as the carbohydrates can excellently supplement but never supersede food-stuff containing nitrogen.

The various effects of all these degrees of Practice may now be expressed roughly enough, but still *quantitatively*, in the following tabular form:¹

*Median threshold for Discrimination of Pitch among adults.
(Expressed in v. d. for the two centre octaves.)*

	Papuans	European Villagers	Non-Musicians of General Culture	Musicians	Specially Practised Reagents
With a minimum of fore-exercise.	—	30	10	4	$\frac{1}{2}$
After 15 minutes good fore-exercise.	16	8	4	2	$\frac{1}{3}$

¹ The Table appears to bring the various results of the best workers into very good harmony with one another and also with my own. It is regretted, however, that there is one exception to this general reconciliation; for it has been found impossible satisfactorily to accord the above estimates with a noteworthy series of experiments at Iowa in 1896, under the general direction of Seashore. There, the male University students showed a median discrimination of about 10 v. d., and the women one of about 9 v. d. These values, taking all things into consideration, appear at least twice as low as the analogous ones of Delezene, Preyer, Gilbert, and myself. Moreover, it is even harder to reconcile Seashore's results from adults with his own simultaneous ones from children 10-14 years old; for the latter perfectly coincide with our table, showing a discrimination of about 4 v. d., and would therefore have double the sensitivity of the adults!

It must also be mentioned that Seashore arrives at a very different opinion from the above as regards the general effect of Practice. He

(b) *Sight.* In visual discrimination there is evidently less scope for further improvement by special training, seeing that the ordinary necessities of daily life are already sufficient to call forth a large portion of most persons' potential faculty in this respect.

But even here a considerable residuum of dormant power usually remains over and may be awakened on sufficient exercise. Schirmer,¹ for instance, found his threshold to continually diminish for about a week of continual practice, after which it seemed to him to have attained its maximum, while Müller-Lyer² and Simon³ continued to make appreciable improvement for many months. Simon further noted the curious fact that practice in judging with both eyes brought with it but little betterment in judging with either eye alone; but when the left eye had been practised by itself to its maximum powers, then the right eye also, although itself unpractised, was nevertheless found to have advanced to its maximum. My own experience fully coincides with this view that the generality of people can considerably reduce their threshold by practice. Here, however, no appreciable influence is manifested by General Culture; the adults of the diverse social classes show just about the same average, and also as regards the children the difference is about what was to be expected from the difference in conditions of test.

This visual disparity between trained and untrained is difficult to estimate quantitatively. For previous experimenters upon unpractised reagents never seem to have attempted more than a rough empirical gradation incapable of being compared with other work. Our evaluation must therefore rest solely upon the present experiments, which, being conducted by daylight, are far from guaranteeing sufficient precision on this head; but still, as they turn out to agree perfectly with the mean of previous records of trained reagents, in all probability they are approximately accurate as concerns untrained ones also. In this case, the median threshold for discrimination of medium shades of gray, after 15 minutes fore-exercise, may be taken as involving the following difference of luminosity:

Untrained Reagents	Trained Reagents
1/30	1/120

holds that neither musical education nor special training can materially affect most people's power of discriminating, and he supports his view by inquiries instituted among the parents of his children and also upon a few experiments.

¹ Arch. f. Ophthalm., XXXVI, 4.

² Arch. f. Anatomie u. Phys., 1889, Phys. Abth., Supplementband, p. 96.

³ Zeit. f. Psych. u. Phys., Vol. XXI, p. 433.

It would thus appear that in this faculty the influence of training is considerable, but still only about one-third of that in discrimination of pitch.

(c) *Weight.* Here, curiously enough, Practice seems to have remarkably little effect; the most highly trained experimenter has shown no advantage over the first comer.

Weber, for instance, declares $1/40$ to be the difference just distinguishable by "quite the majority of human beings, without any long preliminary practice;"¹ yet he himself, in spite of all his long labors, appears to have been much below this standard.² Similarly his successor, Fechner, after prodigiously extensive training, was in the end but a very moderate performer.³ A more direct proof of the inefficacy of prolonged exercise is given by the case of Biedermann and Loewit; for these observers at the very beginning of their experiments found their threshold for one-half pound to be equal to $1/21$; yet at the close of their protracted research, this threshold had only improved to $1/23$.⁴

But however true this may be as regards training of very great duration, I have convinced myself that it does not hold good as regards the few minutes of fore-exercise; for my reagents have almost invariably discriminated better at the end of their fifteen minutes than when they first entered the room; sometimes the improvement is enormous.

We come, therefore, to the general conclusion that a few minutes' practice are necessary for most people to accustom themselves to the test, but that further training has little or no effect.

(d) *Intelligence.* As to the effect of training in this direction, there unfortunately appear to be no available data, except such conclusions as issue from the present experiments themselves. All that could be done before commencing work was to contrive that the reagents should, as far as possible, be on equal terms with regard to previous education.

4. *Irrelevancies from Age.*

This factor is obviously one that must largely influence both Sensory Discrimination and Intelligence, since nobody can suppose that any mental faculties remain at one constant level from first birth to the prime of manhood and on through the last senile decay. The matter acquires peculiar importance owing to the practical reason that we find it much easier to

¹ Tastsinn u. Gemeingefühl, p. 541.

² Annotationes anatomicae et physiologicæ, Prol. XVII.

³ Psychophysik, Vol. I, p. 186.

⁴ Berichte d. kais. Akademie d. Wissensch. zu Wien, Math. Naturw. Classe, Vol. LXXII, p. 343.

subject children to our experiments than to induce the same number of adults to let themselves be tested. Consequently, every research of this kind so far has dealt exclusively with reagents not yet beyond adolescence. We will again consider each of the four variants in turn.

As regards *Pitch*, the first notable inquiry into the development of discrimination appears to have been that of Gilbert, who comes to the conclusion that from the age of six to at least that of seventeen years there is a continual though irregular improvement.¹ But the next investigator, Seashore, arrives at the very different opinion that no betterment takes place after ten, at which period he finds children to be fully equal to average adults; he even goes so far as to pronounce that "the organ of Corti reaches its maximum efficiency at the age of about ten, and that it then begins to deteriorate."²

By the light of the previous section, however, it seems possible satisfactorily to reconcile these very divergent results. For we have just seen the immense effect of Practice, especially in this matter of pitch, and we may not unreasonably suppose such faculties to arrive at maturity earlier or later according as they have been more or less fostered by education, special or general. This hypothesis will be found frequently corroborated in the present work. The high class school perfectly harmonizes with Seashore's elder children in presenting no further increase of capacity after nine, at which age the average threshold already equals that of adults; and this agreement with Seashore extends not only to the relative powers but even to their absolute performances, for both lots of children show a mean of about 4 v. d. But when we turn to the villagers, then the accordance is no less entirely with Gilbert: from 5½ to 7 years, they average 40 v. d.; from 7 to 8, 23 v. d.; from 8 to 10, 15 v. d.; and from 10 to 14, 6·7 v. d.; and here again also the absolute values fairly coincide with Gilbert's, although the latter used such a different testing instrument as an adjustable pitch-pipe.

Even into very advanced life, the influence of Age appears still interreflected with the previously mentioned factor of Culture. For those adults who exercise their higher intellectual functions abide to almost the end of their days at their best level; nothing is observable to correspond with the generally believed diminution of their audible range from 11 octaves to 10. To myself, it was no small surprise to find that the persons of 60, 70, and even 80 years could discriminate quite as acutely as those in their physical prime. But among the grown-up villagers there seems to occur a marked bifurcation: those

¹ Stud. Yale Psych. Lab., 1893.

² Stud. Psych. Iowa, Vol. II.

of them who on leaving school take to violin playing, hand-bell making, etc., continue to develop the discriminative function until it eventually becomes quite as fine as if it had enjoyed a better start; but those who have not had such special later advantages will not be saved by occasionally hearing music from lapsing into great dullment.

Turning next to *Sight*, Gilbert's very extensive experiments again present a continual growth perceptible even up to the seventeenth year; between nine and fourteen, the ages chiefly entering into the present work, the average threshold becomes reduced by about one fourth. My own results turned out quite similar as regards the village children, but in the higher class school development appeared to occur earlier. With respect to Old Age, this again appeared to bring with it no impairment whatever; the supposed "sharpness of young eyes" did not make itself manifest.

In respect of *Weight*, Gilbert's curves show a somewhat quicker arrival at maturity, no improvement being evident after about thirteen years old.¹ The present experiments quite concorded in that the younger children were almost equal to the older ones and both were not far from adults. In Old Age, we once more find no appreciable loss of power.

Finally, we come to the effect of Age upon Intelligence. On this matter general opinion seems to be sharply divided; for some assert that, other things equal, a child must necessarily get cleverer as he grows older; while others hold Intelligence to be equally manifest at all ages after infancy and to be easily distinguishable from the gradually amassing stock of acquirements and proficiencies. As before, in default of any positive antecedent data the present experiments could only keep the point in view and endeavor themselves to throw light upon the question.

5. Irrelevancies from Sex.

This factor evidently requires consideration, seeing that in some of the previous experimentation males and females have been thrown together without apology, while at other times these have been regarded as obviously heterogeneous to one another.

Commencing with children, the only important evidence concerning *Sight* and *Weight* seems to be that of Gilbert,² who comes to the conclusion that the sexes not only differ but that this disagreement is a quantity perpetually fluctuating from year to year according to the various phases of growth and especially of puberty. As regards *Pitch*, the chief testimony is

¹ Stud. Yale Psych. Lab., 1894, p. 59.

² *Ibidem*, p. 57.

that of Seashore, who declares boys and girls to present not the least appreciable difference.

Now, it is very improbable that the various senses should really be so unlike in such a respect, and accordingly it was not surprising to find that in the present experiments the village boys and girls showed quite analogous differences in Pitch also. Perhaps Seashore's results may be reconciled in the same way as before, that is, by considering that his children appear to have been better educated generally and therefore to have passed beyond the perturbations characteristic of the growing phases.

With respect to the influence of Sex upon adults, Seashore's women present a slightly better discrimination, but the difference is only such as may well be attributed to their more general study of music at the age in question. My own results coincided, in that both sexes appeared perfectly equal in all three senses tested; no support whatever was given to the popular assertion that men are much superior.

6. *The Elimination of these Irrelevancies.*

Thus we have come upon a considerable number of factors which evidently disturb the relations that we are about to investigate; and the more we sharpen our criticism, the smaller aberrations shall we be able to bring to light, until the latter begin to appear infinitely numerous and hopelessly unavoidable; we successively encounter illegitimate deviations proceeding from hour of test, from temperature, from fatigue, from state of health, from fullness of stomach, from habitual and occasional consumption of alcohol and caffeine, and so on without end. Eventually, our experimentation will arrive at the condition of the hysterical person who has found out some or other medical objection to every description of food—and so dies of self-starvation. To undertake any investigation at all, our attention cannot be confined to the detection of these impurities, but must take the further step of ascertaining which of them we can and cannot afford to neglect in practice. With this view, we have in every case insisted upon obtaining *quantitative* estimates of the differences actually produced by the various intruders, and we have found them to vary from the extravagant proportion of 1:60 down to inappreciability; from these general computations, it will usually not be hard approximately to pick out the amount applicable to any particular experiments.

So far, however, our information is still almost worthless. For the perturbation into which we are inquiring does not depend simply from the above average difference produced in the measurements by the intruder, but upon the *ratio* of this special

difference to the total average difference found in an assemblage of individuals from both the disparate classes; this latter difference will clearly be the greater of the two, seeing that it derives from many other differentiating causes in addition to the special one; the ratio will very approximately present the correlation between the disturbant and the term disturbed.¹ Even yet, though we have thus obtained all the requisite data, we cannot utilize them to measure the perturbance until we have further obtained a mathematical equation capable of performing this office; this will be given at the end of the next chapter; by its aid and that of the foregoing inquiry, we are finally enabled to decide which of the irrelevancies produce a falsification of appreciable size as compared with the probable error; all of less magnitude may be left out of consideration, as their elimination would not be of the slightest practical advantage.

To avoid more tediousness, we will here take only two sample cases to illustrate how the decision is actually reached as to whether an irrelevant factor be really of formidable nature. In tests of Pitch, say, it is desired to know whether the reagents may properly consist of non-musicians, mixed with musicians. Our Table (page 231) shows us that the former have an average threshold about twice the size of that of the latter; next, on taking a number of thresholds of both non-musicians and musicians thrown together, we find that the ratio between the averages of the worse and better halves respectively comes to something very near 3:1 (see any of the quoted works); hence the irrelevant correlation is about $2/3$ or 0.62. Applying the appropriate formula, we find that under such circumstances a real correspondence of say 0.50 would be unduly deflected by about 0.14. As the experiments were being designed on such a scale as to give a probable error of only about 0.03, the perturbance must be pronounced very appreciably falsifying.

Let us consider the analogous case as regards the sense of Sight. Here also, it may safely be assumed that diversity of daily occupation among the reagents will have led some more than others to practise this particular faculty, and further that the extent of this discrepancy will on the whole certainly be less than that between musicians and non-musicians (seeing that very few people spend hours a day in discriminating small variations of light and shade); again, the general effect of practice in this sense has been found to be about one third of that in Pitch (p. 232), while the total average difference turns out to be about the same as for the latter; hence this time, the irrelevant correlation will be at any rate much less than one

¹See "The Proof and Measurement of Association between Two Things." *This Journal*, Vol. XV, 1904, pp. 72-101.

third of 0.62. On once more applying the suitable corrective formula, we find that any irrelevant correlation of 0.21 would falsify a real correspondence of, say, 0.50 by an amount of only 0.01. Thus we have ascertained that the deviation caused by difference of Practice is in this sense not likely to affect our result by more than a very small fraction of the probable error and may well be neglected.

In this way, a large number of irrelevant factors may be summarily dismissed from consideration; others, either from their amount or from the doubtfulness of the evidence concerning them, must be carefully kept in view; while some will be found of an unquestionably falsifying nature. The first employment of our knowledge is so to order our experiments from the beginning, both as regards selection of reagents and procedure in testing, that appreciable perturbation may be escaped. Next, and yet more important, we have learned the directions from which serious danger is to be apprehended, and forewarned is forearmed; we are enabled to so conduct our work, that the residuum of deviations which could not be excluded actually may instead be afterwards eliminated with mathematical exactness from the final results.

7. *Alternatives and Equivocalities.*

Notwithstanding the exhaustive selective process which will have had to be executed in order to sufficiently eliminate all these irrelevant factors, there will generally still remain over a certain number of experimental methods that at first sight appear almost equally eligible.

As regards Intelligence, to begin with, there is a considerable variety of ways by which estimates can be gained, and most people would appear to enjoy very decided views as to which are the most satisfactory; but since the convincement of their opinion usually stands in inverse proportion to its evidential value, it is perhaps safer to begin with the methods nearest to hand and gradually discover for ourselves their relative advantages. In doing so, two points must always be carefully kept asunder: first, the *reliability* with which any system of measurement represents any particular form of intelligence; and secondly, the claims of the said form of intelligence to merit the name. The former point must be definitely ascertained in the course of the experiments, while the latter, though a very desirable piece of information, may or may not be eventually elucidated by the whole investigation.

Sensory Discrimination will usually offer a still greater field of choice. About the mode of interrogation, it here need only be mentioned that the present wants are not necessarily identical with those of other psychological work. For instance, it is

well known that the sensory threshold has really two different values, according as the examination proceeds from greater to less differences or *vice versa*; now, for questions specially concerning this threshold as such, it is desirable to take the mean between the upper and lower values; for the present purpose, on the other hand, more regular results appear obtainable in a given time by determining only the lower threshold (the reason seems to be that the reagents become less confused when the changes are confined to the one direction).

Then comes the question as to what precise portion out of the whole sensory range should be chosen as the theatre of our experiments. For our pioneering work, at any rate, it would appear unquestionably best to commence with that part where the results are least exposed to irregularities, complications and unknown factors; this will always lead us to somewhere near the centre of the range.

Lastly, we must consider the influence of different apparatus. It is a much more prominent factor in some senses than in others. In Pitch, it appears to be minimal; highly trained observers, however variously tested, show little more diversity than can be satisfactorily accounted for by individual disposition. Delezenne,¹ Seebeck,² Weber,³ Preyer,⁴ Appunn,⁵ Stumpf,⁶ Luft,⁷ and Meyer,⁸ though using such widely differing instruments as monochords, reed-pipes, and tuning forks, obtained results of which the largest is little more than double the smallest, and even this small difference appears chiefly attributable to difference in mode of interrogation. Among less practised reagents, indeed, sensibility will sometimes be found to vary enormously with the kind of timbre, one person doing much better when there are no appreciable overtones and another *vice versa*; but still such discrepancy is generally traceable to diversity of previous habit and continually diminishes with further training. In Visual Discrimination, the results have been considerably less uniform. It is true that the majority of competent observers, Kraepelin,⁹ Volkmann,¹⁰ Aubert,¹¹ Masson,¹² Merkel,¹³ etc., though employing such various in-

¹ Recueil des travaux de la soc. de Lille, 1827, p. 4.

² Poggendorf's Annalen, Vol. LXVIII.

³ Pogg., XIV.

⁴ Die Grenze der Tonwahrnehmung.

⁵ *Ibidem*.

⁶ See ref. to Meyer, note 8.

⁷ Wundt's Phil. Studien, Vol. IV.

⁸ Zeit. f. Psych. u. Phys., Vol. XVI.

⁹ Wundt's Phil. Stud., Vol. II.

¹⁰ Fechner's Psychophysik, Vol. I, p. 148.

¹¹ Physiologie der Netzhaut, 1865, p. 58.

¹² Ann. chim. phys., 1843.

¹³ Wundt's Phil. Stud., Vol. IV, p. 566.

struments as Masson's disc by daylight, the same by artificial light, the shadow-method, the episcotister, and illuminated matt glass—all agree that the most favorable medium shades of gray can just be distinguished from one another when the illumination differs by about 1/120. But on the other hand König with Brodhun,¹ Helmholtz,² and Bouguer³ arrive at a threshold twice as big, while Schirmer⁴ and Volkmann (in his later experiments) come to one nearly twice as small. That, however, the lion's share of these dissimilarities may be solely attributed to difference of instrument is shown by the largest variation occurring in one and the same person: Simon, when using the platinum lamp of König, has a still worse threshold than the latter; but when employing a disc like Schirmer, surpasses even this observer. The discordance assumes far greater proportions when we come to Discrimination of Weight: for widely apart are the values successively obtained by Weber,⁵ Fechner,⁶ Biedermann and Loewit,⁷ Hitzig,⁸ Jacoby,⁹ Merkel,¹⁰ Müller and Schumann,¹¹ Martin and Müller,¹² etc.; on the one extreme we have Biedermann and Loewit who could clearly perceive a difference in a 2 lb. weight if it were increased or diminished by so little as $\frac{1}{2}$ oz.; on the other side, we find C. Jacoby whose five thoroughly trained and competent reagents required for the same purpose an increase or reduction *twenty-four* times greater, a discrepancy that for the main part may safely be attributed to difference of apparatus.

This multitude of gross fluctuations requires careful analysis. A certain number merely imply greater or less difficulty in the performance and consequently a simultaneous rise or decline for the generality of reagents. Others, again, derive from difference of individual disposition; the larger these are in any procedure, the more accurately will any functional correlation be able to be calculated; but they must not be too summarily confounded together, for possibly one procedure may exhibit greatest sensitivity to previous practice, while another may be the most delicate barometer of fatigue or ill health, and yet a

¹ Sitz. Ber. d. Berl. Akad., 26-7-1888 and 26-6-1889.

² Physiol. Optik, p. 315.

³ Aubert, Phys. Optik, 488.

⁴ Arch. f. Ophthalm., Vol. XXXVI, p. 121.

⁵ Tastsinn u. Gemeingefühl.

⁶ Psychophysik, Vol. I.

⁷ Berichte d. kais. Akademie d. Wissenschaft. zu Wien, Math. Naturw. Classe, Vol. LXXII.

⁸ Neurol. Centralblatt, 1888.

⁹ Arch. f. Path. u. Pharmak., 1893.

¹⁰ Phil. Studien, V.

¹¹ Pfüger's Arch., XLV.

¹² Zur Analyse der Unterschiedsempfindlichkeit.

third may reveal the largest correlation with intellect. The remaining fluctuations derive from causes so obscure, that for the present we must call them chance irregularities tending only to shroud and disguise our results.

From these considerations, it is clear that an alteration in method of experiment may result in transforming the correlation. But this further complication, far from being a real disadvantage, opens out a rich and virgin field to investigation, for it is the most promising means of penetrating into the correlation's essential nature. Eventually results will be gained of such greater generality, that they will stand out constant among these superficial changes, as the main features of a landscape abide immovable in the passing variety of cloud and sunlight.

CHAPTER IV.

DESCRIPTION OF THE PRESENT EXPERIMENTS.

I. *The Choice of Laboratory Psychics.*

We will now turn to a description of the experiments that form the basis of the present article. Their many deficiencies can scarcely be clearer to any one than to their author; so true is it that we first learn how properly to conduct any experiment—when we have ended it.

As regards the nature of the selected Laboratory Psychics, the guiding principle has been the opposite to that of Binet and Ebbinghaus. The practical advantages proffered by their more complex mental operations have been unreservedly rejected in favor of the theoretical gain promised by utmost simplicity and unequivocality; there has been no search after condensed psychological extracts to be on occasion conveniently substituted for regular examinations; regardless of all useful application, that form of psychical activity has been chosen which introspectively appeared to me as the simplest and yet pre-eminently intellective. This is the act of distinguishing one sensation from another.

With respect to the particular senses preferred, the present experiments have been confined to Hearing, Sight, and Touch. The other five, Taste, Smell, Pain, Heat, and Cold, do not admit of such practicable or satisfactory examination; also, probably on this account, they have as yet been investigated very incompletely, and therefore do not form a good unequivocal foundation for research of more advanced order. Further, in the chosen three we have already the widest range of type: for Touch is the most direct of the senses, the physiological organ being apparently of such a simple structure as to convey

the stimulus to the brain in a purely mechanical manner; Sight, on the other hand, offers the most perfect example of peripheral transformation, seeing that our visual presentations are but very remotely derived from the really external ether waves with which they are popularly confused; while Sound gives us a half-way stage between the above extremes.

In all three cases, the test has been of relative, not of so-called "absolute," Discrimination; the trial has not been as to how small an external stimulus can cause a sensation perceptible at all, but as to how great the difference in the external stimuli has to be for the reagent to notice any difference in the sensation; every one knows the uncertainties attending the former kind of investigation. Similar motives have in Sound made Pitch seem preferable to Intensity; and in Light, Luminosity to Color. It is perhaps less easy to justify Touch being represented by that form of it often termed the "muscular sense," which despite its notoriety and historical importance is now well known to be really most complex and obscure.¹ Among other reasons for this choice, it was desired to see whether any correlation of Discrimination with Intelligence might not reasonably be attributed to adroitness in outward approach to the distinction rather than to superiority in the essential act itself; should this be the case, then the correspondence should be much more manifest in an active practical comparison of two weights than in the purely passive acceptance of two tones.

In short, the experiments were so chosen that any conclusions very uniformly attained in these three ways might be provisionally considered to hold good for Sensory Discrimination in general. —

It may here be remarked that the present experiments were not primarily designed and executed for the sake of Discrimination itself, but as the indispensable basis for an investigation into Memory and Imagery. Only afterwards was it decided to publish the results on the former head separately, so as to keep the present article within manageable length and at the same time to secure for it greater unity and lucidity.

2. *The Instruments.*

The apparatus with which these inquiries were conducted were not of a very new or elaborate nature, my object at present being merely to make a practical and convenient application of the existing laboratory methods. Moreover, while naturally

¹ It is here by no means intended to assent to the identification of the feeling of Effort with any combination of touch sensations, but only to admit that the former is in practice chiefly estimated through the mediation of the latter.

endeavoring to minimize all errors proceeding from mechanical imperfection, I have become quite convinced that faults arising from such a cause are in this case vanishingly small as compared with those proceeding from other sources.

(a) *Sound.* Like the experiments of Delezenne and Cattell mine were conducted by means of a monochord, but of a special kind made under my direction for this purpose; it is furnished with a Vernier scale, whereby differences of pitch can readily be produced down to $\frac{1}{3}$ of a vibration; as the experiments were limited to the immediate neighborhood of the E above the middle C (reckoning the latter at 330 vibrations per second), the smallest securely obtainable difference amounted to $1/132$ of a musical tone.

To the accuracy of such an instrument many theoretical objections may be raised, as unevenness of the wire, inequalities of tension, etc. However, experience has shown it to be reliable to the range above mentioned, which is ample for the present purpose. It was first tested in the usual way, that is, by simultaneously sounding two notes at a given small interval and then counting the interferences in the sound-waves; I had no difficulty in producing good reliable "beats." The instrument has further been well checked empirically, for more than one reagent has proved able correctly to discriminate down to the extreme limit. I have also been fortunate enough to obtain assistance in this manner from Dr. Krüger; this psychologist and acoustician very kindly allowed himself to be tested with my machine and found his threshold to be only such a small amount greater than with tuning forks, as would naturally arise from his having had much more practice with the latter instrument.

The principal crime generally charged against monochords is that they give far stronger overtones than do tuning forks and that consequently tests upon the former may not be comparable with those upon the latter. This point has already been discussed, and our conclusion was that all the different instruments hitherto tried had led to very similar *average* thresholds (in this respect being very unlike discrimination in Sight and Weight). In any case, there is no reason whatever to suppose that the results given by the overtoned instruments are at all less regular, and this is here the only matter of importance.

As a set-off against this one merit of comparative freedom from overtones, tuning forks are in so many other respects difficult to manage, that I should not personally prefer them except for use in an exceptionally well equipped laboratory. To bring them to sound, they have to be either struck with a hammer or stroked with a bow, and the consequent accessory

Noises and tones can hardly be reduced to such a minimum as is easily attainable in plucking the monochord; also, the tone from the fork only remains constant for a very short time, after which it rises very deceptively; further, it is very difficult to obtain two tones of sufficiently equal loudness, without using a special striking apparatus which again involves other disadvantages. Moreover, it is almost impossible to obtain enough range of tone in any satisfactory manner, especially as even the best tuning forks have appreciable overtones unless placed upon resonators to preferentially intensify the ground-tone; the device of a different fork for each grade of pitch is of very dubious merits; and when the common custom is followed of using one fork with the sliding regulator on it while the other has none, the overtones become so dissimilar as frequently to vitiate the whole experiment. The last objection can be partially met by having on each fork a separate regulator very light and low; but the first and second difficulties are very insufficiently overcome by the usual expedients, such as moving the fork to and from the reagent's ear; to fully obviate them, it is at least necessary to be able to conduct the sound from one room to another, so that the experimenter remains with his forks and opens the conductor just when the tone is pure and true, while the reagent sits alone and hears only what he is intended to hear; but such a procedure is quite inapplicable to the inexperienced subjects requisite for experiments like the present ones.

(b) *Light.* As we have seen, a great number of instruments have been at different times used for this purpose. Each of them possesses various advantages and disadvantages of its own. For my purpose, I constructed a graduated series of cards, each being slightly darker than the preceding one. The required delicacy was obtained by photographic means, on the principle that the darkness of a print will (for a certain short range of half-tints) vary almost proportionately to the time of exposure to light. Various ways of printing were tried; silver gave a very fine gradation, but necessarily introduced the disturbing element of color; carbon was found difficult to develop with sufficient evenness; final preference was given to platinum on the smoothest possible paper, and a series of prints were obtained of such differences that the extreme ranges would measure the dullest normal sight, while two neighboring cards could not be distinguished by the acutest vision.¹ The prints, each about 2" by 1", were then mounted on smooth cardboard, num-

¹ This method of constructing cards as visual tests seems to have been very minutely anticipated as long ago as 1896 by Marbe (see Zeits. f. Psych. u. Physio., Vol XII, p. 62).

bered on the back, and given a border about one inch wide of even black, so that the conditions of contrast should in all cases be approximately equal.

The evenness of the series was then tested practically on many reagents, by trying whether all pairs of cards at the same number of grades apart presented precisely the same difficulty of distinguishment.

As to the illuminating source, it unfortunately was not found practicable to conduct the experiment by the uniformity of artificial lighting. However, small differences of the absolute intensity of the illumination are known not to appreciably affect accuracy of judgment, and such was found to be actually the case as regards the inevitable minor fluctuations of daylight. The most important point was that the two cards to be compared should be illuminated to exactly the same degree. This was sought by placing them side by side on precisely the same plane and opposite the centre of an evenly lighted window. Most reagents had a marked inclination to consider one particular side darker; in some cases this was the right, and in others the left (I could not find that this tendency corresponded with right and left handedness respectively, as might perhaps have been expected in view of the experiments of Van Biervliet¹).

The general luminosity of the cards was determined in the following manner. The lightest and darkest were placed in a dark room, and two candles were arranged in front of them in such a way that each threw a circular ray of light about one inch in diameter upon one of the cards. It was then experimentally ascertained how much closer to the illuminating source the darker card had to be, for the latter to appear just as light as the other. The square of the greater distance divided by the square of the lesser thus gave a measure of the relative brightness of the two cards under equal conditions. It would be superfluous here to enter into the various details adopted to make this test satisfactory, seeing that the result is at best only roughly approximate; for, among other things, the luminosity was thus tested by artificial light, while the discrimination was tried by the different luminosity and tint of daylight. Within the actual limits, however, it does not appear that the consequent error will be very considerable; hence, the value found seems sufficiently accurate to be of some interest, especially as it is difficult anywhere to discover quantitative estimates of the normal unpractised person's power to distinguish differences of light.

(c) *Weight.* For this test a graduated series of weights were

¹L'asymétrie sensorielle, Bulletins de l'Académie des Sciences de Bruxelles, Vol. XXXIV, Série III.

constructed on Galton's convenient cartridge pattern,¹ all of precisely the same size, appearance, and balance; the lightest was 1,000 grains, and the others continually increased their weight in geometrical proportion. I found it necessary to considerably extend the range of differences; for some of my reagents could accurately distinguish Galton's finest degree, namely 1/100; while many perfectly normal persons showed an obtuseness that could not be measured even on the extremes recommended by him as being adapted for "morbid" cases.

3. *Modes of Procedure.*

(a) *Experimental Series I.* The reagents were the twenty-four oldest children of a village school in Berkshire, taken without any selection; they were tested in Light, Weight, and Sound. This school was particularly favorable for my purpose, as it was within 100 yards of my own house; all the children and their families resided in the immediate neighborhood, so that I could easily obtain any information concerning them; the rector and schoolmaster most obligingly gave their valuable co-operation, for which I hereby tender hearty thanks.

Each child was separately interviewed in my house, on a different day for each different sense. The test of Discrimination lasted fifteen minutes (and was then followed by test of Memory; the coincidence between these two is a good measure of the accuracy of both). No single trial was ever unusually prolonged, for fear of admitting the disturbing effects of fatigue, but further interviews were obtained when any doubtful points seemed to require clearing.

As regards the manner of interrogation, there was little hesitation in rejecting the method known as "minimal changes" in its purer form, which is dependent upon the reagent saying whether or not he can distinguish the difference between the two sensations offered to him; such a procedure appeared totally unfitted for such inexperienced persons. But still less applicable appeared the strict method of "right and wrong cases." The compromise was therefore adopted of searching for that threshold where the subject seemed able to give about 80% of his answers right. For similar reasons, preference had to be given to the "procedure with half knowledge" (halbwissenschaftliches Verfahren) in spite of certain disadvantages;² in this, the reagent is informed that the two stimuli are different, but is left to decide for himself as to the direction of the difference.

The beginning of the test is not devoted to recording the largest possible number of answers, but to quietly affording

¹Inquiries into Human Faculty, Appendix.

²See Kämpfe, Wundt's Phil. Stud., Vol. VIII.

the reagent a maximum of fore-exercise and at the same time to gaining a general idea of his threshold. Then, there is a steady progression from greater to smaller intervals, until eventually a threshold is found where he can just give eight right answers out of ten. He is further tried at a still smaller interval, to see if he makes still more mistakes and thus to confirm the fact that he has really reached his limit. And finally he is once more tried at a slightly larger interval than the believed threshold, to corroborate the former observation that he here makes less than two errors out of ten. This constant progression in only one direction appears to very much reduce the mental distraction especially inherent in all procedure "without knowledge," which is very great in unpractised reagents if tested with the usual oscillations to and fro between greater and smaller intervals. Against it may be urged that it finds only the lower instead of the mean threshold, but this is of no importance for our present purpose.

It was further considered that more regularity would be obtained by only recording those answers which were given under the most favorable conditions. Before taking down each reply, a chance of reconsideration was given by repeating the test in such a manner as to reverse the constant error of time and space.

The two stimuli followed one another in the manner found to be most adapted to accurate judgment and to effectually eliminating the influence of Memory. In the case of pitch, the interval from the beginning of the first tone to that of the second was found best at about three-quarters of a second. In some earlier experiments, the first tone was just dying away when the other began; but later, I stopped the first one altogether just before giving the other. There is some knack required to do this satisfactorily, and the practical effect of the change was, to my own surprise, inappreciable with most reagents.

The final measurements, just as they were obtained, are set down in the column under "Sensory Threshold" in Table I.

(b) *Experimental Series II.* This was executed in the same village school, but upon the next thirty-six oldest children, the tests being only in sound. Unlike the previous twenty-four, these were examined collectively, the total interview lasting about $1\frac{1}{2}$ hours. The chief part of this time was devoted to instructing and practising them, and to finding out what was the lowest age fit for such a collective experiment. It became eventually evident that no usable results could be obtained at any rate from those below $5\frac{1}{2}$ years, and thereupon all those under this age were excluded, leaving thirty-six for the real tests.

The latter were carried out in the following manner. Every

boy and girl was provided with a pencil and a piece of paper and had simply to write down 1 or 2, accordingly as he considered the first or the second tone to be the higher. The headmaster as well as the other teachers were present; a small prize was offered to stimulate attention, and energetic measures were found necessary to prevent cribbing. Ten test pairs of tones were given at about the following eight differences of pitch, 50, 33, 26, 20, 16, 10, 6, and 3 v. d.; thus in all, each child answered 80 times; in half the cases, the first tone was really the higher, and *vice versa*.

The marking was done by fully considering each child's whole paper, and then deciding as to what was the limit at which he might be expected to give about eight right answers out of ten. This method has seemed to me in such cases the most satisfactory, provided that the person marking has acquired the requisite experience by having previously examined in a thorough manner a great number of similar reagents, and provided the marking of the paper be done before receiving the intellective gradings (or else there is a great danger of "self-suggestion"). If this method be not adopted, recourse must be had to some purely formula reckoning of errors.

Despite all precautions to secure reliable results, I was unable to quite convince myself that such uncultured children could be treated adequately without elaborate individual attention.

(c) *Experimental Series III.* These experiments — confined to Sight and Weight — were made in a preparatory school of the highest class, which principally trained boys for Harrow. To the Principals, themselves old Harrovians, my hearty thanks are due for their kindness and cordial co-operation. As may well be imagined, the social standing and general culture of the reagents were the opposite extreme to that in the village school.

Unfortunately, these tests of Light and Weight had to be arranged at a few hours' notice and consequently were carried out under very unfavorable conditions. Out of the thirty-seven boys constituting the school, only twenty-four could be present, and of these again it was necessary to withdraw one from the results as being mentally too abnormal to be properly included with the others. In the next place, no masters were in the room after the first few minutes; in spite of the general excellent behavior, this relaxation of discipline must always be admitted to be a momentous circumstance. And finally, the visual and muscular senses—never well adapted to collective examination—had on this occasion to be tested with the apparatus intended only for individual work: the weights and the cards were continually passed round and round by pairs, ticketed 1 and 2 respectively; each boy then wrote down which of the two he considered to be the heavier or darker.

It was impossible to control whether they all handled the apparatus in precisely the same manner, or even to insure that they invariably gave their greatest possible attention to their tasks; moreover, some were inevitably more favored than others as regards intensity and evenness of illumination.

Owing to these facts, I brought away the impression that the experiments were chiefly interesting as enhancing the effect of *Goodwill*. For under better conditions, every reagent can be brought pretty well to try his best; but here, on the contrary, a wide range was observable in this respect, some being very zealous while others were visibly indifferent.

(d) *Experimental Series IV.* This series, which was in Sound only, took place in the same high class preparatory school, but now the circumstances were as propitious as above they were the reverse. Pitch is a sensory quality especially susceptible of collective test. The experiments were arranged and prepared with full deliberation. The entire school were available with only the exception of the above mentioned abnormal case, of one boy who had to leave the room before the end of the hour, and two who were that day absent; thus there were thirty-three complete results. Several masters attended, so that the strictest discipline was maintained throughout; there appeared no inclination to crib; every boy seemed perfectly to understand what was required and to be intent upon doing as well as possible. I am therefore inclined to attach as much value to this series as to Series I; for the cultured intelligence and long habit of examination possessed by these boys should compensate the individual attention given to the villagers.

The conditions entirely resembled those described in Series II, except that only 48 final tests were made, 6 at each of the following differences: 20, 15, 11, 8, 6, 5, 3, and 1 v. d. The intellective grading was not received by me until long after the sensory grading had been completed, so that the latter is free from danger of self-suggestion.

(e) *Experimental Series V.* These were executed upon 26 male and female adults (thus bringing the total number of reagents throughout the present experiments to 123).

The method was individual and precisely the same as that already described for Series I. But in arranging the composition of the reagents, instead of trying to obtain as homogeneous a set as possible, it was here rather sought to include the greatest variety; for although little can be proved in such a manner, much can profitably be suggested.

4. *The Estimation of "Intelligence."*

As regards the delicate matter of estimating "Intelligence,"

the guiding principle has been not to make any a priori assumptions as to what kind of mental activity may be thus termed with greatest propriety. Provisionally, at any rate, the aim was empirically to examine all the various abilities having any *prima facie* claims to such title, ascertaining their relations to one another and to other functions.

Four such different kinds of Intelligence have been introduced into the present work. First, there is that revealed in the ordinary classification according to school order (based here upon examinations). This clearly represents *Present Efficiency* in such matters as Latin, Greek, Mathematics, etc. Examples of this kind will be found in experimental series III and IV.

The next sort of Intelligence derives from the same school order, but so modified as to exclude all influence of Age. Such a corrected order may be provisionally accepted as representing, not Proficiency, but *Native Capacity*. It has been arrived at by taking the difference between each boy's rank in school and his rank in age. For obvious reasons, it has been preferred to consider the absolute and not the relative differences; a boy, for instance, who was 20th by examination and 22nd by age would be placed just above one who was 15th by examination and 16th by age, the former being two places and the latter only one better than would have been expected with greatest probability.

The resulting order is clearly but a first approximation, to which we may apply any number of further corrections. For our present purpose, the following has appeared the most that can be practically required (and even this makes no appreciable change in the final values obtained). Evidently, the top boy is prevented from proving his full capabilities by want of competitors; let us suppose that he happens also to be the oldest; then, on our above method, he will seem no better than a boy of middle age and at the same time of middle school order; but the latter will in reality always be found below many younger than himself, compensating this by being also above about an equal number of older ones; now, our top boy has not let himself be surpassed by any single one of his juniors, and therefore would certainly have gone above a great many of his seniors, had the school included such. The top boy's true position may be roughly estimated by making him an extra allowance of a number of places equal to the general mean deviation of actual from average rank (which in this case comes to 5 places); clearly, also, such allowance may with equal right be claimed by the top boy, even if he does not happen to be the oldest; further, the same correction is applicable in slighter degree to the second boy, in still slighter to the third, and so on in a

rapidly diminishing curve up to the centre of the school. For practical purposes, it has seemed sufficient to allow the next four boys, 4, 3, 2, and 1 places respectively; naturally, the whole of this correction must be repeated inversely for the bottom end of the school. Though this explanation is rather complicated, the correction is very easily carried out and, as stated, its effect is hardly appreciable.

The third kind of Intelligence is that represented and measurable by the general impression produced upon other people. This forms the basis of the common broad assortment of the children by their teachers into "*bright*," "*average*," "*dull*" respectively; and with such an assortment I have had to content myself for the elder children in the second series of experiments, while for those under 7 years of age, I have not obtained any intellective grading at all.¹ But for the more important Series I, a list of relative rank was procured of satisfactory completeness. It may here be noted that teachers, if directly asked for such a detailed list, frequently begin by asserting it to be impracticable. It will be generally found, however, that if they be merely requested to pick out the brightest pupil of all, they can do so without any great trouble; and when they are next requested to select the brightest of the remainder, they are still able to perform the desired feat; and so on, until the classification is complete.

The fourth and last sort of Intelligence which has here been estimated is that known as *common sense*. To this end, the oldest of the children of Series I was interviewed and interrogated concerning her comrades in precisely the manner described above, except that the criterion was not to be "*brightness at school work*" but "*sharpness and common sense out of school!*" and she seemed to have no great difficulty in forming her judgments concerning the others, having, indeed, known them all her life. As a check, and in order to eliminate undue partialities, it had been arranged that as she left the house, the second oldest child should enter it and thus be able to give an as far as possible independent list, since neither had beforehand had any idea of what was wanted. Finally, a similar list was obtained from the Rector's wife, who also had always lived in this village; but her graduation is unfortunately incomplete and therefore unusable, for she professed inability to pronounce verdict upon some few children who had not come much under her notice; as far as it went, it appeared perfectly homologous with the other two lists.

¹ The day after these experiments I left the neighborhood of the school.

5. Procedure in Deducing Results.

(a) *Method of Correlation.* So far this chapter has been occupied with obtaining estimates as to the reagents' respective abilities in the several sensory and intellective functions. This is an operation requiring the fullest use of psychological insight; and, therefore, based on the long preliminary investigation previously described, every effort has been made to ferret out and evade all circumstances tending to make our little sample of facts appreciably misrepresentative of the real general relations or psychologically superficial and misleading. But the next portion of our problem is of a very definite objective nature; we wish to ascertain how far the observed ranks in the several abilities tend to correspond with one another; this, it is believed, is no longer a task to be effected by exertions of psychological ingenuity; instead of constructing complex arbitrary tables and plausible but more or less fanciful explanatory stories, we now are in need of such a procedure as will impartially utilize all our information in the *demonstrably most complete manner* and will focus it to a plain quantitative value; for the moment, psychology has to give way to mathematics.

Accordingly, all the more important correlations in the present work have been worked out by the best method hitherto evolved, that of "product moments," as Pearson terms it; only instead of using the actual measurements obtained for the reagents' respective thresholds, the change has been made of employing the numbers denoting their relative ranks; a full explanation of the advantages of this modification may be found in the article specially devoted to the topic (the chief being a reduction of the probable error equivalent to doubling the quantity of cases observed). Merely subsidiary results have often been reckoned by the much more convenient method of "rank differences," while a few correlations were for various reasons not amenable to either of these more exact methods, and therefore had to be worked out by Pearson's auxiliary method or mine of "class averages" (the latter has generally been preferred, on account of its smaller probable error). All these are but different ways of more or less closely arriving at the same measure of correlation, and thus all the results can be freely compared with one another.

The method of "product moments," though sometimes involving lengthy calculations, is so simple in principle that it can be worked by any moderately intelligent schoolboy. Explanation and illustration are given in the above article; here, nothing more than the general formula can be stated, which is as follows:

$$r = \frac{S_{xy}}{\sqrt{S_x^2 \cdot S_y^2}}$$

where x = any individual deviation from the general median as regards one of the compared characteristics,

y = the deviation of the same individual as regards the other characteristic,

Sxy = the sum of such products for all the individuals,

Sx^2 = the sum of the squares of all the various values of x ,

Sy^2 = the same for y ,

and r = the required correlation.

(b) *Elimination of Observational Errors.* This necessitates a further mathematical operation, which, however, is very brief and does not involve anything more than elementary arithmetic. There are two formulæ, one theoretical and the other empirical:

$$1) \quad r_{pq} = \frac{r_{p'q'}}{\sqrt{r_{p'1}r_{p'2} \cdot r_{q'1}r_{q'2}}}$$

$$\text{and } 2) \quad r_{pq} = \frac{\sqrt[4]{mn \cdot r_{p''q''} - r_{p'q'}}}{\sqrt[4]{mn - 1}}$$

where $r_{p'q'}$ = the mean correlation between the various gradings for p and those for q ,

$r_{p'1}r_{p'2}$ = the average correlation between one and another of these several independently obtained series of values for p ,

$r_{q'1}r_{q'2}$ = the same as regards q ,

$r_{p''q''}$ = the correlation of an amalgamated series of measurements for p with an amalgamated series for q ,

m and n = the number of independent gradings for p and q respectively,

and r_{pq} = the required real correlation between the true objective values of p and q .

It will be found exceedingly important to employ both formulæ simultaneously, for they are independent of one another and each has different sources of fallacy, so that the most essential information is gained by a comparison between their respective results.

When we say that a series of objects correlates entirely with a second series, we do not assert that every set of measurements of the one will absolutely coincide with those of the other, seeing that discrepancies must inevitably arise from errors in measuring; we only mean that whatever all sets of measurements of the one series have in common with each other will also be found common to all measurements of the other series; then, either of the above formulæ will exactly eliminate the observa-

tional discrepancies and thus present the correlation in its entirety.

But much more often the measurements for the same series are connected with one another by *more* than connects them with the measurements of the other series, and then the case is ambiguous. Either the surplus really lies in the series measured, which is equivalent to saying that this series contains elements not common to the other series and that the correlation is to this extent incomplete; here, once more, both formulæ will produce the properly corrected amount. Or, as is usual, the excess of agreement between the measurements for the same series may partly (or wholly) derive from their having the same constant fallacies; and now it will be found that both formulæ give a correction still in the right direction but too small in quantity; further, this deficiency will be much greater for the theoretical formula than for the empirical one, so that *when both formulæ give the same result, we can assume that the latter has not been appreciably falsified by any constant fallacy common to the several sets of measurements for the same series.*

Under special circumstances, the contrary case may occur of the sets of measurements for the one series being connected with each other by *less* than connects them with those of the other series. This will happen whenever several sets of measurements supposed to be taken from the same lot of objects are really procured from different ones and their several correspondences with the second series have arisen from independent causes. In physical matters, this danger is not serious; if two persons decide independently to measure a fossil cave-bear, they are unlikely to make the mistake of going to different animals. But in psychology it is otherwise; persons may honestly endeavor to appraise the same mental faculty, and yet, owing to diversity of procedure and ignorance of organic uniformities, they may really obtain measurements of quite independent function. In such case, the sets of measurement, however accurate they may be, will show no correspondence with one another; and if the functions are even only partially different, the measurements will correspond with one another to that extent less than they would by reason solely of errors of observation.

The effect will be to falsify any corrections by the theoretical formula, for the latter begins by assuming only one lot of objects to have been measured and therefore the correspondence between the sets of measurements to be at least as great as might be expected from their accuracy—an assumption generally fair enough, but under the special conditions delusively reducing the denominator and thus producing a final value proportionally too large. Now, this same fallacy affects cor

rections by the empirical formula in exactly the opposite direction; for the latter bases itself upon the fact that an amalgamation of several sets of measurements constantly emphasizes whatever elements are common to them all and simultaneously obliterates all that are not common; thus in the normal case of only one lot of objects underlying the sets and determining their correlations to the other series, amalgamation will continually raise the correlation towards its full amount; but if there be more than one underlying lot of objects, each correlating with the other series independently, then amalgamation will not emphasize but obliterate these independent influences and consequently not raise but lower the correlation. Hence, when several functions really corresponding with the second series independently have been confounded together and taken for different measurements of a single correspondence, the results, as corrected by the respective formulæ will sharply diverge. Conversely, if, when a double set of measurements has been made the empirical corrective formula produces an increase of correlation, then these sets of measurements may be regarded as certainly deriving from some single common faculty (any influences specific to each set of measurements being theoretically subtracted from the faculty and viewed as merely so many sources of observational error); and if the two corrective formulæ lead to the same final amount of correlation, then this latter concerns wholly and solely the common faculty.

Further, it is of great importance to remark that the last fallacy, namely the case when measurements believed to be taken from the same function really derive from different ones correlating with the other series independently, may, by the first corrective formula, easily come to any values greater than 1 (and therefore impossible, seeing that 1 represents entirety). By the empirical formula, on the other hand, this can never occur; for whether the sets of measurements be connected with one another by either anything more or anything less than connects them with the measurements of any other compared series, then the correspondence between the two series will in both cases be reduced and therefore must necessarily be less than 1;¹ in other words, the empirically corrected correlation can only amount to full unity when all the sets of measurements for both series have one common element and differ in every other systematic constituent.

Fuller explanation and illustration are given in the article devoted to the topic of measuring correlation.

(c) *Elimination of Irrelevant Factors.* This is the final opera-

¹ Except for deviations due to mere chance, whose range will vary with the probable error.

tion necessary to obtain a true result. Unlike the preceding one, it may often be altogether escaped; for if the conditions are favorable and if the preliminary investigation has been sufficiently thorough, the experiment need not be affected by any irrelevant factor of large enough magnitude sensibly to vitiate the result. Here, also, the necessary mathematical work has been reduced to brief and elementary arithmetic; for more explanation, the reader must again be referred to the special article.

If the irrelevant factor be connected with *only one* of the two compared series, the equation is:

$$r_{pq} = \frac{r'_{pq}}{\sqrt{1 - r_{pv}^2}}$$

where r'_{pq} = the apparent correlation of p and q, the two variants to be compared,

r_{pv} = the correlation of one of the above variants with a third and irrelevantly admitted variant v,
and r_{pq} = the required real correlation between p and q, after compensating for the illegitimate influence of v.

If the irrelevant factor be connected with *both* series compared, the equation becomes :

$$r_{pq} = \frac{r'_{pq} - r_{pv} \cdot r_{qv}}{\sqrt{(1 - r_{pv}^2)(1 - r_{qv}^2)}}$$

where all the terms have the same meaning as before.

CHAPTER V.

THE PRESENT RESULTS.

I. *Method and Meaning of the Demonstration.*

As the reader will have noticed, the formulæ given at the end of the previous chapter are equations whereby from several observed correlations we are able to deduce a single true one. This latter alone is of real scientific significance, and under the ordinary unsystematic conditions — such as governed the great majority of work reviewed in the second chapter — the actually observed correlations will rarely be of much interest in their primitive raw state; for after passing through the proper corrections, they would come forth transfigured in every conceivable manner; some would increase in size, some diminish, some entirely disappear, and some even become inverted. Nevertheless, our true correlation in no way deserves the reproach of being a theoretical abstraction, for it only represents the limit to which the observed correlation itself will continually

approach as improvement is made in the experimental procedure; and not even the most perfect methodics can afford to dispense with the formulæ, seeing that these are the sole means by which the perfection may be adequately ascertained.

Our method of demonstration implies four distinct steps, all of which are believed to be absolutely indispensable for work intending to be more than merely suggestive. First, we must exactly determine the quantity of correlation actually observable, and we must compare it with the probable error; then, if the former be no more than about twice as large as the latter, the whole experiment may indeed have produced a substantial negative result, but cannot possibly warrant any positive conclusion other than to suggest the desirability of extending the investigation until it acquires more evidential value; but if, on the other hand, the observed correlation be four or five times greater than the probable error, we may then consider a *prima facie* case of correspondence to have been established and we may legitimately go on to the corrective processes so as to bring our raw figure to its most probable real amount. Accordingly, the second step will be to form an estimate of the errors in observing the two series compared; for this purpose we must have obtained two or more independent sets of measurements for each series, or at least must be acquainted with the relations found between other such sets under sufficiently similar circumstances; the influence of these errors can then be eliminated by the formulæ given on page 253, and at the same time an opinion can be formed as to the presence or not of the grave fallacies discussed on pages 253 ff. The third proceeding is to look for any factors irrelevantly admitted (or, more rarely, excluded); any suspicions must be carefully verified in succession, and, if necessary, employment must be made of the eliminating equations given on page 256. Finally, we have to critically review the whole argument, paying particular attention to such disturbing factors as have not been disposed of very satisfactorily; in this way we come to a final estimate, not only as to the most probable amount of real correspondence, but also as to the degree of confidence to which our evidence is entitled; for these two things are by no means always parallel, a high apparent correlation often having but small evidential value and *vice versa*.

A few words may now be said concerning the eventual *meaning* attachable to the result which we hope to obtain. To put it briefly, the usual direction of inquiry is in the present work reversed. The customary procedure consists in determining some matter of research subjectively, say, “Perception,” “Attention,” “Imagination,” “Fatigue,” etc., and then ascertaining its relation to other similarly pre-determined psy-

choses or neuroses. Here, on the contrary, although every effort has been made to render the mental phenomena as unequivocal and significant as possible, yet in the beginning not the least note is taken of any psychological import beyond such as is barely necessary to define the subject of discussion in the most positive and objective manner; while the structure of language necessitates the continued use of such terms as Discrimination, Faculty, Intelligence, etc., these words must be understood as implying nothing more than a bare unequivocal indication of the factual conditions of experiment. For the moment we are only inquiring how closely the values gained in the several different series coincide with one another, and all our corrections are intended to introduce greater accuracy, not fuller connotation; the subjective problems are wholly reserved for later investigation. It is no new thing thus elaborately to deal with and precisely measure things whose real nature is concealed from view; of this nature, for instance, is obviously the study of electricity, of biology, and indeed of all physical science whatever.

Let us, then, consider the extent of connection between two series of things implied by this sole fact of their presenting a numerical correlation with one another; such a correspondence, when beyond the range of mere chance coincidence, may be forthwith assumed to indicate and measure *something common to* both series in question. Such a community may often consist of a definite so-called "substance;" A's changes of wealth will show some correlation with those of B, if both possess some shares in the same stock. Or, on the other hand, the community may derive from a more complicated interaction of forces; thus, the weather is supposed to correlate with the state of the spots on the sun. But this distinction is superficial even in physical matters; thingness may well be an indispensable crutch for popular thought, and indeed in metaphysics becomes a serious enough topic, but it has no place in strict natural science and still less in psychology, where fast limit has never been securely traceable between things, qualities, and conditions.

But the same simple mathematical formulæ which have brought us so far will take us yet farther. As from several sets of inaccurate measurements it has been found possible to arrive at the accurate correlation of the two real series, so now in a similar manner from any number of real series we can proceed on to dealing exclusively and precisely with any element that may be found common to these series; from ascertaining the inter-correlations of, say, auditory discrimination, visual discrimination, the capacity for learning Greek, and that for playing the piano, we can arrive at estimating the

correspondence of whatever may be common to the first pair of faculties with whatever may be common to the second pair. By combining such correlations of higher order, it is feasible to execute any required amount of elimination and selection, so that eventually a dissociation and exactness may be introduced into psychology such as can only be compared with quantitative chemical analysis; even in the present work, it is hoped to obtain results of sufficient fineness to be independent of local conditions of experiment, and therefore to be precisely verifiable by any other workers. All the time, the relations discovered by us will wholly retain their impartial objective character; however accurately we may learn the distribution of community, it will remain as a later and very different task to detect and analyze its psychical nature. But we shall find that the successive positive ascertainment of objective relations continually reduces and simplifies the thinkable explanatory hypotheses, so that practically our method of investigation is bringing us towards the introspective psychological solution also—and perhaps in the end by the shortest route.

2. Correspondence between the Discriminations and the Intelligences.

(a) *Experimental Series I.* We will begin by dealing with the subject broadly and considering the general average correlation between the various forms of Discrimination and those of Intelligence. To establish our *prima facie* case, we note that Discrimination has been tested in three senses and that Intelligence has been graded by three different persons; thus we have nine correlations which, if no correspondence exist, should all be small (about half of them under 0.09) and approximately as many should be inverse as direct. Far from this being the case, we find that every single one is direct, that the smallest amounts to 0.25, and that the average comes to 0.38 with a probable error of about only 0.02. Now, a correlation thus more than nineteen times the size of its probable error would not occur by mere accident in millions of trials, so that chance, as a possible cause of the apparent correlations, may at once be put completely out of court. Our result has thus made good its right to further elaboration.

But when we consequently proceed to discount the errors of measurement, it unfortunately becomes clear that our data are far from being adequate for the purpose; we have, indeed, a duplicate set of observations for Common Sense, but none for School Cleverness nor for any of the Sensory Discriminations. The excuse for the deficiency lies to some degree in practical difficulties, but still more in the fact that at the time of the experiments I was only just beginning to realize the necessity of

such duplication. In default, then, of better information, the other errors will provisionally be taken as being about the same size as those for Common Sense (later on we shall have opportunity of partially checking this assumption); on this basis, an elimination of the observational errors by our first or theoretical formula brings the required correlation to 0.60.¹ This result can now be compared with that given by the second formula; for this purpose we utilize the fact that Common Sense and School Cleverness prove to be not very different criteria, so that all three lists may be used as measurements of practically the same intellectual faculty; if we accordingly amalgamate these three lists into one, the latter shows an average correlation with the Discriminations amounting to 0.44 and our required correlation comes to 0.54,² or somewhat smaller than by the other way. Such a decrease by the empirical as compared with the theoretical formula could be produced by two causes: either the estimation of School Cleverness might have been more accurate than those for Common Sense, or else there could have occurred the fallacy a priori feared by us, namely that the three critics had been warped by the same prejudices and therefore not able to judge with sufficient independence of one another. As, however, the total divergence only amounts to 0.06, we can conclude that neither of the above disturbances can have existed to any appreciable degree, and we might well ascribe even the small apparent difference to mere chance variation; but to be on the safe side we will adopt the lower value, 0.54.

We next pass to the inquiry into irrelevant factors and will commence with the conspicuous one of Sex. This to all appearance manifests a connection with both Discrimination and Intelligence, and therefore might conceivably be the sole cause of the two latter being congruous with one another. But closer inspection alters the aspect of affairs; for while Sex and Discrimination only show a correlation with one another to the

$$\frac{0.38}{\sqrt{0.64 \times 0.6}} = 0.60$$

² The two forms of intelligence coincide to the extent of 0.84, so that the equation becomes approximately

$$\frac{\sqrt{2+0.84^2} \times 0.44 - 0.38}{\sqrt{2+0.84^2} - 1} = 0.54.$$

Even if we neglect the slight discrepancy between the two sorts of intelligence, the result will not be very different, for

$$\frac{\sqrt{3} \times 0.44 - 0.38}{\sqrt{3} - 1} = 0.53. \text{ Note that as we have here an amalgamation for only one of the two compared series, we must take the square instead of the fourth root of the number of amalgamated lists.}$$

extent of 0.26 (after correction for errors), they respectively correlate with Intelligence to the amounts of 0.59 and 0.54; so that the true correlation between Sex and Discrimination comes to -0.07,¹ that is to say it entirely disappears. Thus it would seem that the correspondences of Discrimination and Intelligence with Sex are in no degree the causes but merely effects of their correspondence with one another and that *the fluctuating differences of Sensory Discrimination observable in connection with Sex at the various stages of growth are chiefly and perhaps altogether a mere consequence of similarly fluctuating differences in their Intelligence.* This hypothesis tallies well with other indications; in the experiments of Gilbert, for instance, as to the most characteristic discrepancies between boys and girls at the various ages, the two senses tested (visual and muscular) present an almost identical progress, as if both were depending on some common influence. The same conclusion can be more directly derived from the fact that either the boys or the girls, taken separately, present correlational values very similar to the above; but such a subdivision so reduces the number of conjoined cases that the probable error becomes too formidable for the attainment of sufficiently regular results. Hence it has been thought both permissible and advantageous to throw the boys and girls together into the one collective experiment.

The next obvious irrelevant factor is that of Age, which apparently exhibits correlations with Discrimination and Intelligence of 0.37 and 0.42 respectively. But as regards the former, first, the true value obtained in the same way as before descends down to only 0.18,² thereby not indeed this time disappearing but still coming almost within reach of some bias in the operation of grading; possibly, then, the connection of Age with Discrimination is at least in the main, like that of Sex, no more than an effect of their common correspondence with Intelligence; this would accord with the strange phenomenon noted above, that Intelligence appears temporarily to diminish about the eleventh year, for precisely the same occurs to their powers of Sensory Discrimination (a fact first pointed out by Gilbert and noticeable in the present experiments also). To turn to the other above indicated correspondence, that between Age and Intelligence, all those who furnished me with their personal estimation of the children's comparative intellect had been particularly requested to do so entirely regardless of Age, and they had anticipated no difficulty on this head declaring

$$^1 \frac{0.26 - 0.59 \times 0.54}{\sqrt{(1 - 0.59^2)(1 - 0.54^2)}} = -0.07$$

$$^2 \frac{0.37 - 0.42 \times 0.54}{\sqrt{(1 - 0.42^2)(1 - 0.54^2)}} = 0.18$$

that their opinions were naturally formed quite independently of any such consideration. But we have seen good reason for being very strict in this respect, and when we actually examine the figures, the observed deviations are often far greater than can be attributed to mere chance; from seven to ten years it is the little ones who are favored and to the large extent of 0.65, which is, however, in this case only three times the abnormally big probable error;¹ but from ten to fourteen the above correlation of 0.42 is found in the opposite direction and upon being corrected only descends to 0.38,² which is more than seven times its probable error and therefore would not occur by mere coincidence in many thousand times. Thus we are impelled to believe, either that judgment of Intelligence is to a great though unconscious extent biased by consideration of Age, or else that there is a stage of development somewhere near the eleventh year where Intelligence temporarily declines; in connection herewith we have the curious fact that from about eleven to twelve years there appears to ensue a suspension in the growth of children's heads. Returning finally to our main topics with these two values, 0.18 and 0.38, the corrected correlation between Discrimination and Intelligence now comes to the slightly reduced amount of 0.52.³

Passing to the next irrelevant factor, Practice, this influence has in the third chapter shown itself to be only moderate as regards distinguishing Light, and still smaller as concerns Weight; moreover, there is no reason to suppose that the children differed appreciably from one another in their amount of previous practice with these two kinds of activity. But in the matter of Sound all this is reversed, for even the most homogeneous school presents a wide diversity of musical education, and we have already seen that such circumstance is of enormous influence. Taking the quantitative estimates given on page 231, in conjunction with the average and its mean deviation in Table I, the correspondence of Discrimination and Practice can be reckoned out to probably amount in such a school to something like 0.70. This very large factor must therefore be eliminated be-

¹ This result depends upon only one short set of observations; also no detailed rank had been furnished, but merely that favorite but particularly bad classification into "bright," "average," and "dull."

$$\frac{0.42 - 0.54 \times 0.18}{\sqrt{(1 - 0.54^2)(1 - 0.18^2)}} = 0.38$$

This method of correcting three inter-correlated terms in succession beginning with the smallest of them, though far from being theoretically exact, nevertheless appears sufficiently applicable to the large majority of actual cases including the present one.

$$\frac{0.54 - 0.38 \times 0.18}{\sqrt{(1 - 0.38^2)(1 - 0.18^2)}} = 0.52$$

fore we can hope to obtain an approximately true result. As it only affects Auditory Discrimination, the total average will thereby be finally increased to about 0.58.

So far, we have broadly taken the general average correlation between all the kinds of Discrimination and those of Intelligence. Let us next briefly consider the individual relations between the several specific sensory and intellective faculties. Common Sense and School Cleverness present practically the same amounts, their raw correlations being 0.39 and 0.36 respectively, so that there appears no object in treating these separately. The various sensory departments show apparently much larger discrepancies: for Weight, which has a raw correlation of 0.34, after correction we eventually get 0.43; for Light, with a raw correlation of 0.44, we come to 0.58; and for Sound, with a raw correlation of 0.37, we arrive at no less than 0.71. To this disparity, however, no great evidential value can be attached, until more positive estimates have been obtained as to the errors of measurement; the above rank might really only mean that the accuracy of experimentation had been least in the thresholds for Weight and greatest in those for Pitch.

Lastly comes the process of reviewing the whole argument. Clearly enough, it has in many respects been of a rough character. The arbitrarily assumed observational errors for School Cleverness and for Discrimination were but inadequately checked in the former case and not at all in the latter; while the irrelevant influence of previous practice on Auditory Discrimination was based upon solid but too general data. Also, there is always the danger that other unsuspected irrelevant factors may exist in harmful magnitude; on this head, however, the precautions taken appear fairly adequate; chapter III only shows the results of the preliminary investigation as regards those factors which finally proved most formidable, but a great number of others have been thoroughly examined and their influence has been found to be inappreciable for our present purpose. Among all these sources of inaccuracy, it will be manifest how insignificant is here the rôle of the probable error of the raw main correlation (0.02); thus, though the reagents were only twenty-four in all, it would have been worse than useless to augment their number at the expense of correctness in other respects; for the present, increased precision must be chiefly sought by other means. To sum up, the most likely value for the average correlation between the Discriminations and the Intelligences comes to about 0.58, but this final conclusion must be considered as having a large total probable error, say, 0.10.

(b) *Experimental Series II.* On first inspection, the results here would seem diametrically opposed to our last ones, the

correlation turning out to be on the *minus* side, so that Discrimination would appear correlated to the extent of 0.39¹ not with Intelligence but with Stupidity.

But we again notice that both Intelligence and Discrimination are irrelevantly correlated with Age, and to the large amounts of 0.69 and 0.81 respectively.² On applying the corrective formula, the true correlation turns round to +0.41,³ whereby the paradox is readily resolved into a result under the circumstances perfectly normal; we should expect the amount to be smaller than in the former experimental series, because the observational errors must have been greater and have required larger compensation than before, whereas we have given the same.

Thus it is once more evident that the influence of irrelevant factors, though sometimes of moderate magnitude (as in the preceding series) may at other times assume such dimensions as to wholly reverse the conclusion. Nor would such an effect be in the least diminished by increasing the number of reagents experimented upon.

Except for this lesson concerning the danger of irrelevant factors, there is little information to be gleaned from the present series; for while the observed correlation is only —0.25, the probable error comes to no less than 0.18; in other words, such a correlation would turn up about every third time, either when no correspondence existed at all, or when there really was one twice as large. Hence we see that though a short series of cases *may* be managed in such a way that the rôle of the probable error becomes insignificant, yet this is not necessarily so; when the data are gained in the customary unscientific manner, the results of these brief experiments are worthless for persons versed in correlational methodics and delusive for those who are not so.

(c) *Experimental Series III.* Here the probable error has been reduced to more reasonable dimensions (0.06 for the average of two correlations); also the disturbances from Sex and Age have been eliminated. But on the other hand (as we have seen on page 248), the sensory tests were so unsatisfactory that the attenuation by errors must be estimated of enormous magnitude. Quite accordingly, the observed correlations of Intelligence with Visual and Tactual Discrimination are no more than 0.13 and 0.12 respectively. Upon this diminutive basis we cannot attempt to pile the very large and vague corrections that would be necessary.

Thus we see that an inadequate way of testing the reagents,

¹ The raw correlation is —0.25.

² The raw correlations are 0.55 and —0.65 respectively.

³
$$\frac{-0.39 - 0.69 \times (-0.81)}{\sqrt{(1 - 0.69^2)(1 - 0.81^2)}} = +0.41.$$

whether due to circumstances, want of skill, or undue hurry, will so increase the attenuation by errors that the raw observable correlation is brought down to a vanishing minuteness. This effect, so far from being remedied by increasing the range of the experiments and the number of cases, will almost inevitably be augmented thereby.

The principal information to be gleaned from this series is that the correspondence between Discrimination and Intelligence cannot well be appreciably due to Zeal, for the latter faculty had upon this occasion become the paramount differentiating influence.

(d) *Experimental Series IV.* On this occasion circumstances were as favorable to accuracy of experimentation (see page 249) as in the last two cases they had been the reverse. Accordingly, we find that Discrimination correlates with Talent in the four branches of study, Classics, French, English, and Mathematics, by an average of 0.51 with a probable error of only 0.03. This proportion of over 17 to 1 is amply sufficient to warrant us in proceeding to determine the real correspondence with considerable precision.

We will again take first the errors of measurement. As regards Intelligence, there was no difficulty in obtaining the required data; for the gradings were based upon the regular school examinations, so that if several of these be taken simultaneously into consideration, each will constitute a sufficiently independent set of measurements. The errors prove much smaller than those in the first experimental series; one examination paper correlates with another in the same branch of study by an average of 0.86; and one total examination correlates with another total examination by an average of 0.95; so that whereas before we had to make a compensation of 24%, we now only require those of 7% and 2%. From this it would appear that examination papers form a test that is far more constant and free from accidental errors of judgment than are teachers' subjective impression as to the "brightness" of their wards; but still this only refers to the reliability of the testing process and does not prejudice the question to be subsequently discussed as to whether the kinds of intelligence tested are different and of unequal value. As concerns the grading of Discrimination, unfortunately the same impediment again frustrated the attainment of several independent sets of observations, so that we once more have no precise measure of compensation; we shall therefore have to make the same free estimate as before. Correcting, then, for both Intelligence and Discrimination, we get a true value of 0.69.¹

¹

$$\frac{0.51}{\sqrt{0.86 \times 0.64}} = 0.69$$

Proceeding next to irrelevant factors, these appear to have been reduced in the present series to such a minimum that there is no more any necessity for theoretical corrections. The matter of Sex does not again come into the question, since this school is only of boys. Nor is Age this time a disturbant, for its correlation with Discrimination works out to the insensible amount of -0.07. The latter circumstance completely disposes of another possible objection, for it might be urged against Experimental Series I that Age is by no means identical with stage of Growth, some children being more precocious than others, so that our correction to eliminate the former factor does not necessarily suffice to nullify variety in the latter one; hence, it might be argued, the whole correspondence between Intelligence and Discrimination could conceivably be due merely to the brighter children being also more forward with their senses. But though it must certainly be admitted that Age and Growth do not always keep level with one another, yet they at any rate correspond to the extent that when the former has ceased to exercise any influence at all the latter also must have become entirely inoperative; and as throughout the school the older boys do not in the least surpass the younger ones in Discrimination of Pitch, we may safely say that their faculty in this respect no longer depends to the smallest degree either on their Age or even on the stage of their Growth.

Still it may be interesting to know what would have occurred had the school lists not been artificially modified, but allowed to retain the factor of Age. In such case, the correlation comes to only 0.45.¹ Thus by eliminating Age we had increased the correlation from 0.45 to 0.69; but to obtain such a rise it can easily be calculated that we must remove an irrelevant factor amounting to about 0.76;² now, in the unmodified lists the actual correspondence of Age with Class Order turns out to be precisely this amount. Thus it would appear that *the influence of Age was wholly irrelevant*; not actual Proficiency but Native Capacity is the factor directly correlated with Discrimination; our apparently somewhat theoretical modification of the original school order was no empty abstraction but had after all a solid enough basis in present fact; if with regard to the educational curriculum it merely represented future possibilities, yet in other directions it showed itself to correspond with already efficient powers.

It is now also evident that the whole process of modifying

$$^1 \quad \frac{0.33}{\sqrt{0.86 \times 0.64}} = 0.45$$

$$^2 \quad 0.69 = \frac{0.45}{\sqrt{1 - x^2}}, \text{ so that } x = 0.76.$$

the school lists could have been avoided; we could have left them in their raw state and simply have eliminated the irrelevant factor of Age by our usual formula; in this way, indeed, the theoretical precision would have been far greater, for our artificial treatment of the lists is only even approximately correct when Age is a paramount influence in deciding school place, and if applied, say, to university students, would produce an improper decrease (never increase) in the observed correlations; but practically the advantage is generally the other way, as in the present case, for the above theoretical incorrectness is more than compensated by a reduction gained in the probable error, owing to our being able to actually observe more of the correlation and thus leaving less to obtain by calculation.

As regards possible irrelevant correspondence with Practice, the present experimental series is again favorably situated; for though this factor has its usual large influence upon Discrimination of Pitch, there are in this case some positive data whereabouts to measure it. Out of the thirty-three children it was ascertained that twenty-two were taking lessons in music, and these not unnaturally showed a much finer Sensitivity (a median of 2.3 v. d. as against one of 5 v. d.). We can therefore reject the eleven not learning music, thus confining the experiment to reagents on nearly the same level as concerns Practice; upon doing so, the correlation makes a further rise to 0.78.¹ Or else we can work out the correlation of Discrimination with Music Lessons, and then remove this irrelevant factor by means of our formula; thereby we arrive at a similar value.

Let us now sum up this fourth series. With respect to the correlations of Discrimination with the School Studies *separately* we have arrived at an average of 0.78, of which figure 0.57 has been actually observed; to this aggregate 0.89 is contributed by Classics, 0.88 by French, 0.73 by English, and 0.61 by Mathematics; here the order may be considered as well enough evidenced, the ambiguity present in the first series having disappeared, since the observational errors have been calculated separately for each study. The correlation of Discrimination with the *Total* School Ability can be calculated with equal ease; the average raw correlation is 0.68, while the successive Total School Orders correlate with one another by 0.95, so that the corrected required correlation becomes 0.87.²

In this series we have had the good fortune, not only to ob-

$$^1 \frac{0.57}{\sqrt{0.86 \times 0.64}} = 0.78$$

$$\frac{0.68}{\sqrt{0.64 \times 0.95}} = 0.87$$

tain reliable estimates of almost all the perturbing influences, but even to eliminate them in practice so that an average of no less than 0.68 can be actually observed. In the further calculation the only value that cannot be approximately relied upon is that of the errors in measuring Discrimination, and even this cannot well be very much smaller than the amount here adopted and, as is subsequently proved, cannot possibly be much larger. The existence of other sources of fallacy yet lurking, either to reduce or to still further augment the total cannot of course ever be categorically disproved; but at any rate a careful search has been made and has so far failed to reveal them. The whole system of results exhibits such regular unconstrained compliance with the definite laws governing correlations (and also present such other remarkable uniformities to be discussed later on) that they appear to offer every guarantee of being perfectly normal.

The correlational value is, however, considerably larger than that found for the same sense in Experimental Series I. The readiest explanation of this discrepancy would appear to lie in a fact which I have often had occasion to notice, namely that when the reagents are very unpractised in any form of sensory Discrimination, the latter correlates with their Intelligence by a much smaller amount.

(e) *Conclusions.* On the whole, then, the results of all four experimental series appear sufficiently concordant with one another. Whenever we have succeeded in obtaining a fairly pure correlation between Sensory Discrimination and Life Intelligence, we have found it amount to a very considerable value. In the case of Pitch, it came to as high as 0.87. Very possibly other discriminative functions would show similar results, while some would prove much more specific (and usually dependent on factors peripheral to the nervous system).

3. Correspondence between General Discrimination and General Intelligence.

Up to now, we have only discussed the correspondence of the various Intelligences with the various sensory activities, Hearing, Sight, Touch, etc. Such isolated facts are interesting enough, but quite otherwise important is the relation of *any common and essential element in the Intelligences to any common and essential element in the Sensory Functions*. For brevity, we will term these common elements "General Intelligence" and "General Discrimination," but always with the reservations made in the first section of this chapter.

Curiously, this more general correspondence can in the present case be settled with much greater precision than was possible for the specific relations. This is due to our now

having adequate data wherewith to measure the errors of observation, seeing that all the experimentally obtained gradings of specific Discrimination constitute so many one-sided independent attempts to grade the General Discrimination; the amount of observational error will be quantitatively revealed in the correlations between one grading and another.

(a) *The Village School.* Here our calculation is as follows. The average of the nine correlations between the Intelligences and the Discriminations comes, as we have seen, to 0.38;¹ the two kinds of intellective gradings correlate with one another by an average of 0.55; and the three gradings in Discrimination do so by 0.25.² Therefore by the theoretical formula the true correlation between General Intelligence and General Discrimination comes to

$$\frac{0.38}{\sqrt{0.55 \times 0.25}} = 1.01.$$

Checking this by the second or empirical method, we find that on taking an amalgamation of the three intellective gradings with an amalgamation of the three gradings in Discrimination, the correlation rises to 0.66. Therefore the true correlation between General Intelligence and General Discrimination comes in this way to

$$\frac{\sqrt{3} \times 0.66 - 0.38}{\sqrt{3} - 1} = 1.04$$

This again may be further checked by taking our amalgamation two instead of three lists at a time; in this way we get nine different correlations which present an average of 0.55, so that our required result now becomes 0.96.³ Therefore an average again gives us as nearly as possible 1.00.

Thus we arrive at the remarkable result that the *common and essential element in the Intelligences wholly coincides with the common and essential element in the Sensory Functions.*

(b) *The High Class School.* Here, also, the children were tested in the three senses, but unfortunately, as we have seen, the results for Light and Weight are not seriously usable,⁴ so that we no longer have sufficient material for constructing a "General" Discrimination.

¹These correlations are here taken as actual measurements, and therefore are obviously required raw, not corrected; the correction then issues from their joint product according to the formula.

²This value is precisely the same as that found for adults: see Table V.

³

$$\frac{\sqrt{2} \times 0.55 - 0.38}{\sqrt{2} - 1} = 0.96$$

⁴As far as they go, they indicate results entirely similar to those above.

This default, however, has been made good by what appears to be a very happy substitute. Our main correlations have dealt with reagents all undergoing musical instruction, and I have kindly been furnished with a complete order of their relative abilities in this department. Musical talent has always been recognized as being not so much an intellective as a sensory function; whole nations appear almost devoid of it, without therefore showing themselves any less intelligent; lunatic asylums, on the contrary, often contain a surprising share of the faculty. We will, then, take this as our second sensory function, will note whether it presents any community with Discrimination of Pitch, and if so will compare this common element with that obtaining between the intellective functions. As regards the first point, it may be noted that hitherto very conflicting opinions have been stoutly maintained; the great majority of writers have held Musical Talent and Pitch Discrimination to be very intimately connected and even go so far as to directly term the discriminative power "musical sensitiveness;" while a few, but including perhaps the ablest judges, flatly deny any such correspondence whatever. The actual facts would at first sight seem to lie wholly on the side of the former tenet, seeing that the correlation works out to the substantial amount of 0.40 (or about 0.63, when corrected for errors). Next, these two auditory functions correlate with the Intelligences by 0.57 and 0.55 respectively, and the latter correlate with one another to the amount of 0.71. Thus the relation between the element common to the two former and that common to the four latter will be given by

$$\frac{0.56}{\sqrt{0.40 \times 0.71}} = 1.04$$

We can now check the result by the empirical formula; for we find that the amalgamated order derived from the two sensory faculties correlates with the amalgamated order derived from the four Intelligences by 0.72; so that the required correlation comes to

$$\frac{\sqrt[4]{8} \times 0.72 - 0.56}{\sqrt[4]{8} - 1} = 0.96$$

Taking as usual the mean,¹ we again reach a final correlation of precisely 1.00, and therefore once more must conclude that the element common to the sensory activities also wholly coincides with that common to the intelligences.

¹If this small difference of value between the theoretical and empirical results be minutely investigated, it can be clearly proved to be solely attributable to mere chance, as indeed might well be expected from its small dimensions.

Before passing, it may be remarked that thus after all those were virtually in the right who maintained Musicality and Pitch Discrimination to have no correspondence with one another; for though a correspondence really does exist, yet it is not to the smallest degree of the specific character contemplated by those who talk of "musical sensitivity." It must here also be noted that this surprising intellectuality of musical talent by no means annihilates the many well-evidenced phenomena seeming to indicate the contrary; one fact cannot destroy another, and any apparent conflict merely proves our imperfect acquaintance with their true nature.

(c) *Practical Verification of the Argument.* The conclusion above arrived at is so important and the method of argument is so new, that I have endeavored to reproduce analogous circumstances artificially, so that any one may easily test any portion of the reasoning.

The main argument was repeated as follows. A target was constructed of a great many horizontal bands, numbered from top to bottom. Then a man shot successively at a particular series of numbers in a particular order; clearly, the better the shot, the less numerical difference between any number hit and that aimed at; now, just as the measurement of any object is quite appropriately termed a "shot" at its real value, so, conversely, we may perfectly well consider the series of numbers actually hit in the light of a series of measurements of the numbers aimed at. When the same man again fired at the same series, he thereby obtained a new and independent¹ series of measurements of the same set of objects. Next, a woman had the same number of shots at some set numbers in a similar manner. If, then, our above reasoning and formulæ are correct, it should be possible, by observing the numbers hit and working out their correlations, to ascertain the exact resemblance between the series aimed at by the man and woman respectively. In actual fact, the sets of numbers hit by the man turned out to correlate with those hit by the woman to the extent of 0.52; but it was noted that the man's sets correlated with one another to 0.74, and the woman's sets with one another to 0.36; hence the true correspondence between the set aimed at by the man and that aimed at by the woman was not the raw 0.52, but

$$\frac{0.52}{\sqrt{0.74 \times 0.36}} = 1.00,$$

that is to say, the two persons had fired at exactly the same series of bands, which was really the case. I repeated this ex-

¹ Provided, of course, that there be no appreciable constant error.

periment, testing three times by the first or theoretical formula and four times by the empirical one; by both methods the average came to just upon 1.00, with a mean variation above and below of precisely similar dimensions to those in our instances of Discrimination and Intelligence. Thus the experimental justification of our method of argumentation was as complete as could well be desired.

(d) *Conclusion.* On the whole, then, we reach the profoundly important conclusion that *there really exists a something that we may provisionally term "General Sensory Discrimination" and similarly a "General Intelligence," and further that the functional correspondence between these two is not appreciably less than absolute.*

Besides its intrinsic value, such a general theorem has the enormous advantage over the specific results of the last section of being independent of any particular conditions; it has nothing to do with the procedure selected for testing Discrimination and Intelligence, nor even with the accuracy of its execution, nor indeed even with the homogeneousness of the experimental subjects; if correct, the proof should be reproducible in all times, places, and manners—on the sole condition of adequate methodics.

4. *Universal Unity of the Intellective Function.*

In view of this community being discovered between such diverse functions as in-school Cleverness, out-of-school Common Sense, Sensory Discrimination, and Musical Talent, we need scarcely be astonished to continually come upon it no less paramount in other forms of intellectual activity. Always in the

present experiments, approximately, $\frac{r_{pq}}{\sqrt{r_{pp} r_{qq}}} = 1.$ ¹

I have actually tested this relation in twelve pairs of such groups taken at random, and have found the average value to be precisely 1.00 for the first two decimal places with a mean deviation of only 0.05. All examination, therefore, in the different sensory, school, or other specific intellectual faculties, may be regarded as so many independently obtained estimates of the one great common Intellective Function.

Though the range of this central Function appears so universal, and that of the specific functions so vanishingly minute, the latter must not be supposed to be altogether non-existent.

¹Where r_{pq} = the mean of the correlations between the members of the one group p with the members of the other group q,

r_{pp} = the mean of the inter-correlations of the members of the group p among themselves,

and r_{qq} = the same as regards group q.

We can always come upon them eventually, if we sufficiently narrow our field of view and consider branches of activity closely enough resembling one another. When, for instance, in this same preparatory school we take on the one side Latin translation with Latin grammar and on the other side French prose with French dictation, then our formula gives us a new result; for the two Latin studies correlate with the French ones by an average of 0.59, while the former correlate together by 0.66 and the latter by 0.71; so that the element common to the Latin correlates with the element common to the French

$$\text{by } \frac{0.59}{\sqrt{0.66 \times 0.71}} = 0.86 \text{ only.}$$

That is to say, the two common elements by no means coincide completely this time, but only to the extent of 0.86² or 74%; so that in the remaining 26%, each pair must possess a community purely specific and unshared by the other pair.²

We therefore bring our general theorem to the following form. *Whenever branches of intellectual activity are at all dissimilar, then their correlations with one another appear wholly due to their being all variously saturated with some common fundamental Function (or group of Functions).* This law of the Universal Unity of the Intellectual Function is both theoretically and practically so momentous, that it must acquire a much vaster corroborative basis before we can accept it even as a general principle and apart from its inevitable eventual corrections and limitations. Discussion of the *subjective* nature of this great central Function has been excluded from the scope of the present work. But clearly, if it be mental at all, it must inevitably become one of the foundation pillars of any psychological system claiming to accord with actual fact—and the majority of prevalent theories may have a difficulty in reckoning with it.

Of its objective relations, the principal is its unique universality, seeing that it reappears always the same in all the divers forms of intellectual activity tested; whereas the specific factor seems in every instance new and wholly different from that in all the others. As regards amount, next, there seems to be an immense diversity; already in the present examples, the central factor varies from less than 1/5 to over fifteen times the size of the accompanying specific one. But all cases appear equally susceptible of positive and accurate measurement; thus we are becoming able to give a precise arithmetical limitation

¹The influence of an element is measured by the *square* of its correlational value. See "The Association between Two Things."

²Of course this specific community is further resolvable into natural talent and favoring circumstances of which factors the latter may often be paramount.

to the famous assertion that "at bottom, the Great Man is ever the same kind of thing."

Finally, there is the exceedingly significant fact that this central Function, whatever it may be, is hardly anywhere more prominent than in the simple act of discriminating two nearly identical tones; here we find a correlation exceeding 0.90,¹ indicating the central Function to be more than four times larger than all the other influences upon individual differentiation. Not only the psychical content but also the external relations of Sensory Discrimination offer a most valuable simplicity; for it is a single monotonous act, almost independent of age, previous general education, memory, industry, and many other factors that inextricably complicate the other functions. Moreover, the specific element can to a great extent be readily eliminated by varying and combining the kind of test. For these reasons, Discrimination has unrivalled advantages for investigating and diagnosing the central Function.

5. *The Hierarchy of the Intelligences.*

The Theorem of Intellective Unity leads us to consider a corollary proceeding from it logically, testing it critically, and at once indicating some of its important practical uses. This corollary may be termed that of the Hierarchy of the Specific Intelligences.

For if we consider the correspondences between the four branches of school study, a very remarkable uniformity may be observed. English and French, for instance, agree with one another in having a higher correlation with Classics than with Mathematics. Quite similarly, French and Mathematics agree in both having a higher correlation with Classics than with English. And the same will be found to be the case when any other pair is compared with the remainder. The whole thus forms a *perfectly constant Hierarchy* in the following order: Classics, French, English, and Mathematics. This unbroken regularity becomes especially astonishing when we regard the minuteness of the variations involved, for the four branches have average correlations of 0.77, 0.72, 0.70, and 0.67 respectively.

When in the same experimental series we turn to the Discrimination of Pitch, we find its correlations to be of slightly less magnitude (raw) but in precisely the same relative rank, being: 0.66 with Classics, 0.65 with French, 0.54 with English, and 0.45 with Mathematics. Even in the crude correlations furnished by the whole school without excluding the non-musicians, exactly the same order is repeated, though with

¹ See page 276.

the general diminution caused by the impurity: Classics 0.60, French 0.56, English 0.45, and Mathematics 0.39.

Just the same principle governs even Musical Talent, a faculty that is usually set up on a pedestal entirely apart. For it is not only correlated with all the other functions, but once again in precisely the same order: with Classics 0.63, with French 0.57, with English 0.51, with Mathematics 0.51, and with Discrimination 0.40. Ability for music corresponds substantially with Discrimination of tones, but nevertheless not so much as it does with algebra, irregular verbs, etc.¹

The actual degree of uniformity in this Hierarchy can be most conveniently and summarily judged from the following table of correlation; the values given are those actually observed (theoretical correction would modify the relative order, but in no degree affect the amount of Hierarchy or otherwise). Each number shows the correlation between the faculty vertically above and that horizontally to the left; except in the oblique line italicized, the value always becomes smaller as the eye travels either to the right or downwards.

	Classics.	French.	English.	Mathem.	Discrim.	Music.
Classics,	0.87	0.83	0.78	0.70	0.66	0.63
French,	0.83	0.84	0.67	0.67	0.65	0.57
English,	0.78	0.67	0.89	0.64	0.54	0.51
Mathem.,	0.70	0.67	0.64	0.88	0.45	0.51
Discrim.,	0.66	0.65	0.54	0.45		0.40
Music,	0.63	0.57	0.51	0.51	0.40	

Altogether, we have a uniformity that is very nearly perfect and far surpasses the conceivable limits of chance coincidence. When we consider that the probable error varies between about 0.01 for the ordinary studies to about 0.03 for music, it is only surprising that the deviations are not greater. The general Hierarchy becomes even more striking when compared with

¹Of course, notable instances will easily be found where musical ability is apparently divorced from General Intelligence; in this very school, for example, the best musician is far from standing high intellectually. But not even the most extreme cases necessarily contravene the above rule. A correlation does not state any absolute coincidence between two faculties, but only a limited and precisely measured tendency in this direction; so far from excluding deviations, it proclaims them and even estimates their exact probability. If we may assume the normal law of frequency to approximately hold good and may abstract from further influences, then the proportion of persons with any given amount of musical talent who will attain to any given degree of stupidity (*or vice versa*)

$$= \frac{I}{2} - \frac{I}{\sqrt{\pi}} \int_{-\infty}^{ah} e^{-t^2} dt$$

where h is a measure of the correlation between Musicality and Intelligence, and a = the given inferiority in the latter faculty.

the oblique line, which is no measure of the central Function and where consequently the gradation abruptly and entirely vanishes.¹

The above correlations are raw, and therefore do not tell us either the true rank of the respective activities or the full absolute saturation of each with General Intelligence. For the former purpose we must eliminate the observational errors, and for the latter our result must further be *squared*.² Thus we get:

Activity.	Correlation with Gen. Intell.	Ratio of the common factor to the specific factor.
Classics,	0.99	99 to 1
Common Sense,	0.98	96 4
Pitch Dis.,	0.94	89 11
French,	0.92	84 16
Cleverness, ³	0.90	81 19
English,	0.90	81 19
Mathematics, ⁴	0.86	74 26
Pitch Dis. among the uncultured, ⁵	0.72	52 48
Music,	0.70	49 51
Light Dis., ⁵	0.57	32 68
Weight Dis., ⁵	0.44	19 81

It is clear how much the amount of any observable raw correlation depends upon the two very different influences: first, there is the above intellective saturation, or extent to which the considered faculty is functionally identical with General Intelligence; and secondly, there is the accuracy with which we have estimated the faculty. As regards the ordinary school

¹ The only other data of this kind with which I am acquainted are some comparisons made between the different branches of study at the Columbia University in the course of the research quoted on page 218. The correlations there obtained, which were throughout somewhat smaller than the above, manifest only a limited concordance with our above principle of Hierarchy. But a university is clearly not the place in which to look for natural correspondence between functions; at that time of life, strong ties of a wholly artificial sort have intervened; each student singles out for himself that particular group of studies tending to his main purpose and devotes to them the most judicious amounts of relative energy. To determine natural correlations, we must go to where the pupils meet each other in every department on relatively equal terms.

² See note to page 273.

³ Here so termed for brevity; really that quality is meant which causes a person to be regarded by his teachers as "clever."

⁴ The opposite and more usual view, namely, that mathematics form an entirely independent faculty, will be found expounded in 331 pages "Ueber die Anlage zur Mathematik" by the well-known psychiatrist, Möbius. Similar evidence is brought by him to the effect that this talent is proportional to the development of the upper outer orbit of the eye, especially the left.

⁵ As has been before mentioned, the rank of these three faculties remains ambiguous until their observational errors have been ascertained.

studies, this accuracy is indicated by the oblique italicized line, and therefore appears about equal in all cases (not in the least following the direction of the Hierarchy); but in other cases there is a large divergence on this head, which leads to important practical consequences. Mathematics, for example, has a saturation of 74 and Common Sense has one of about 96; but in actual use the worth of these indications becomes reversed, so that a subjective impression as to a child's "brightness" is a less reliable sign than the latter's rank in the arithmetic class; almost as good as either appears a few minutes' test with a monochord.

In the above Hierarchy one of the most noticeable features is the high position of languages; to myself, at any rate, it was no small surprise to find Classics and even French placed unequivocally above English (note that this term does not refer to any study of the native tongue, but merely to the aggregate of all the lessons conducted therein, such as History, Geography, Dictation, Scripture, and Repetition).

However it may be with these or any other special facts, here would seem to lie the long wanted general rational basis for public examinations. Instead of continuing ineffectively to protest that high marks in Greek syntax are no test as to the capacity of men to command troops or to administer provinces, we shall at last actually determine the precise accuracy of the various means of measuring General Intelligence, and then we shall in an equally positive objective manner ascertain the exact relative importance of this General Intelligence as compared with the other characteristics desirable for the particular post which the candidate is to assume (such as any required Specific Intelligences, also Instruction, Force of Will, Physical Constitution, Honesty, Zeal, etc.; though some of these factors cannot easily be estimated separately, there is no insuperable obstacle to weighing their *total influence* as compared with General Intelligence). Thus, it is to be hoped, we shall eventually reach our pedagogical conclusions, not by easy subjective theories, nor by the insignificant range of personal experiences, nor yet by some catchpenny exceptional cases, but rather by an adequately representative array of established facts.

6. *Outer Factors Determining the Amount of Correlation.*

The values given in the preceding section show the correlations found for various specific activities. These amounts, however, are not wholly constant; the apparent or raw correlations, as we have abundantly seen, deviate in every direction, depending almost entirely upon the number and kind of impurities suffered to enter and vitiate them; but even the true corrected correlation appears to admit of no inconsiderable variation,

according to the conditions of experiment. Generally speaking, the amount seems to be always greater in proportion as the reagents are enabled to *manifest their finest powers*; nor is such factor readily resolvable into the greater regularity thereby obtained. It is especially conspicuous in the following phenomena:

The correlation is augmented when Discrimination is calculated according to the reagents' acutest perception under the most favorable circumstances, rather than according to their freedom from accidental slips.¹ The correlation is larger when all the reagents have the function in question well developed, either by general habit or by careful fore-exercise, than when all are in a comparatively backward stage.² So far, this increase has only manifested itself for the lower grades of practice; it might conceivably be reversed, on approaching the higher grades of special training.

The correlation increases, when the conditions of examination are such as least to distract or puzzle the reagent.

The other factor appearing at all likely to determine the amount of correlation is Age. Within the narrow range of most of our experiments, however, namely from nine to fourteen years, this influence is not apparent. To measure exactly any such change, correlations have been calculated of a secondary order, that is, between the above correspondence on the one hand and Age on the other; even among the village children, who are still in the process of developing their sensory acuteness, this correlation of secondary order only amounts to -0.15, thus indicating that the correspondence between Discrimination and Intellect is almost as great among the older children as among the younger ones. For the boys of the preparatory school, who have already reached their full sensory powers, this secondary correlation comes to -0.07, testifying that the above correspondence is as nearly as possible uniform throughout the different ages. Finally we have seen evidence that this correspondence is the cause of the correspondence between one kind of Discrimination and another, and the latter has proved of exactly the same value for adults as for children (0.25 raw).

It must, however, be mentioned that the opinion has been arrived at and stoutly defended by Wissler, that any corre-

¹This seems to indicate an opposition between the sensory acuteness due to Intelligence and that arising from Practice. This, again, would evidently conflict with Fechner's principle of measuring the sensory threshold by means of Gauss' formula.

²It must be mentioned that Binet has arrived at the contrary conclusion, namely that correlations with Intelligence are only observable on first trial and almost disappear when the reagents are again tested. See *L'année psychologique*, Vol. VI.

spondence between Intelligence and Mental Tests can only exist among young children and must disappear with advancing adolescence. He bases his view firstly upon the fact that his Columbia experiments with University students show no appreciable correlations of this nature; this argument will be dealt with in the next section, criticising these same experiments. And secondly, he truly enough remarks that even the correlations found by Gilbert in children from six years upwards are no longer evident in their seventeenth and eighteenth years. But Wissler does not seem to have been able to measure these correlations of Gilbert quantitatively, the latter not having furnished the data required for the standard formula; they may, however, easily be reckoned by means of the method of "class averages";¹ they will then be seen to be all very irregular and throughout in dangerous proximity to the amount of the probable error; though the seventeenth and eighteenth years do indeed show little correspondence, yet the sixteenth year exhibits the very highest of all while the twelfth year has the lowest; as regards the *general* tendency of these correlations, it is really if anything to *increase* with Age (such tendency only amounts to 0.15 ± 0.08 and therefore is probably a mere chance).

7. Previous Researches Conflicting with the Present Results.

The great bulk of past experiments do not admit of direct comparison with the present ones; for while the latter have been expressly confined to the most elementary forms of laboratory procedure, the former have continually striven to cover vast territory so as to summarily exhaust the whole problem. These more ambitious researches, therefore, can only be criticised in the general manner attempted in Chapter II.

But three investigations—luckily such as to represent the best work accomplished in this department—have also included our present topic, simple Sensory Discrimination. These three will now be discussed in more detail.

Gilbert. The first is that of J. Gilbert, whose valuable and already frequently mentioned experiments upon over 1,000 school children in 1893 included an inquiry into "muscle sense" and "color differences." The apparatus for the former, as for my own, consisted in a number of small boxes which looked and felt exactly similar but were really a set of finely graduated weights. Also his "color differences" were very analogous to my sight tests, for both utilized a series of objects each slightly darker than the preceding one and both were executed by day-

¹ See "The Association between Two Things." *This Journal*, XV, 1904, pp. 72-101.

light.¹ His grading of "mental ability" lay in the usual classification of the children by their respective masters into "bright," "average," and "dull."

As regards his results, Gilbert is perhaps the most conspicuous among those investigators who *do* find an appreciable correspondence between "mental tests" and "general ability." Concerning the above two Sensory Discriminations, he unfortunately confines himself to the following indirect information: "The curves for reaction-time gave the most positive results showing that the brighter the child the more quickly he is able to act. *In discrimination the same relation is noticeable but to a less degree.*"² Thus we are referred on to the correlation for reaction-time, which is everywhere emphasized. "The difference between the reaction-time of those who were bright, of average mental ability and dull respectively" becomes "very noticeable." Again, "the bright children react much more quickly than the dull," and "it is shown here that we judge of a child's mental ability by the quickness or rapidity with which they were able to act." Though Gilbert contents himself with these utterances of rather vague character, he here carefully furnishes most of the essential data for more precise conclusions; he tells us the average reaction-time of "bright," "average," and "dull" for each of the twelve years tested; also he gives us all the mean variations.

With this knowledge, we are easily enabled to work out the exact correlation for ourselves,³ and are surprised to find that it after all averages no more than 0.19 (± 0.04).⁴ The correlation of general ability with sensory discrimination, being even less noticeable, must indeed be minute.

To explain this low figure we must note that from the described mode of procedure the conditions would seem to have here been at least as unfavorable as they were in the present Series III. In confirmation of this view, it may be seen that

¹This similarity of apparatus by no means implies any proportional similarity in mode of proceeding.

²Studies Yale Psych. Lab., II, p. 94. The italics are mine.

³By the method of "class averages."

⁴Here we have an illustrative instance of operating with large numbers of cases. This correlation of Gilbert was based upon an examination of no less than some 1,100 children; therewith we may compare the chief correlation in the present Series IV based on examination of only 22. To the layman, the latter result would seem the immensely more exposed of the two to the danger of being a mere chance coincidence. But when the matter is worked out precisely according to the true and established laws of chance, a correlation like that found by Gilbert, being less than four times its probable error, would occur by mere accident about once in 200 times; whereas the correlation in Series IV, being over twenty times its probable error, would not so occur in millions upon millions.

the average threshold found by Gilbert for discrimination of weight is as coarse as $1/11$, whereas even that in Series III comes to $1/15$ and the more carefully executed Series I shows $1/20$. The natural consequence would necessarily have been to produce a similar attenuation by inadequate representativeness and therefore a similarly small correlation as in that series. Moreover, such an effect would have been materially enhanced by the heterogeneity of Gilbert's reagents, which must inevitably have introduced serious irrelevant correlation. If these considerations be justified, his observed correlations must be taken as being very much smaller than the true ones would have been.

Seashore. The next results to be criticised are those of Seashore, which include an important investigation into the same Pitch Discrimination with which our own experiments have been so much occupied. Seashore, as we have seen, comes to the resolutely negative conclusion "*that there is no functional relation*" between any one of these (mental tests) and general mental ability."

To support this verdict in the case of Pitch, he draws up the following table, remarking, "the distribution of the results practically coincides with the most probable distribution according to chance, which is indicated in the parentheses."

General Mental Ability.

	I	II	III	IV	V
Discriminative	I II	5 (5) 16 (14)	8 (6) 18 (16)	3 (7) 18 (17)	5 (4) 7 (10)
Sensibility	III	6 (8)	7 (10)	15 (10)	5 (6)
for Pitch.	IV	6 (7)	7 (8)	8 (9)	8 (5)
	V	4 (3)	3 (3)	3 (4)	2 (2)
					1 (1)

But this ingenious mode of calculating correlation is of a somewhat disseminated nature, and one may be pardoned for thinking it hardly adapted for giving very accurate results. If, instead of trying to consider fifty amounts all at the same time, we sum them up a little and compare averages we shall find after all a fairly definite tendency for the higher ability to be also accompanied by higher place in Discrimination. If we desire a still more unified and really usable result, we can easily obtain it by any of Pearson's or my auxiliary method;¹ it works out by Pearson's method to 0.24 (± 0.07). Thus here, where functional relation has been so categorically denied, it is in reality greater than in Gilbert's reaction-times where it was held up as being so complete.

¹The italics are Seashore's.

²Seashore's data are not sufficient for us to apply the standard formula.

But this value, though plain and positive enough, must nevertheless be admitted to be somewhat small as compared with that in my own experiments. Nor can this be altogether explained in precisely the same way as in the last example, for here the method of procedure seems to have been much more deliberate; the experimentation was skillfully designed, and the seven minutes allowed to each reagent was perhaps sufficient for its proper execution, though about a quarter of an hour appears to me better. But when we consider the composition of the sample of persons experimented upon, we come upon an irrelevant correlation of great magnitude; for we find that though Seashore remarks a great discrepancy between the children under and over ten years of age, yet he throws them all together into the same correlation and thus introduces an irrelevant connection with Age which comes to no less than 0.73 (reckoned again by Pearson's auxiliary method). This irrelevancy has evidently just the same effect, whether it be really due to difference of age; or, as I have suggested, partly to disparity of culture; or even, as Seashore himself supposes, to imperfection of experimentation. In all cases alike, the real correlation comes to about 0.35,¹ which is almost exactly the same result as obtained in my own Series I before allowing for further errors.²

Columbia. We now pass to the third and last series of experiments bearing on our particular question. This is the very extensive and in many respects important one that has been continuously conducted at Columbia University for the past ten years. For our present purpose, it includes "Perception of Weight" and "Perception of Pitch," while an intellectual grading is obtained from the students' average standing in the various university courses.

Again the general conclusion is an unqualified negative, reading as follows:

"The markings of the students in college classes correlate with themselves to a considerable degree, but not with those made in the laboratory."

Here our critical review is rendered much simpler by the correlations having been properly calculated and plainly stated. Wissler's statement is fully borne out by the values quoted,

$$\frac{0.24}{\sqrt{1 - 0.73^2}} = 0.35$$

²Once more Dr. Seashore has an easy opportunity of practically testing these theoretically gained conclusions concerning this work; if he will exclude the disparate cases under ten years of age, and still more if he will then confine his consideration to such as have learnt music, he will be able definitely to ascertain whether or not the correspondence does not thereupon become very apparent.

the average correlation of the tests with Intelligence being only 0.06 and that of the tests among themselves being 0.09, thus in neither case much exceeding the size of the probable error, though the great majority of instances are at any rate positive. The amount for Perception of Weight is unfortunately not given, but that for Perception of Pitch comes to 0.16.

Above we have seen that Wissler would account for this minuteness of correspondence by the greater age of the students as compared with the children usually experimented upon. But so far our available evidence is not at all that correspondence diminishes with age, but rather that it is completely independent thereof. Moreover there does not appear any urgent need of introducing such a factor, seeing that perhaps sufficient explanation is forthcoming otherwise. To begin with, university students are not exactly the average from the schools but more or less a selection of the most able; hence they intellectually form a more homogeneous class, so that all their intellective correlations will be somewhat reduced in amount. More serious, probably, is the fact that the test of Discrimination has here been so impurified by alien elements, that even Wissler himself prefers to call it a test of Memory.

But perhaps the chief source of the lowness of the correlations will be found in the following circumstances, namely: that the subjects were examined three at a time, each being managed by some "student or officer of the department;" that no less than twenty-two different tests were carried out, many of a most difficult character, besides measuring the length and breadth of each reagent's head; that during the leisure moments afforded him in the course of these tests the observing "student or officer of the department" had to note in writing the contour of the reagent's forehead, the character of his hair, the nature of his complexion, the color of his eyes, the shape of his nose, the description of his ears, of his lips, of his hands, of his fingers, of his face, and of his head—and that this whole procedure is considered to be satisfactorily completed in forty-five minutes.

On the whole, then, the apparent conflict of previous researches with the present ones does not appear fundamental or such as in any way to invalidate the evidence now produced.

8. *Summary of Conclusions.*

To conclude, the following is a brief summary of the principal conclusions indicated by the foregoing experiments:

I. The results hitherto obtained in respect of psychic correlation would, if true, be almost fatal to experimental psychology as a profitable branch of science. But none of these results, as at present standing, can be considered to possess any

value other than suggestive only; this fact is not so much due to individual shortcomings of the investigators, as to the general non-existence of any adequate system of investigation.

II. On making good this methodological deficiency, there is found to actually occur a correspondence—continually varying in size according to the experimental conditions—between all the forms of Sensory Discrimination and the more complicated Intellectual Activities of practical life.

III. By this same new system of methodics, there is also shown to exist a correspondence between what may provisionally be called "General Discrimination" and "General Intelligence" which works out with great approximation¹ to *one or absoluteness*. Unlike the result quoted in the preceding paragraph, this phenomenon appears independent of the particular experimental circumstances; it has nothing to do with the procedure selected for testing either Discrimination or Intelligence, nor with the true representativeness of the values obtained by these tests, nor even with the homogeneousness of the experimental reagents; if the thesis be correct, its proof should be reproducible in all times, places, and manners—on the sole condition of adequate methodics.

IV. The above and other analogous observed facts indicate that *all branches of intellectual activity have in common one fundamental function (or group of functions)*, whereas the remaining or specific elements of the activity seem in every case to be wholly different from that in all the others. The relative influence of the general to the specific function varies in the ten departments here investigated from 15:1 to 1:4.

V. As an important practical consequence of this universal Unity of the Intellectual Function, the various actual forms of mental activity constitute a stably interconnected Hierarchy according to their different degrees of intellective saturation. Hence, the value of any method of examination as to intellectual fitness for any given post is capable of being precisely ascertained, since it depends upon:

- (a) the accuracy with which it can be conducted;
- (b) the hierarchical intellective rank of the test;
- (c) the hierarchical intellective rank of the duties involved in the post.

Methods have been given whereby all these three points can be sufficiently ascertained.

VI. Discussion as to the psychical nature of this fundamental Function has been reserved until a more complete acquaintance has been gained concerning its objective relations. Among the latter, the principal and determining one is its

¹ In the present experiments, as far as the second decimal place.

unique position as indicated in paragraph IV. The chief further evidence is to the following effect:

The function appears to become fully developed in children by about their ninth year, and possibly even much earlier. From this moment, there normally occurs no further change even into extreme old age.

In adult life, there would seem no appreciable difference between the two sexes.

The Function almost entirely controls the relative position of children at school (after making due allowance for difference of age), and is nine parts out of ten responsible for success in such a simple act as Discrimination of Pitch.

Its relation to the intellectual activity does not appear to be of any loosely connected or auxiliary character (such as willingness to make an effort, readiness in adaptation to unfamiliar tests, or dexterity in the fashion of executing them), but rather to be intimately bound up in the very essence of the process.

APPENDIX.

EXPERIMENTAL SERIES I.

Village School, 24 Oldest Children.

A. Original Data.

Sex.	Age.	Discriminative Threshold.			Intellectual Rank.		
		Pitch	Light	Weight	Common Sense out of School.		Cleverness in School.
		1/3 v. d.	1:200	1:200	(A)	(B)	
f	II 6	8	4	4	6	5	2
m	12 II	15	3	4	11	7	22
f	12 8	14	6	4	16	10	7
f	13 8	13	4	9	1	1	1
m	II 4	5	14	7	3	2	3
f	II II	25	7	4	10	14	9
f	II 3	10	19	8	8	19	12
f	13 1	10	12	10	2	4	6
m	12 5	18	11	9	5	6	11
m	12 7	14	30	7	21	22	19
f	12 8	60	3	10	12	9	4
f	13 10	20	12	10	13	12	18
f	13 1	40	5	12	4	8	8
f	12 1	45	12	9	9	13	14
f	10 6	33	5	15	15	18	10
m	II 5	25	4	28	17	11	17
m	10 0	90	15	5	22	21	5
f	II 9	17	15	20	14	20	15
m	13 7	24	26	13	19	17	24
m	12 6	18	35	14	18	3	6
m	10 4	70	10	14	23	24	20
m	11 7	17	42	16	24	23	23
f	II 2	28	20	17	7	15	13
m	II 2	90	25	18	20	16	21

EXPERIMENTAL SERIES I.

Village School, 24 Oldest Children.

B. Calculated Correlations.

Description.		Amounts.		
First Term	versus	Second Term.	Raw.	Prob. Err.
Sound Discrim.	Common Sense (A)	+0.44	0.07	} +0.71
" "	" (B)	+0.41	0.07	
" "	School Cleverness	+0.25	0.09	
Light	Common Sense (A)	+0.42	0.07	} +0.58
" "	" (B)	+0.44	0.07	
" "	School Cleverness	+0.47	0.06	
Weight	Common Sense (A)	+0.38	0.07	} +0.43
" "	" (B)	+0.27	0.08	
" "	School Cleverness	+0.37	0.07	
Com. Sense (A)	Common Sense (B)	+0.64	0.05	} +0.86
Common Sense	School Clever. Av.	+0.55	0.04	
Discrimination	Discrimination "	+0.25	0.05	
"	Intelligence "	+0.38	0.02	+0.86
"	Amal. Intell.	+0.66	0.04	
"	Sex "	+0.21*	0.09	+0.00
Sex	Intelligence "	+0.47*	0.06	+0.59
Age	Sound Discrim.	+0.41	0.07	} +0.18
"	Light Discrim.	+0.30	0.07	
"	Weight Discrim.	+0.20	0.09	
"	Intelligence Av.	+0.34	0.05	+0.38
"	Corr. Discrim.-Intell.	+0.15	0.09	+0.00
General Intell.	Pitch Discrim.			+0.72
"	Light "			+0.57
"	Weight "			+0.44
"	Common Sense			+0.98
"	School Cleverness			+0.90
"	General Discrim.			+1.00 (p. e. about 0.02)

The probable error may be taken at about 0.10

*These two correlations have had to be calculated by an auxiliary method, the standard formula not being applicable; hence, the probable error is considerably larger than it would otherwise have been.

EXPERIMENTAL SERIES II.

Same Village School, 36 next Oldest Children.

A. Original Data.

Sex.	Age.	Pitch.	Intellectual Class estimated by Schoolmaster.
m	9	7	6
m	8	7	6
f	7	3	16
m	8	0	24
f	8	0	26
m	8	4	35
m	9	9	35
m	8	2	38
m	9	0	42
f	7	II	48
f	8	2	48
f	7	2	67
f	9	II	67
*f	5	8	70
*m	5	7	74
m	10	2	74
*m	7	0	77
f	7	1	77
*f	6	9	80
m	9	6	80
m	7	10	96
m	8	2	96
*m	6	7	104
*f	6	10	112
m	7	II	112
*m	5	6	120
*m	5	II	120
*m	6	6	120
*m	6	5	> 120
f	7	0	> 120
*f	6	5	> 120
*m	6	1	> 120
m	7	3	> 120
*m	5	6	> 120
*m	6	6	> 120

Those marked with a star are in the infant school; their relative intellectual rank has not been ascertained, and therefore they are not included in the correlations below. Their average threshold is about 40 v. d., or more than double that of the remainder.

B. Calculated Correlations.

Description.			Amounts.		
First Term	versus	Second Term.	Raw.	Prob.Err.	Corrected.
Pitch Discrim.		Intellectual Class	-0.25	0.18	+0.41
Age		Pitch Discrim.	+0.55	0.18	+0.69
"		Intellectual Class	-0.65	0.18	-0.81

