**The correlation between various mental abilities**

by

F. Krueger and C. Spearman

**Contents**

*I. Preliminary considerations*

1. The general problem
2. Why it is necessary to determine degrees of dependence quantitatively
3. A method for compensating for random errors. The “disattenuation formula”
4. A method for avoiding and eliminating constant extraneous factors. The “correction formula”

*II. Description of our experiments*

1. General comments
2. Details of the individual experimental methods

*III. Our results*

1. The raw correlation values
2. Application of the “disattenuation formula”
3. Application of the “correction formula”
4. The “central values” of the various abilities

*IV. Oehrn’s results*

1. Oehrn’s experimental methods
2. The effects of practice and fatigue on the correlation values
3. Application of the “disattenuation formula”
4. The “central values” derived from Oehrn’s data

V. Interpretation of the overall results

VI. The main results

**Appendices**

1. Example protocol
2. Example text for the combination method used by Ebbinghaus
3. Appendix III. Example calculation of Bravais’ correlation coefficient and the formula for its probable error

Since throughout this article frequent reference is made to two previous studies published in the *American Journal of Psychology*, this journal is abbreviated as *Am. J. Psych*.

**I. Preliminary considerations**

*1. The general problem*

The following experimental study aims to determine and more closely examine the quantitative relationships between various mental abilities.

Popular opinion recognizes the existence of such very general connections. For example, when an individual is considered “intelligent”, this typically implies more than a reference to the abilities actually demonstrated so far. Rather, such an attribution additionally expresses the expectation that the individual will also, to some degree, excel at tasks that differ substantially from the ones tested so far. However, scientific psychology meets such connections with the highest reservations, or often even outright dismissal; for many psychologists, “intelligence” is but a name for a random coincidence of several favorable predispositions.

Perhaps even stronger is the popular conviction that there exist connections of a particular, limited nature. For example, children who excel at arithmetic are expected not to lag behind in algebra, either; those recognized as good singers are assumed to have good prospects of learning to play the violin as well. Such presumably, not completely unfounded views, reached their peak at the beginning of the last century, when phrenology posited its familiar fourteen particular intellectual “faculties”. This system of faculties soon collapsed under critical objections.

However, curiously scientific psychology seems to have contented itself with this negative success. In the last twenty years, numerous experimental studies have attempted to measure the connection between mental abilities quantitatively. However, their success has so far remained very modest; the results contradict one another; and especially recently there has been a tendency to deny the connections in question altogether.

This unsatisfactory result, in conjunction with the extraordinary theoretical and practical weight of the problem, led us to investigate such interrelations in a new and, in our view, more thorough way. Unfortunately, to this end we could not avoid trying the reader’s patience with a rather dense discussion and mathematical method that is in principle simple, yet time-consuming.

*2. Why it is necessary to determine degrees of dependence quantitatively*

When attempting to investigate the dependences between various abilities with a fresh view, one first of all has to recognize they are in no way absolute. There are in fact cases in which an individual considered “intelligent” shows only modest intellectual achievements in some domains; e.g., the gifted arithmetician may actually not exhibit a particular talent for algebra. At most, one may be able to find the former type of talent has a greater or lesser *tendency* to accompany the latter. Thus, before moving on, one first has to be able to *calculate the degree of a partial dependence*.

To this end we are going to use the relatively well-known method developed by Bravais, Galton, and Pearson. A sufficiently large number of cases is selected carefully as a suitable sample representing the complete class of subjects. The two variables being compared will sometimes turn out to be more, sometimes less proportional; with a simple calculation we can determine a single number – the so-called correlation coefficient – which is a general measure of proportionality between the two variables. This coefficient, typically represented by the symbol ***r***, is conveniently equal to one when the two data sequences are perfectly proportional, equal to -1 when they are inversely proportional, and zero when they are fully independent. Instead of real numbers, one may as well compare two *ordinal series*; in that case ***r*** reflects the degree of their tendency towards agreement. This method makes dependences between even the most diverse variables quantitatively comparable.

However, it is of course possible for the dependence between two variable to manifest not as simple proportionality, but in a unique and complicated form. This unique form would not be reflected by the coefficient ***r*** (per se). Nevertheless, ***r*** accurately measures the degree of dependence in almost all cases; only when its form deviates from simple proportionality in an extreme (and thus easily noticeable) way will this measure of dependence produce substantial error.

Fortunately, we can disregard such issues in the present study, *because here only these following three conditions regarding the meaning of* ***r*** *must be met: 1. The correlation can be considered greater in cases in which* ***r*** *is significantly greater. 2. When* ***r*** *is close to one, the correlation is almost perfect. 3. On the other hand, when* ***r*** *barely exceeds zero, there is almost no correlation.* With respect to the correlations that actually found here (i.e., rankings of subjects according to their mental abilities), serious objections to these three propositions seem unlikely to be well-founded.

In all cases, in addition to ***r***, a so-called probable error has to be calculated. All results (or differences between them) whose value isnot at least double the probable error (or probable difference) cannot be considered scientifically significant. To exclude chance findings beyond doubt, the relevant value should exceed the probable error at least five-fold. These rules regarding the probable error are especially important in experiments comparing short data series, in which chance plays a correspondingly great role; as is the case for the experiments reported here. (See the formula for the probable error on p. 114.)

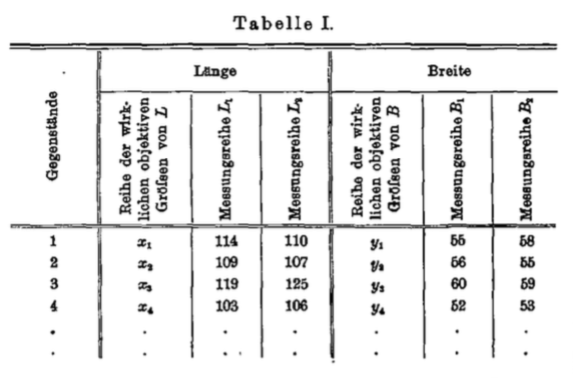
*3. A method for compensating for random errors. The “disattentuation formula.”*

Establishing a correlation is especially complicated by the fact that at the outset such a relationship can never refer directly to the facts compared, but only to the measurement data derived from them.Every kind of measurement, however carefully performed, remains subject to so-called random errors; in the case of psychological measurements, or even measurements of mental abilities, such errors can reach significant sizes.

Let us assume, for example, that one has measured the length and width of a certain number of objects, and further, that each measurement can yield a value that is either too great or too small by the same amount with equal probability. Under these conditions, it is clear that – regardless of the sizes of the individual measurement errors – the mean of either series of measurements may be either higher or lower than the mean of the actual values; and if the number of objects is large enough, the discrepancy between the measured and real mean length (or width) becomes vanishingly small.

The case is completely different when one aims to calculate the *correlation* between the length and width of the objects. Because the more strongly the data points are governed by random errors, the more a potentially present correlation between length and breadth will be *masked*; thus the apparent correlation will never appear too high, but always in according measure *too low* (or course within the range of random fluctuations defined by the probable error). And this kind of disturbance can in no way be eliminated by extending the experimental series, or even by repeating the whole experiment. In fact, this illusory decreaseof the correlation through measurement error appears to be the main cause of the contradictions found in previous results; in experiments with a very large measurement error, even a factually existing correlation cannot become visible; consequently, its existence is erroneously denied. The finding of a small correlation is initially ambiguous; it could point to the actual absence of a relationship, or merely to large measurement errors.

To overcome this ambiguity, measurements have to be carried out at least *twice*. Let us assume for the sake of illustration that this has happened in our above example (the correlation between the length and width of objects); the following table could result:



Thus we obtaintwo series of measurements, *L1* and *L2* (or *B1* and *B2*) reflecting the objective series *x* (or *y*, respectively). The correlation coefficient between *L1* and *L2*is of course the higher, the smaller the random measurement errors are; it thus serves as a “*reliability coefficient*” for the method used in the length measurements. In the same manner, the correlation between *B1* and *B2* constitutes a “reliability coefficient” for the method of width measurements.

These “reliability coefficients” can be used in a precise quantitative manner to restore a “raw” correlation (lowered by random errors) to its full value. To this end, the following “disattenuation formula”, as we would like to call it, is used:



where *A1* and *A2* (or *B1* and *B2*) are two equally accurate, independent series of measurements of **A** (or **B**, respectively); *A1B1* (or *A1B2*, *A1A2*, etc.) is the raw correlation between *A1* and *B1* (or *A1* and *B2*, *A1* and *A2*, etc., respectively); *M* is the mean; and **AB** is the sought *fully restored, thus pure* correlation between **A** and **B**.

Finally, it should be noted that when correcting correlation coefficients for attenuation, usually the concept of “measurement error” has to be understood in a very broad sense; especially when we are not dealing with concrete objects (like in our hypothetical example), but rather (as is much more common) with constructs of a more or less abstract nature. For example, in our study we wanted to measure the subjects’ mental abilities under perfectly comparable conditions, thus in abstraction from their random “current disposition;” however, since random fluctuations in momentary disposition cannot be physically eliminated, they have to be, for our purposes, included under random “measurement errors”.

Because of the above considerations, each subject’s performance at the various abilities to be compared were first tested by Krueger. One week later, the subject was tested by Spearman at the same time of day using the exact same methods. Neither of us learned anything about the other’s results before the end of the experiments.

*4. A method for avoiding and eliminating constant extraneous factors. The “correction formula”*

The deviations of the measurements from the true values discussed in the previous section were of a random kind, i.e., the deviation for each object and each measurement was independent of all other deviations. However, it is also possible for deviations to affect all objects equally; e.g., in our hypothetical example above there could have been a fixed tendency to overestimate the length at, say, 5% of its value; in this case, the individual measurements and their means are disrupted, but the correlation remains intact.

Finally, it is possible – and this case is to be discussed in this section – that the measurement deviations from true values vary from one object to the next, but remain constant across repeated measurements of the same object. This case again disrupts the correlation. And such a disturbance is not eliminated together with the random errors; rather, it points to an extraneous factor to be considered separately, which can be rendered harmless only by special measures. Further, such a factor can not only decrease the resulting correlation (like random errors), but under certain conditions increase it, too.

*All influences that are not strictly part of the relationship investigated are to be regarded as such extraneous factors*. E.g., in our experiments we wanted to determine the relationship between various *innate* abilities. But let us assume a case in which the subjects belong to different age groups, and further, in which the two abilities compared improve with age. Because of their common dependence on age group, the values for the two abilities would also acquire a certain dependence on each other. However, this creates a spurious correlation between the two abilities when no correlation in the sense intended in these experiments is present; on the other hand, if a true correlation *is* present, it *magnified* in an illusory manner.

Let us now assume that *only one* of the two abilities is dependent on the given extraneous factor. Clearly, this does *not* cause an artificial association with the second ability. On the contrary, in as far as the values for the first ability have been forced to conform to the extraneous factor, they have been diverted from any potentially present proportionality with the second ability; thus in this case the correlation is *reduced* in an illusory manner.

Now it is clearly impossible to completely evade the countless disruptions of this kind. We can only hope to reduce them to a negligible level; and to this end, we must have a way to *measure* them. For this purpose, each of the various disturbances discussed here can be expressed in terms of the following “correction formula:”



where **AB** is the directly calculated, hence “apparent” correlation between **A** and **B**; **AC** (or **BC**) is the directly calculated correlation between **A** (or **B**, respectively) and an arbitrary extraneous factor **C**; and is the sought “true” correlation between **A** and **B**. 

In the most common case one aims to eliminate the influence of the dependence of only *one* of the two attributes compared on an extraneous factor, and then the above equation reduces to



The reader may have noticed that the difference between the extraneous factors discussed in this chapter and the random errors discussed in the previous one is in essence of a methodological rather than empirical kind. So e.g., in the present study, we have been able to regard fluctuations in momentary disposition as mere random errors because we re-tested each subject after one week; thus there was time for the subject to arrive in a completely different state the second time. On the other hand, had we re-tested immediately, a thorough investigation of the influence of momentary disposition as an extraneous factor in itself would have been unavoidable.

All these considerations concerning extraneous factors lead to two conclusions of decisive importance. First, especially in the field of correlation research, one has to *pose the question* with far greater clarity than has been customary; i.e., one has to define the states of fact between which a relationship is posited as unambiguously as possible from the outset. This requirement leads the second conclusion: one should never attempt to establish a correlation before one believes to have determined, in a *thorough preliminary investigation*, all extraneous factors that could significantly influence the attributes to be compared. Thus a relationship can never be established merely by mechanically calculating a correlation coefficient. While one must of course have the mathematical tools at hand, a thorough understanding of the relevant subject matter is equally required.

To simplify this preliminary study, for our experiments we have chosen four mental abilities that have been studied extensively in previous work: first, the spatial resolution of the tactile sense, about which numerous studies are known to exist. Next the abilities of addition and rote memorization, regarding which particularly Kraeplin’s school has established the findings most important for our purposes. Fourth, the ability of pitch discrimination, which has already been investigated by one of us in an earlier correlation study. Finally, we added Ebbinghaus’ well-known “combination method” because of the noteworthy results it has previously delivered.

Using historical information on extraneous factors is inconvenient because it is given in very diverse formats, few of which enable us to calculate the correlation between the dependent and independent variables using the ***r***-method. Rather, one needs to have at hand a whole range of mathematical methods, using which one can then produce at least a rough approximation of ***r***. We cannot discuss these numerous auxiliary methods here; note only that it is not the *absolute* influence of the extraneous factor that is crucial, but rather the *ratio of this influence with the mean variation in the values of the relevant attributes*. Further, it should be remembered that disturbances in the results can be neglected only when they are significantly smaller than the probable error. Thus if one wishes to reduce the probable error by extending the experimental series, one has to ensure that extraneous factors are kept at bay with proportionally increasing rigor.

Even though the abilities chosen for this study are well-known, this preliminary research has remained the most extensive and laborious part of the entire work. Based on a comparative literature review we eventually concluded that the abilities of *addition* and *rote memorization* seem least affected by disturbances. Considering the planned scale of our experiments, many of the most well-studied extraneous factors could not possibly cause any disturbance here. In our sample, differences in age and education were too small to merit consideration. Similarly, we did not need to worry about the fact that the subjects consumed alcohol, tobacco, and caffeine to varying degrees. It was not relevant for our purposes that some subjects were in a more favorable working disposition in the morning, others in the evening. It did not matter whether some subjects had eaten an earlier or less satisfying meal than others. Even mental and physical fatigue, within the expected limits, did not require particular consideration. Individual differences regarding practice seemed more relevant; but even in this respect it seemed sufficient to exclude persons with a particular advantage (e.g., those who had previously taken part in similar experiments) from our study.

The results were similar for *pitch*, except that the factor of practice required even more consideration. We had to exclude not only subjects who had previously participated in similar experiments, but also those who were exceptionally *unpracticed* in music.

We encountered greater difficulties with the *tactile resolution*. According to previous results (which, however, remain contentious) from several researchers, not only practice, but even fatigue and minor indispositions seemed to have a problematically major effect. Thus we had to carefully avoid testing subjects in a fatigued or otherwise strained state.

While participating in our study, subjects were not allowed to take part in similar experiments elsewhere. They were also asked to abstain from conversation about the experiments during this period.

Finally, any circumstances that appeared potentially problematic during the experiments themselves, were recorded to at least allow for eliminating their influence after the fact, using the correction formula. For each subject we recorded the age, engagement in mathematics and music, general and current state of health, duration of the last period of sleep, physical and mental exhaustion, signs of emotional or other distraction, time elapsed since the last meal, tobacco and alcohol consumption, room and outdoor temperature, humidity, and air pressure. Further, we attempted to estimate the individual levels of *motivation* with which the subjects engaged in the experiments.

**II. Description of our experiments**

*1. General comments*

The *subjects* who kindly volunteered for this study shared a certain commonality in that they all studied psychology. Except for one outside lecturer, who had studied psychological theory in depth for years, they were all participants of an introductory course in experimental psychology taught by one of us at the Leipzig institute. Thus all subjects had some experience with psychological experiments. Persons with special training in particular domains were excluded, as stated earlier. Other cautionary measures taken eliminate extraneous factors discussed in the previous section (p. 63).

In choosing the *scope* and *methods* for this study we were guided by the following considerations. First, as mentioned above, we limited ourselves to the kinds of testsfor which a sufficiently developed methodology already existed, and using which other authors had carried out extensive experimental series. Further, for our purposes we only considered experimental setups and measurable psychological functions regarding which extensive preliminary studies (using subjects *different from* the ones finally tested) had convinced us that they could generate comparable, relatively stable data over a relatively short time span. While for practical reasons, the experiments carried out by one researcher on a single day could not substantially exceed an hour’s duration, they still required testing five diverse abilities sequentially within one session. This is why we had to exclude a number of *per se* viable and valuable procedures, e.g., complicated memorization tasks, from the final setup.

During the preliminary experiments we decided on the following sequence of experiments within each session:

1. *Pitch discrimination*. Performance on this task appeared to be particularly independent of the disposition with which the subjects approached the experiment because of various immediately preceding influences. This test could be completed with sufficient accuracy within 15–20 min for one pitch range. Next was the
2. Ebbinghaus *combination method* test, the completion of which (within 4 minutes after a preliminary training trial) usually did not cause any fatigue and was perceived as rather pleasant. Only then
3. we tested the tactile *spatial threshold* at three (including a few preparatory trials: four) locations on the skin; this took about 20 minutes on average. These experiments could not be carried out at the beginning of the session because they required skin adaptation to room temperature, which remained relatively constant (air heating) at 15°C on average. This was followed by
4. the *addition* of single-digit numbers, which was strictly time-limited (to 7 minutes) and will be discussed in more detail below; and then finally, in another room of the institute, the most tiring task –
5. the *rote memorization* of sequentially presented series of numbers.

The apparatus and materials for the combination method, addition, and rote memorization were alwaysset up ready for use before the subject entered the room. For each session and each subject the researcher used one copy of the experimental protocol reproduced in Appendix I. The routine questions and answers preceding the experiments helped us to put the subjects into a uniform, calm mood. Part of the questions answered for good in the first session could be skipped when re-testing one week later (by Spearman). After the end of each experiment, the observations under points seven and eight of the protocol were recorded.

*2. Details of the individual experimental methods*

We repeatedly practiced all tests mentioned above – particularly also on each other – in the final sequence; during this practice we developed and rehearsed the details of particularprocedures, to which we then tried to conform as consistently as possible.

a) Pitch discrimination.

The acoustic apparatus used was a monochord with a millimeter scale, constructed by ourselves; both its bridges were partitioned using a nonius, so that distances of 0.1 mm could be read off. The string was tuned to an equal tension before each experimental series using a constant tuning fork. When both string partitions used (left of one bridge, right of the other) were tuned to the same pitch using the overtone method, the string lengths read off very rarely differed by more than 0.1 mm. Because in the pitch range used (350–370 v.d.) 1 mm of string corresponded to one vibration on average, the strings could be tuned with an accuracy of approximately 0.1 v.d. Due to time limitations we had to content ourselves with determining the qualitative threshold for a single pitch range. However, in order to avoid habituation to a specific pitch and certain kinds of random errors, we irregularly “wandered” by a few vibrations within the above mentioned pitch range on each trial.

We proceeded similarly in the tactile threshold experiments: in order to avoid fatigue etc., on each trial the location of the esthesiometer was horizontally shifted by a few millimeters within the skin region currently tested. The procedure for the tactile threshold resembled that for pitch experiments in two further respects. First, in both cases we started with clearly above-threshold stimuli and proceeded in a *descending* manner, at first in greater steps, closer to the threshold in approximately equal smaller steps (at most 1 vibration or 2 mm). Second, in both tests we assumed the threshold had been reached when, for any ten stimuli of the same size, more than two incorrect responses were made; with three incorrect responses, we considered the stimulus to be already sub-threshold, and where necessary defined the arithmetic mean between two stimulus values as the threshold.

The protocol reproduced in Appendix I illustrates the course of a session with the actual information and responses from one of our subjects (of approximately average ability). For simplicity, we used the following symbols: “/” for correct responses, “–” for incorrect responses, and “o” for ambiguous ones.

In the pitch discrimination experiments – point two of the protocol – either the objectively higher or lower pitch was presented first, in an irregular sequence. On each trial, the subject had to judge whether the second pitch heard was “higher” or “lower” than the first one. As can be seen in the protocol, to save time we usually asked for less than 10 responses on suprathreshold stimuli; this was the case only as long as responses were made with extraordinaryconfidence and regularity. Occasionally we had to regress a step back from the threshold.

b) For Ebbinghaus’ *combination method*, we followed the instructions tested and carefully documented by the author in all details. Prof. Ebbinghaus was so kind as to provide us with the original texts authored and on many occasions verified by him. We selected six paragraphs from the group designed for *superordinate categories*, three for the experiments to be conducted by Krueger, and three for those led by Spearman. The texts were copied mechanically and presented in a form identical to that of the originals from Breslau.

In each experiment a short introductory discussion was followed by the subject’s filling in a text reserved for this specific purpose under the supervision of the researcher, until the task was completely clear to the subject. The most important requirement was filling in the gaps *meaningfully*; the second was avoiding gaps, i.e., completeness; *speed* was only a tertiary concern.

This training trial was followed by the main trial, using a new text, which was terminated by the researcher after exactly 4 minutes.

A second text was kept at hand as a backup to be used in case of unforeseen interruptions to the main experiment – which, however, never occurred.

In evaluating the data thus acquired, we again followed the calculation method developed by Ebbinghaus in all details. Thus we counted 1. the total number of syllables filled in; 2. the number of syllables omitted, 3. the number of syllables filled in without discernible meaning. Each syllable omitted contributed half a mistake, each meaningless syllable a whole one; the number of mistakes thus calculated was subtracted from the total number of syllables (1); the resulting value was regarded as the *quantum* of the subject’s performance. Its *quality*, on the other hand, was measured by the total number of mistakes expressed as a percentage of the gross performance (1).

The combination experiments were the only ones in which not all subjects could participate. The foreigners, i.e., the (four) non-Germans had to be excluded, because we had found in preliminary experiments that the performance of foreigners, even when they had mastered German to a high level, lagged behind the mean performance of the German participants significantly (as was to be expected with German texts).

Incidentally, because of the characteristics of our subjects and the experimental setup, certain difficulties Ebbinghaus had to contend with in his group experiments with school children could be left aside; such as deliberately not following the instructions due to lack of motivation, copying from other students, etc.

An example text illustrating the method is shown in Appendix II; it has been *filled in* by a subject of (for this task) average ability. Mistakes and omissions are underlined.

c) The tactile thresholds were measured, as shown in the protocol, in turn on three moderately sensitive skin regions. For practice a short, initially overt series of trials were carried out on the back of the *left* hand. At each level of stimulus intensity we alternated irregularly and covertly between double-tip and single*-*tip touch.

The esthesiometer used, constructed by one of us specifically for this purpose, was particularly light (25 g, made of magnalium), convenient to handle, and no longer than necessary. The two identically tapered tips were made of hard rubber. An essential innovation was including — in addition to the two parallel tips — a third one, oriented diagonally, which allowed us to perfectly match the direction and quality the double touch with the single tip.

In order to control for practice and other factors that could have changed during the experiment, at the end we tested the right cheek bone immediately after the left one. By determining three threshold values and averaging them in each subject, we hoped to cancel out random fluctuations and thus increase the reliability of this measurement.

d) The *addition* experiments were carried out using machine-typed columns of 70 single-digit numbers each. The procedure resembled the methods used by Kraeplin’s school. The only difference was that every 10th number was set apart from the preceding ones with a horizontal line, and the subject was to add up each group of 10 numbers thus created and write down the result next to it. We believe to have sufficiently reduced the influence of mere writing ability in this manner, while still preserving a measure of performance *quality* (the accuracy of addition). This measure of accuracy was the percentage of correct sums.

Each training trial lasting 1 minute was followed by two main trials, each lasting exactly 3 minutes; we conducted two main trials for additional control and to increase the reliability of the test. If at the end of a main trial the subject was not just finishing up a group of ten numbers, he was to write down the last sum calculated next to the number he was just working on. These incomplete final groups were treated just like the completed groups when calculating addition accuracy.

e) *Rote memorization*. The single-digit numbers to be memorized were printed on the cardboard pieces used with Wieth’s memorization apparatus in such a way that the break between every two subsequent numbers in a series was as long as their respective presentations. For the presentations (i.e., showing of the individual numbers) we chose a duration of 0.75 seconds, which according to both Reuther’s extensive observations and our own preliminary experiments seemed to be the most convenient and favorable. Before the beginning and at the end of each series the rotating disc showed a special signal shape.

In designing the number series we used the experiences of earlier researchers and avoided series of numbers favored by associative connections.

We worked with series of 6, 8, 10, and 12 numbers, respectively, in each session in this order; the series had been determined by preliminary experiments to present moderate to high difficulty to learners. Three series of each length were memorized in turn. Thus 12 series of numbers were presented in total; in this way we hoped to control for factors that could potentially change during the experiments, as well as achieve sufficient reliability.

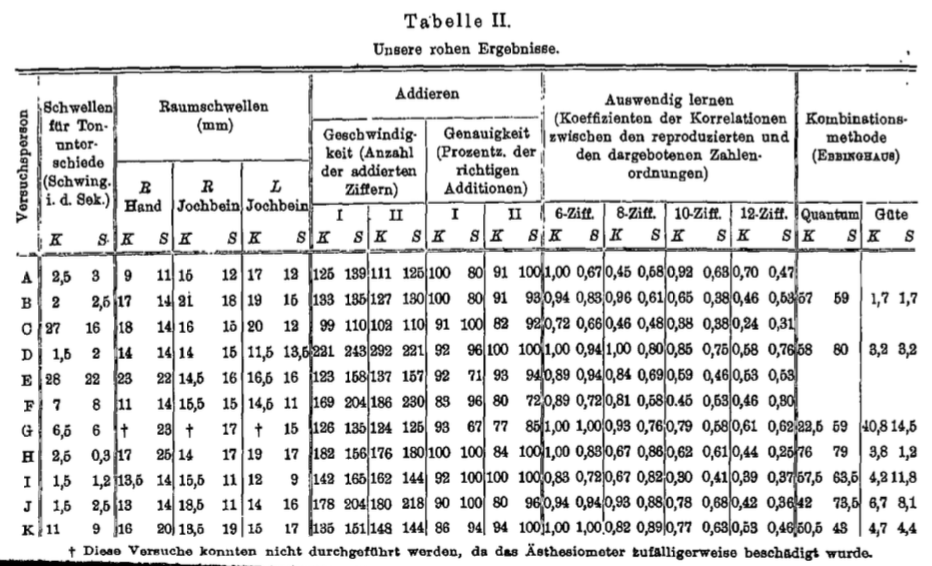
Each series of numbers was presented *once*. After the closing signal the subject was asked to immediately *write down* all numbers memorized on a sheet of paper prepared for this purpose. We did not want to forego testing what was memorized by reproduction, and preliminary experiments had shown that writing was less disruptive than speaking out loud.

When reproducing the numbers, subjects were asked to assign each one a specified place within the sequence of 6, 8, 10, or 12 numbers. Essentially only this sequential order played a role in our evaluation. It was necessarily the most important element because the constituents of the series (especially the longer ones) had to remain more or less identical; by considering exclusively the sequential order the results only gained in clarity. Additionally, in this way the usual, complicated and more or less arbitrary calculation methods could be circumvented. We only needed to calculate the correlation between the real (presented) series of numbers and that written down by the subject. To this end we used a method previously described by one of us, which allows for an incomparably easier and faster calculation than does the ***r***-method. For omitted or incorrect numbers, probable random values were inserted into the equation, so that the subjects gained nothing by guessing.

**III. Our results**

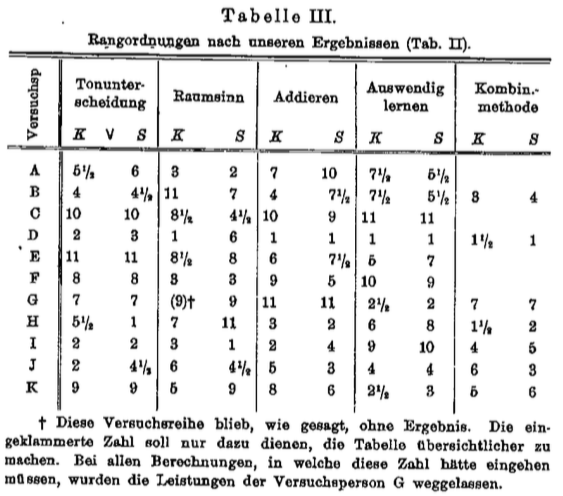
*1. The raw correlation values*

The raw results of our experiments are presented in Table II.



First, the multiple values obtained for a subject’s performance on a specific ability had to be reduced to a single value. When such values appeared homogeneous enough, we simply took their sum; e.g., for tactile thresholds, we added up the threshold values for the hand and both cheekbones in each subject. When, on the other hand, the values to be combined were completely heterogeneous, we rendered them comparable by converting them into *ordinal values*. E.g., concerning the speed of addition, the subject who had added up the highest number of digits was assigned the ordinal value of 1; the subject with the next highest number was assigned the ordinal value of 2; and so on. Addition accuracy was handled in a similar manner. We then used the sum of the two ordinal values for each subject.

The single values thus obtained for each participant and each ability domain were eventually converted into ordinal numbers as well. Only in this manner is it possible to make the values from one ability domain comparable to those from another. Thus we obtained two ordinal rankings for each of the five abilities tested – as shown in Table III; one is derived from Krueger’s values, the other from Spearman’s.



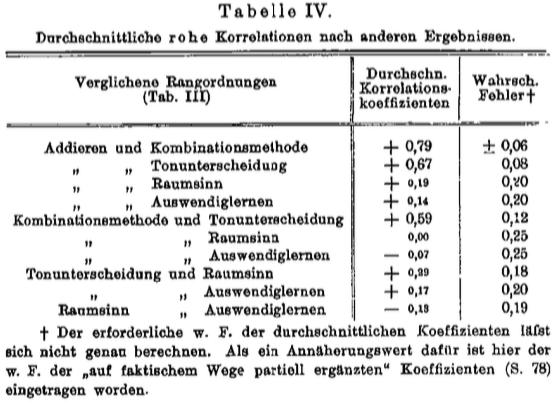
Finally, the correlation coefficients for each pair of rankingswere calculated using the method presented earlier (p. 52–53).

Before discussing the resulting correlations in more depth, it should again be emphasized that the experimental series were short. *While this does not render the results invalid, it makes them less exhaustive*. Specifically, at the outset we have to disregard their *minor* features – whether they are consistent with our interpretations or not – as presumably only random. We only pay attention to numbers and differences which are at least double the probable error (or probable difference). To this end we use means as much as possible, and specifically means taken across all relevant cases.

Note that according to the formula, the probable error is the lower, the higher the respective correlation is. It should however be added that this equation presupposes very long series of experiments with real-numbered values; for ordinal numbers and short series, as in our case, the equation remains sufficiently accurate only for low to moderate correlation coefficients; for *very* high correlation coefficients, even though the probable error decreases with increasing correlation, this happens to a much lesser degree than required by the equation. Unfortunately, the probable error is higher for all correlations involving Ebbinghaus’ combination method, because, as mentioned above, we had fewer subjects at our disposal.

Since every ability was measured twice, we obtained *four* correlations for each pair of abilities. Let us say *A* and *B*: one value for the correlation between Krueger’s results for *A* and his results for *B*; the corresponding value for Spearman’s *A* and *B*; a third value for Krueger’s *A* and Spearman’s *B*, and finally a value for Krueger’s *B* and Spearman’s *A*.

At first we took the *mean* of these four values. This results in the mean raw correlations between the various abilities, which are presented in Table IV next to their probable errors.



Seven out of these correlation values (printed in a smaller type) are clearly not greater than double their probable error; thus in all these cases *no* correlation is apparent at the outset (only later will we be able to ascertain whether a correlation could possibly be present using the disattenuation and correction formulas). The other three correlations, on the other hand, exceed their respective probable errors more than five-fold; with average values of 0.7, 0.79, 0.67, and 0.59 they can thus be regarded as rather certain.

Before going further, we first have to check whether these high correlations exist only for abilities tested on the same day by the same researcher. Because if, let us assume, the correlation between two abilities *A* and *B* largely or completely disappears when *A* is being tested on one day by one researcher and *B* one week later by the other, then any correlation determined by a single researcher could be reduced to the momentary moods of the subjects; it would stem from subjects who happen to be in a good working disposition performing rather well on both tasks compared, *A* and *B*. However, this hypothesis is not supported by the results: the mean of the three reliable correlations as determined by a single researcher is 0.68; and the mean of all other values for these three correlations, i.e., those coming from two different researchers, comes to 0.67; thus they are almost equal. This also removes the suspicion that the measurements could have been influenced by the researchers’ falling prey to “autosuggestion”; because neither learned anything about the other’s results before the end of the study.

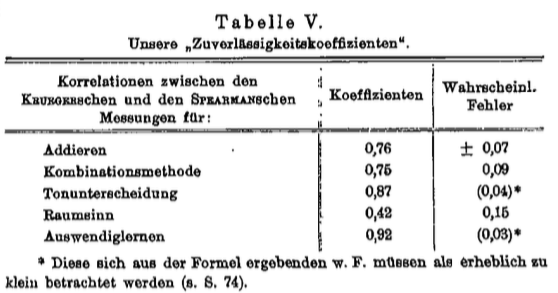
Another question is of course whether the correlation values obtained in the second series of measurements (Spearman) show different patterns than those obtained in the first one (Krueger).

Specifically, Binet has arrived at the view that associations between such simple abilities should strongly decrease with each repeated test, and could eventually even disappear. Our experiments have not revealed any tendency of this kind. The mean of the three correlations under consideration was 0.61 for Krueger’s results, and 0.75 for Spearman’s.

*2. Application of the “disattenuation formula”*

To use the disattenuation formula, a “reliability coefficient” for the measurement method for each ability has to be calculated; as mentioned above, in our case this is achieved by calculating the correlations between Krueger’s and Spearman’s measurements for each ability.

This results in the values presented in Table V.



Another point needs to be considered here, which, while lacking practical significance in the present case, could be of importance under other circumstances. Namely, our reliability coefficients were calculated from two series of measurements acquired on different days. Thus they are strictly comparable only to those correlations between two abilities for which the measurement series were also acquired on different days; i.e., here we are not allowed to use correlations between two series measured by Krueger or two by Spearman, but only mixed correlations between one by Krueger and one by Spearman. In this case, the disattenuation formula simplifies to the following:



When applying this simplified disattenuation formula, e.g., to the correlation between addition and pitch discrimination, we obtain the following equation:



Thus the latter value is to be regarded as the pure correlation between addition and pitch discrimination obtained by disattenuation. In the same way we obtain the pure (i.e., disattenuated) values of 0.93 and 0.81 for the correlations of the combination method with addition and pitch discrimination, respectively. The mean of these three correlations has thus increased, by using the disattenuation formula, from 0.68 to 0.85; this latter value would be expected if the ability of each subject had been determined without errors using long series of and extensive measurements.

A certain approximation to this theoretically complete disattenuation can be achieved by averaging across Krueger’s and Spearman’s measurements, then calculating the correlations between these series of means. This results in the three correlation coefficients of 0.73, 0.84, and 0.64; their mean has now increased to 0.74. This value, partially disattenuated based on empirical data, thus in fact reaches a size halfway between the raw value, and that fully disattenuated with the formula, just as predicted by the theory.

Thus, in conclusion these results seem to show *that pitch discrimination in fact has a high correlation with abilities that appear as fundamentally different as that of addition and that on which Ebbinghaus’ combination method is based. There is also a high correlation between the latter two abilities.*

Finally, let us now turn towards the correlations that appear vanishingly *low*. It can be seen that these are all those in which either the tactile resolution or rote memorization takes a part. Now, as we have discussed above, such an apparent absence of a correlation is initially ambiguous; it can either point to an actual absence, or merely derive fromexcessively largerandom errors. The latter case would be reflected invery low reliability coefficient, i.e., a very low correlation between Krueger’s and Spearman’s values for the same ability (see Table V).

We find that in this regard, tactile thresholds and rote memorization exhibit curiously opposite behavior. In the case of the tactile thresholds, Spearman’s and Krueger’s measurement series match up so imperfectly, their correlation is only 0.42. This alone provides sufficient reason for the low apparent correlations between the tactile resolution and the other abilities; because when inserting this value (0.42) into the disattenuation equation and additionally considering the range of chance fluctuations suggested by the probable error, one finds that the tactile resolution could *potentially* have a rather *high* pure correlation with the other abilities. The current experiments remain inconclusive on this point.

There are two possible reasons for the large discrepancy between the series acquired by the two researchers. On the one hand, it could be due to the two measurements in each subject having been carried out on different days, and thus been subject to different conditions regarding disposition, fatigue, temperature, etc. On the other hand, the measurement method could have been insufficiently accurate for our purposes. Because the potential appearance of such an ambiguous discrepancy could be easily foreseen, we determined the thresholds of the right as well as the left cheek bone as a control measurement (see p. 69); had the aforementioned discrepancy only been due to a change in conditions on the second day of experimentation, then each subject’s two cheekbones would still have yielded closely corresponding values for the same researcher on the same day. However, this is not the case; the correlation between the sensitivities of the right and left cheekbones as measured by the same researcher is only 0.36 (averaged across Krueger’s and Spearman’s data). So here we see an example of how easily a measurement method can be insufficiently accurate to allow any correlation, even one which is actually present, to appear.

The case is completely different for *rote memorization*; here the correlation coefficient between Krueger’s and Spearman’s values is no less than 0.92, i.e., almost perfect. In this case the disattenuation formula cannot produce any noticeable increase in the correlation between this ability and the others; if no correlation is revealed here, then none is actually present. This leads to the important result that *the rote memorization of series of numbers has in fact no significant correlation with any of the other abilities tested*.

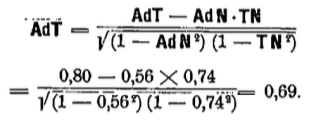
*3. Applying the correction formula*

Finally we have to see in how far these correlation values, despite the cautionary measures mentioned above, have to be regarded as disturbed by *constant* extraneous factors.

Almost all of these factors could easily be ranked according to their weight, which varied from case to case; this was possible either based on objective measurements (temperature etc.) or on the comments made in the protocol.

First we wanted to look at the case in which the extraneous factor may affect *both* abilities compared. This case is especially risky because it *increases* the correlation and can create the illusion of an association where none exists. Here the most relevant factors appeared to be the subjects’ nationality, age, general state of health, motivation, and fatigue or other momentary states of being strained.

The most problematic factor is nationality. Both addition and pitch discrimination showed significant correlations, of 0.48 and 0.68, respectively, with nationality; specifically, the Germans outperformed the foreigners. By using the disattenuation formula, the first value increases to 0.49/M(0.76, 1.00) = 0.56, and the second one to 0.68/M(0.87, 1.00) = 0.74. Initially, a considerable illusory increase should be suspected here; because in applying the correction formula (p. 59), we obtain:



Accordingly, the real value of **AdT** would probably not be 0.80, but only 0.69. However, if we leave out the foreigners’ results and re-calculate the correlation coefficients exclusively for the German subjects, we do not find even the slightest decrease. From this we can conclude that the nationality-based difference in performance is not actually a new factor randomly confounding the association investigated, but rather itself originates from this association. However, here this question can only be touched upon as a suggestion, since the relevant associations between nationality and the various abilities are subject to a very large probable error of 0.28 (this is due to one variable – the nationality – assuming only two possible values, either “German” or “foreign;” thus the ***r***-method was not applicable, and another much less reliable method had to be used instead. See *Am. J. Psych*. 15, p. 85).

The *age* (20–38 years) did not show a noticeable correlation with any of the five abilities tested (the greatest correlation found was 0.22; the average across all positive and negative correlations was 0.02).

Concerning *health*, it seems that the subjects had only one issue: about half of them reported being somewhat “nervous” or “neurasthenic.” However, it turned out that these self-reports regarding neurasthenia etc. had no noticeable correlation with any of the abilities; the mean for all five correlations was zero (here again the calculation method used in the case of nationality was required).

Nor did any of the abilities show any noticeable correlation with the apparent *motivation* of the subjects. The mean value of these five correlations was -0.03. This result is even more remarkable given that Krueger’s and Spearman’s observations concerning the apparent motivation of the subjects showed the relatively high correlation of 0.66, despite having been carried out on different days.

We eliminated *fatigue* and other forms of *strain* as confounding factors by never testing subjects in these states in the first place.

Thus, in any case it seems the correlations are not magnified in an illusory manner. Let us now consider the case of a possible association between an extraneous factor and *only one* of the abilities to be compared. In this case, as mentioned above, the correlation would appear lower than it is.

Regarding the tactile resolution, we particularly considered the influences of physical conditions, which is why we recorded the room temperature (fluctuating between 12.7°C and 17.7°C ), the outdoor temperature (between -1.6°C and +6.1°C ), humidity (40–76%), and air pressure (731–764 Torr). However, none of these circumstances showed any noticeable correlation with the tactile resolution (their correlation coefficients with the latter were 0.04, -0.04, -0.12, and 0.06, respectively).

The correlation between addition ability and a subject’s degree of engagement with mathematics similarly appears insignificant. Here we obtained a coefficient of only 0.22, which is less than double the probable error; apart from that, by inserting this value into the correction formula (c) on p. 60, we obtain



Thus this extraneous factor would have decreased the pure correlation by only 2%, which in our experiments is negligible.

The correlation between pitch discrimination and a subject’s engagement with music turned out to be more significant; despite having excluded both extremely practiced and unpracticed subjects, as described above, here we still found a correlation of 0.46. When inserting this value into the correction formula, we obtain



According to this result, the correlations between pitch discrimination and the other abilities would have been unduly decreased by about 11%. This number may have a certain significance even for such short experimental series; however, as we will see shortly, it has no effect on the arguments developed in this article. In any case there seem to be more reasons for regarding the reported correlations between pitch discrimination and the other abilities as too low, than for seeing them as too high.

*5. The “central values” of the various abilities*

Only now, after having thoroughly tested the purity of the correlations observed, is it permissible to use them for further scientific deliberation.

The question is above all whether the pairwise correlations between abilities are to be understood as isolated phenomena, or whether all of these correlations could rather derive from a *single* cause (it should be noted that previous work has discovered similar correlations between nine other mental abilities).

At first we would like to make use of the common-sense and at least for heuristic purposes valid hypothesis that any two correlated mental abilities may be regarded as dependent, in part, on a *common* underlying factor. This hypothesis leaves space for developing various more specific analyses later; e.g., the two abilities could be partly mediated by the same physiological organ; or their respective organs could be subject to a common influence; or the common factor could be of an even more indirect or abstract nature.

Let us assume that two abilities, *P1* and *P2*, share the common underlying factor **P**, and two other abilities *Q1* and *Q2* in turn share the factor **Q**. If the correlations *P1P2* and *Q1Q2* derive from the same cause, then this is equivalent to saying that *P1* and *P2* have the same underlying factor as do *Q1* and *Q2*, or, in other words, that **P** is the same as **Q**. In this case the effects of **P** and **Q** have to agree perfectly, i.e., the correlation **PQ** has to be equal to 1. Thus we have reduced the problem to a mathematical one; we only need to determine the correlation **PQ** and see whether it is in fact always equal to 1.

Now, excellent performance in ability *P1* seems to suggest that **P** is also rather well-developed in the given person (since according to our hypothesis, *P1* partly depends on **P**). A measurement of *P1* is thus simultaneously a measurement of **P**, although one possibly fraught with very large random errors. In a similar manner, *P2* also constitutes a measurement of **P**, and *Q1* and *Q2* are measurements of **Q**.

Random errors, however large, can be compensated for with our disattenuation formula. In short, if all correlations stem from a single cause, then



must always hold. Since *P1*, *P2*, *Q1*, and *Q2* represent arbitrary abilities, we can replace them with the letters *A*, *B*, *C*, *D*. Then we obtain



or, eventually,



where the letters *A*, *B*, *C*, *D* represent arbitrary (not too closely related) abilities compared.

Now this equation can readily be applied to our results. To this end, in this case we can go back to the *raw* values of the correlation coefficients, because the disattenuation carried out on pp. 77–78 and the correction discussed on p. 82 would only multiply both sides of the equation by the same factors. While corrections carried out according to Eq. (b) on p. 60 would affect the equation, they have turned out to be unnecessary with our results. By returning to the raw values, we not only circumvent any uncertainty regarding the correction methods, but also make it possible for the tactile thresholds to be taken into account.

Representing the five abilities tested by their initials – if the hypothesis about a single common factor were true – it would have to hold that, e.g.,

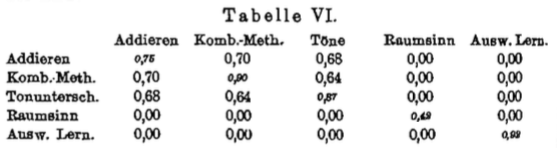


or, when inserting the actual values,



The right side comes to 0.39 and the left side to 0.46, the difference thus being 0.07. Using the complete set of correlation values given in Table IV, fifteen equations of this kind can be formulated; we find mean and maximum differences of 0.08 and 0.15, respectively. Since the probable random difference is about 0.10, our results, as far as they go, stand in good agreement with the theory that all correlations derive from a *single common factor* – which we shall call the “*central factor.”*

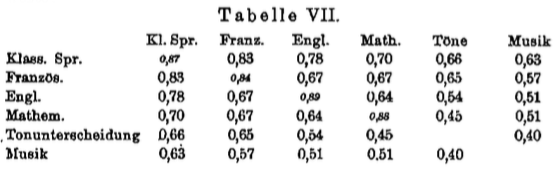
For short series of experiments, this hypothesis regarding a single central factor can be tested in a more efficient and perspicuous way using another mathematical relationship that is easily derived from the one just discussed. When any correlation *AF* is greater than the correlation *BF*, then – according to our hypothesis – every other correlation *AG* (or *AH*, *AI*, …) must also be greater than *BG* (or *BH*, *BI*, …); in the case of *BG* (or *BH*, *BI*, …) being zero, *AG* (or *AH*, *AI*, …) must also be zero. Based on these relationships, it must be possible to bring all correlation values into a hierarchical spatial arrangement in which each value (except for zero values) is greater than all values to its right or below it; and whenever there a single zero value comes up, the remainder of the relevant row or column must also be filled with zero values. Let us check this for our results (setting all correlations lower than double the random error to zero). For comparison, we would like to include the reliability coefficients as well (in *italics*). This gives us the following picture:



In creating this table (just like in applying the disattenuation formula) we have used only correlation coefficients derived from measurements taken on different days; for only these coefficients are, as discussed above, strictly comparable with the reliability coefficients. The latter form the downward diagonal in the table, which so sharply contrasts with the course of the other correlations (between *different* abilities).

One might think that such hierarchical arrangements could easily be produced for random combinations of values as well. However, this is not the case. E.g., the arrangement of the above values would have been impossible if *any one of the three* correlations which are high – according to our results – had been equal to zero; and equally so if *any one* of the zero values had been replaced by a high correlation.

For comparison we would like to cite similar, but more extensive results from an earlier study. The study had 36 school children as participants. Four of the abilities compared were school subjects, specifically classical languages, French, English, and mathematics; the relative abilities of the pupils were determined based on their school exam rankings. A fifth ability compared was musical talent; the children were ranked in this respect by a music teacher. Finally, as a sixth point, their pitch discrimination ability was tested. This resulted in the following correlation values:



It is evident that the hierarchical arrangement discussed above can again be obtained (the only small exception presumably remains within the random fluctuation range).

Using the hypothetical central factor and the disattenuation formula, we can further determine a value of particular interest. Because as soon as one has acquired two measurements series *A1* and *A2* for any ability **A**, and two measurement series *B* and *C* for two other abilities, respectively (one series each), it becomes possible to determine a *constant* value related to **A**; namely, its correlation with the central factor, in short its “*central value.*” Comparisons of **A** with any two other abilities **D** and **E**, at any time and accuracy level, should – according to the central factor hypothesis – *always yield this same central value*.

The procedure is the following. The series of values *B* and *C* can be conceptualized as measurements of the central factor, let us say **Z** (see Note 1, p. 84); and the series *A1* and *A2* are by definition measurements of **A**. Then the sought constant correlation value **AZ** is defined by the following equation:



Tables VI and VII enable us to immediately test for this relationship, because the ability of pitch discrimination was tested both in the current experiments and in the older study conducted under completely different circumstances in school children. By inserting the values from Table VI into Eq. (g), for the present experiments we obtain



In the older experiments, finding **TZ** is complicated by the fact that at that time only *one* series of measurements was taken, so that the value *T1T2* cannot be determined accurately. We must use a freehand estimate of *T1T2* which was already produced and inserted into the disattenuation formula at the time (for details see *A. J. P*. 15, 1904, p. 65); this estimate came to 0.64. By inserting this value and the correlations (Table VII) of pitch discrimination with classical languages and English, respectively, into Eq. (g), we obtain



If instead of classical languages and English we insert French and mathematics (again Table VII) into the equation, this results in



Such a *perfect* agreement of the current central value for pitch discrimination with the two older ones is, given the present circumstances, of course a chance occurrence. However, it is important that these applications of Eq. (g) have at least not yielded any strongly inconsistent values. And of even greater importance is the fact that this equation provides an accurate and very convenient control measure for more extensive experimental series in the future.

When applying the equation to the current results on addition and the combination method, we obtain even higher central values: **AdZ** = 0.97; **KZ** = 0.97. Such high values suggest that the abilities concerned stand in very close relationship with the central factor; thus they appear particularly suitable for helping to more closely elucidate the nature of this curious central factor, which seems to be involved, to some degree, in all abilities studied so far (except rote memorization). We will return to this point below.

**IV. Oehrn’s results**

*1. Oehrn’s experimental methods*

The association between various simple abilities has been the subject of quite a high number of previous experimental studies. Since most of the conclusions reached above are of a very general nature, they should – one might think – hold for older results as well, if only one could submit them to the same mathematical analysis.

However, unfortunately older studies rarely communicate their original results; we may be given elaborate and ingenious tables, but precisely the data that would allow us to calculate the correlations accurately are missing. Further, most of the results are muddled by various apparently highly disturbing extraneous factors.

A brilliant exception in both respects is the very first study of this kind ever conducted, Oehrn’s elegant work on “individual psychology.” The author documents his raw data in detail, and from the outset focuses on avoiding “greater differences between the subjects with respect to age, education, and habituation to the experimental conditions.” His experiments show a level of thoroughness rarely matched by his successors in the field.

We would now like to submit Oehrn’s results to the same treatment used on our data. Should we succeed at deriving new and important regularities from these highly reliable and widely known (for seventeen years) data, then our methods would have passed a decisive test. Additionally, should these regularities be in consistent agreement with the results of our own experiments, then this would confirm the latter in a remarkable way.

Let us first briefly describe Oehrn’s experimental methods. Each subject was tested individually.

A. “In order to learn about the individual nature of the *process of perception*” the subjects were tasked with counting letters. Each subject received a copy of the same book with the instruction to count its letters as quickly as possible, word by word and line by line, starting at a designated paragraph. In evaluating the results only the speed of counting was taken into account, not its accuracy.

a) In the first series of experiments, successive letters were to be counted individually. Whenever 100 letters were counted, the relevant place was marked with a pencil and counting resumed from 1.

b) It turned out to be impossible to count without pronouncing the numbers at least in thought. In order to eliminate the delay this caused as far as possible, in a second series of experiments Oehrn modified the letter-counting procedure in such a way that groups of three letters each were to be counted. Each pencil trace thus marked 100 letter triplets.

B. “In order to study *memory*,” Oehrn asked his subjects

a) to memorize series of nonsense syllables. Memorization took place by reading each series from beginning to end as many times as necessary until the subject could recite it once without errors. After each incomplete reproduction, the missing part of the series was read again. Then the number of syllables memorized per 5 minutes of study was calculated.

b) Another series of experiments was conducted in a similar mannerusing series of numbers instead of syllables.

C. Addition of single-digit numbers was used to study “the *associative process*.” The subjects were instructed to add up series of numbers presented in notebooks as quickly as possible. Errors noticed by the subjects had to be corrected; no other control of accuracy was undertaken (due to technical difficulties).

D. “*Motor functions*.”

a) Writing to dictation. A paragraph of a German book was dictated by Oehrn and written down by the subject at the highest possible speed. Only the speed was taken into account in the evaluation.

b) Reading. For these experiments Oehrn used Freitag’s “Pictures from the German past,” which was printed very clearly and in German script. Subjects were asked to read out loud at maximum speed; however, no syllables were allowed to be “swallowed;” all had to be fully pronounced (in a low voice or whisper). Every 5 minutes the current location in the text had to be marked on a signal.

*2. The effects of practice and fatigue on the correlation values*

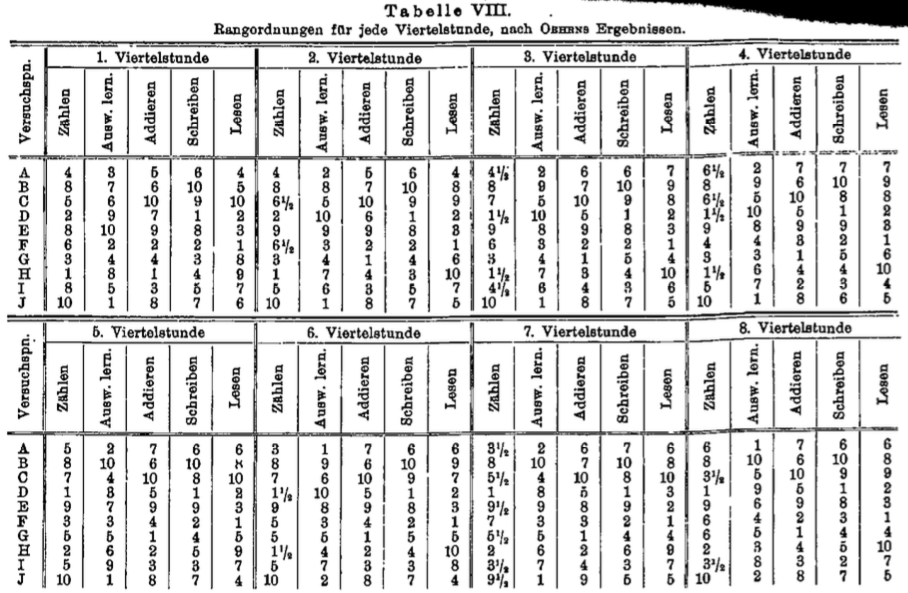
Oehrn himself processed his experimental results in an exceptionally careful and thorough manner. However, since he was proceeding in the usual manner without determining correlations mathematically, he could only draw the most general of conclusions. His overall findings regarding the correlations between the different abilities are limited to the following sentences:

“We see that the functions studied naturally form two groups with respect to absolute duration: while each subject’s levels of performance in counting letters, addition, and motor functions are rather proportional to each other, the same subject’s behavior is significantly different with respect to rote memorization. When ranking the subjects according to the quantity of their performance, for the former functions their order is preserved with only minimal deviations, whereas it changes completely for rote memorization.”

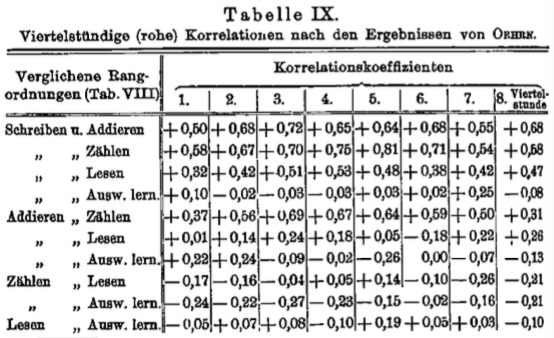
Let us now try to evaluate these same experimental data in more depth using the correlation and disattenuation formulas.

Above all we would like to take advantage of the fact that Oehrn tested each participant continuously for two hours with each of the methods described above.

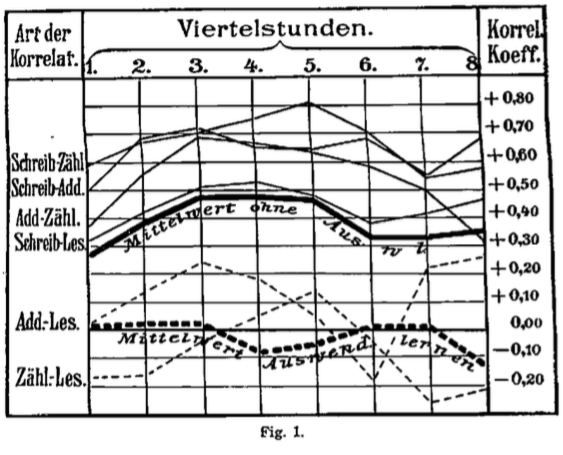
Pages 134–135 of his study show the results for each quarter-hour of testing. Just like in our own experiments, at first we converted these real values into ordinal numbers; thus we obtained rankings of all subjects in each domain of ability for each quarter-hour. We then added up each subject’s ordinal numbers for counting single letters and letter triplets and these sums were then again converted to ordinal numbers; in this way we obtained a single ranking for letter counting in general. At the outset we left aside the rankings for the memorization of syllable series because Oehrn tested subjects with this method for only six quarter-hours, not eight, as he did with all other methods. We eventually produced five rankings for each of the eight quarter-hours, one for each of the following abilities: counting letters, memorizing number series, addition, writing, and reading; the rankings are presented in Table VIII.



We then calculated all correlations between these five rankings for each quarter-hour, and obtained the sought correlation coefficients — shown in Table IX— and their probable errors. Because Oehrn’s sample, just like ours, was not very large, we again to have to constantly keep an eye on the size of the probable error, in order to avoid ascribing meaning to details which could easily be mere chance artifacts.



For the sake clarity we have presented these results in a graph (Fig. 1). It is striking that the *rote memorization experiments* again yield in negligibly low correlations; the bold dashed line representing the mean of all correlations involving rote memorization never significantly deviates from the zero line, and thus from complete independence.



Out of the three other correlations involving *reading*, two remain smaller than double the probable error throughout; even though the third one crosses this threshold, it still significantly lags behind the three remaining correlations.

The remaining three correlations, finally, which only include writing, addition, and counting, have very high values, and all of them exceed the probable error five-fold.

Despite these differences between the correlation curves with respect to their absolute height, their general *relative* time course (from quarter-hour to quarter-hour) is similar (of course excluding the curve for rote memorization). In all cases the second quarter-hour yields higher correlations than the first, and the third one again surpasses the second one. Further, there is always a clear decrease in correlation in the sixth, or at the latest, in the seventh quarter-hour, only to give way to a final increase at the very end in most cases. Thus, on the whole we recognize the familiar shape of the *practice-fatigue curve* in these correlation values.

As mentioned, Binet voiced the view that such correlations manifest clearly only under unfamiliar experimental conditions; with continued practice they would quickly decrease, or sometimes even disappear. However, one of us arrived at the opposite conclusion; when investigated using an exact method, practice (at least in its first stages) actually increases these correlations.

The results just presented support the latter view. *With increasing practice and habituation, the correlations increase without exception.* Thus unfamiliarity is so far from causing correlations, that — on the contrary — it seems to significantly decrease them, either directly or indirectly.

One possible reason for this reduction is quite evident: it is known that results tend to be subject to greater random error under still-unfamiliar experimental conditions; and this, as we have seen, reduces correlations. However, such a reduction can be precisely quantified using the “disattenuation formula:” and in the present case, it turns out to be far from sufficient to explain the differences observed.

Thus, it seems we have to regard *unfamiliarity* as a direct disturbing factor; i.e., some subjects suffer more from it than others, which leads to a slight divergence of the performance values from the correlation that is otherwise present between the abilities.

“During the second hour,” notes Oehren, “*fatigue* usually started to show a pronounced effect.” So here we have another disturbing factor with an effect resembling that of unfamiliarity, as it also affects the subjects in an unequal manner. This factor is also clearly revealedin the plotted curves.

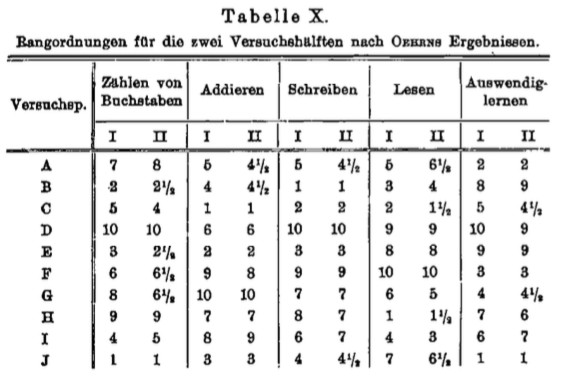
In the last quarter-hour, the “drive” introduced by Kräpelin is expected to help subjects to, in part, overcome the fatigue. And in fact, now the correlations rise accordingly.

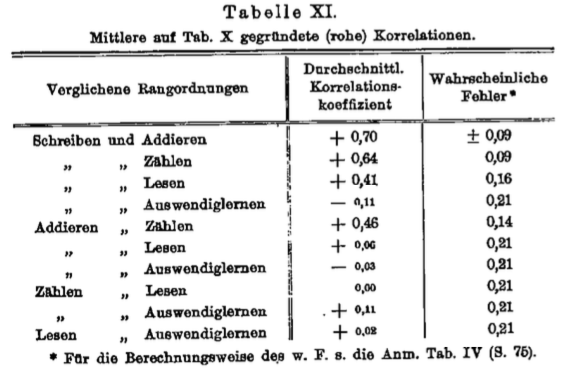
Finally we wanted to have a closer look at the curve for the rote memorization experiments. Despite its modest deviation from the zero line it shows a fairly striking and regularity in taking, curiously, exactly the course of an *inverted* practice-fatigue curve. However, a closer look at this paradoxical phenomenon only provides yet another confirmation of what has been said. Because assuming – based on all reliable results so far – that the correlation between rote memorization and the other abilities is close to zero *in general*, empirically it can still reach positive or negative values (with equal probability) of a certain magnitude for small samples randomly selected as representatives of the whole class. Such a correlation has to be disturbed by extraneous factors in the experimental setup, i.e., by unfamiliarity and fatigue; this occurs in equal measure regardless of whether the correlation happens to be positive or negative. In the latter case, the practice-fatigue curve should appear on the negative (lower) side of the zero line with inverted cusps, just as is the case in Fig. 1.

This thoroughly lawful pattern of the correlation coefficients even in Oehrn’s data, where the author had no knowledge of said coefficients, *should remove any doubts regarding the applicability of these calculation methods (when handled appropriately) to such short experimental series*.

*3. Application of the disattenuation formula*

In order to apply the disattenuation formula, we need two series of measurements for each domain of ability. These are most easily obtained by taking the means across the first and second halves of the experiment, to be used as the first and second series, respectively. This yields the rankings shown in Table X. These form the basis for calculating the correlation coefficients, whose means are presented in Table XI.





The resulting overall picture resembles that found in our own experiments. Six of the correlations (in smaller type) are less than double their probable error, and thus negligible. Two correlations exceed this threshold slightly; and two are again very high. The mean of the four significant correlation values is 0.55. After fully eliminating random errors with the theoretical method used above, this value increases to 0.62.

Note that the disattenuation formula does not significantly change the correlation coefficients with values as reliable as Oehrn’s. Nevertheless, it is still necessary to apply the formula even in such cases, precisely to verify such a sufficient reliability has been achieved.

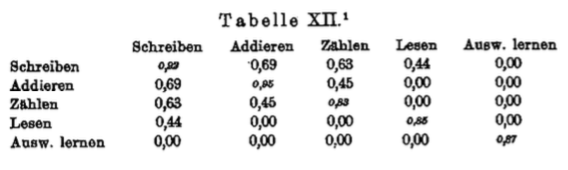
When trying to draw closer comparisons between Oehrn’s results and ours, we find only *one* correlation that appears in both investigations: the correlation between addition and rote memorization; and the same correlation value is in fact found in both cases, namely zero.

Further, the “reliability coefficients” for addition and rote memorization appear in both studies. In Oehrn’s experiments the value for addition is 0.88, whereas in ours it is only 0.76. This suggests our testing method may be less reliable than Oehrn’s. The reason for this discrepancy is easily found, as there was only one essential difference between the two methods: we included both the speed and the accuracy of addition in our evaluation, whereas Oehrn took into account the speed only. However, nothing stops us from carrying out our calculations based exclusively on speed, and thus, removing any reason to expect a difference in reliability; in doing this, our reliability coefficient becomes 0.89, which is in fact almost identical with Oehrn’s.

A comparison between the two rote memorization methods is more interesting. Oehrn’s method was based on the number of stimulus presentations necessary for one error-free reproduction; our experiments, on the other hand, worked with the level of correct reproduction achieved after a single exposure. Despite this difference, the reliability coefficients are almost identical and very high for both methods: 0.95 for Oehrn’s and 0.92 for ours.

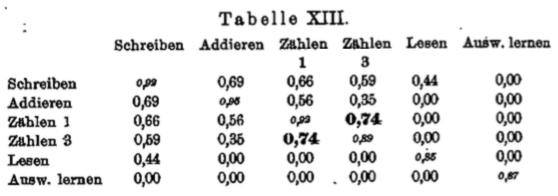
On the other hand, the reliability coefficient for Oehrn’s method using series of *syllables* was very low – only 0.49. It was specifically by calculating the reliability coefficients for Oehrn’s results that we could establish the much higher reliability of the number series method at the outset, and thus, choose it for use in our own experiments. Series of syllables, despite their other well-known advantages, would have delivered the kind of ambiguous results we have in fact obtained for tactile thresholds; in which case, our whole study would have failed (because unambiguous correlation values for at least *four* domains of ability were indispensible). Thus we see the practical importance of reliability coefficients.

*4. The “central values” derived from Oehrn’s data*

Oehrn’s correlations are readily brought into the hierarchical arrangement described above (p. 86). Including the reliability coefficients (in *italics*), we obtain

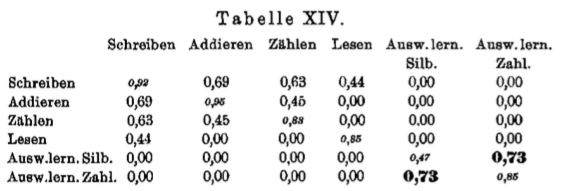
We see that each value (except for the zero values) is greater than all values to its right or below it.

Here we have an opportunity to see how such a hierarchical arrangement becomes impossible as soon as two abilities are related in any way *apart from* sharing the central factor. As mentioned, Oehrn studied counting ability in two ways: at first as counting of subsequent single letters, then of letter triplets. Based on differences that emerged in the course of careful self-observation, he consistently treated these two abilities as equally separate from each other as from the other abilities; only we combined them. If we separate them again, the hierarchical arrangement immediately becomes impossible; we obtain



The counting of single letters and that of letter triplets show a much too high mutual correlation to fit into the general hierarchy. This excess over what is consistent with the hierarchy points to a *special* connection, beyond that mediated by the central factor.

We find the samein an even more striking manner if we – like Oehrn – treat the rote memorization of syllables and that of numbers separately. We then obtain Table XIV, where the correlation between the memorization of syllables and that of numbers is incompatible with its place in the table.



The two types of rote memorization, despite having little correlation with the other abilities, show a mutual correlation as high as 0.85 even before applying the disattenuation formula. After disattenuation, this correlation value does not significantly differ from 1. This particularly high correlation between the memorization of numbers and that of syllables suggests that a rather large group of abilities may be related closely enough to be subsumed under the concept of “rote memorization” as a more or less unitary ability.

Usually one finds, on the contrary, a surprising absence of any special connection between two types of abilities, even when they seem to closely resemble each other at first glance.

For example, based on careful self-observation, Oehrn subsumed both speed writing and speed reading under “motor functions.” As a result, based on the fact that both abilities showed a clear association with addition, he drew the general conclusion that “motor functions” are “quite proportional” with “the associative process.” However, a second look at Table XIII shows writing actually has its lowest correlation with reading (of course apart from rote memorization); and even more importantly, the correlation between reading and writing in no way stands out from the hierarchy, as is the case, e.g., with counting single letters and letter triplets. So two processes falling into the same general category, such as “motor functions,” etc., is by far no justification for assuming a corresponding correlation with respect to their actual time course.

Next we would like to determine how far the “central values” derived from Oehrn’s data agree with our own. We have already seen that these values – according to the hypothesis – should remain constant under all circumstances.

Two abilities appear in both studies and thus produce comparable values: rote memorization and addition. Regarding the former, we readily obtain the desired correspondence; both studies yield a central value of zero.

Regarding addition, on the other hand, inserting Oehrn’s data from Table XII into Eq. (g) (p. 88) yields



Here we initially notice a considerable deviation (by 0.21) from the central value for addition found in our own experiments, which was 0.93. However, we have to remember that Oehrn again took into account only the speed of addition, whereas we included both its speed and accuracy. The speed and accuracy of a function are two different things. Thus, the comparison just made is not valid in the first place.

If we now set up a *pure* comparison by similarly including only our values for speed in the calculation, we obtain



We see that this value corresponds very well with that derived from Oehrn’s data.

**V. Interpretation of the overall results**

We now have to approach the important and difficult question of the *interpretation* of these initially only quantitative relationships; to attempt an assessment of the common *cause* of the correlations established, or, using our terminology, a closer appraisal of the nature of the “central factor.”

At the very least, our results seem to allow us to reject a number of common-senseexplanations, some of which have been suggested previously, for good.

E.g., one experienced psychologist and educator has ascribed the following composition to this cause or central factor in a letter to one of us: “the ability to comprehend instructions, to work with attention and motivation, and to take advantage of all available aids.” This view, despite being quite commonsensical, becomes untenable when confronted with our finding, which states that writing to dictation, addition, and pitch discrimination have high correlations with each other and the other abilities, whereas rote memorization shows no noticeable correlation with any other ability. As it is impossible to claim that motivation, the quick comprehension of experimental conditions, the use of aids, and above all the level of attention are crucial for all other abilities, but not noticeably used specifically in rote memorization! Rather, all previous studies have so far consistently found the opposite. According to Oehrn himself, e.g., when writing to dictation, the central processes are so deeply ingrained they “mostly proceed just as automatically in the absence of conscious attention.” Further: “Finally, regarding rote memorization, no elaborate justification should be needed to argue that it requires a high level of attention, and more so than all other functions.” In order to memorize series of numbers and syllables, “we have to try to grasp with intent attention that which the eye and ear present to us.” The “comprehension of instructions” and “use of aids” are also sure to play a greater role in memorization than in writing or pitch discrimination.

Binet also assumes that simple abilities, like the ones discussed here, are to be regarded as “tests of intentional attention.” However, he adds that he does not use the word “attention” in a completely conventional way, but rather refers to the function that allows us to cope under unfamiliar conditions. Nevertheless, even using this redefined sense of the word, the explanation remains equally at odds with the empirical results discussed above. Oehrn incidentally mentions the first test for all subjects was letter counting; it was during this test, he says, that “habituation to the conditions of the experimental setup” occurred; still, this ability yielded only moderate correlations. Further, Oehrn states: “Practice has the smallest effect on writing, and the greatest on the memorization of numbers;” and it was just writing and number memorization that showed the highest and lowest correlations, respectively. However, the main point is that we have been able to establish empirically that the correlations, rather than decreasing with habituation and practice, on the contrary, even significantly increased in every domain of ability. Therefore, as we have seen, it is necessary to regard unfamiliarity not as a cause of the correlations, but rather as an extraneous factor.

We have further shown that the momentary “disposition” of the subjects cannot be a significant factor either, as the pairwise correlations between abilities did not decrease noticeably if one was tested a week after the other.

On the whole, we have every reason to believe the central factor does not merely show an accidental connection with the relevant abilities, but rather stands in a very close functional relationship with them.

Now it is much harder to construct a new, well-founded interpretation, than to refute previously attempted or common-sense explanations of the correlations.

The most central positive evidence appears to be the following. First, there is the curious contrast between the development of new, arbitrary associations between numbers – rote memorization – on the one hand, on which the central factor shows next to no influence; and, on the other hand, the operation of previously acquired associations of numbers that are interlinked in complex ways – addition – in which the central factor appears crucial. Second, one is struck by the surprising *psychological heterogeneity* of the abilities that have revealed the closest functional relationships. We found very high “central values” for the so-called sensory ability of pitch discrimination, the near-automatic motor function of speed writing, and likewise for academic success at school, which represents a much higher level of mental achievement.

It should also be mentioned that some previous experiments ranked 24 school children according to their reputation of “intelligence” among their classmates. One student was selected and asked: “Who do you think is the brightest of your classmates?” Then: “And apart from this one, who is the brightest?” And so on, until a complete ranking was produced. For validation, a second student and a lady who knew all children well were questioned in an identical manner. The ranking produced in this way showed a near-perfect correlation with the central factor.

Let us try to follow up on these vague hints. On the one hand, the exceptionally high central value of “intelligence” points towards the central factor strongly affecting the reputation that a person enjoys among his peers regarding his mental acuity; on the other hand, the similarly high central value of the semi-automatic function of writing suggests the central factor is, despite all, quite separate from true intelligence in a higher sense. Further, the great psychological heterogeneity of the correlated abilities seems to point to the central factor in question having to be, at least for the time being, regarded not as purely psychological, but rather as *psycho-physiological*. This gives rise to the speculation that the underlying cause could be a general functional quality of the nervous tissue. The manner in which this general quality functions could perhaps be conceptualized as a “plasticity function.” A nervous system with an increased plasticity function would not be distinguished by its pathways being able to enter into arbitrary new connections more rapidly – as would be required for forming random new associations more quickly (e.g., memorizing meaningless series). Rather, it would be able to develop more precise and durable complexes of abilities across all psychophysiological domains, and accordingly to function in a more accurate and stable manner (in the sense of systematic regularities) – which would manifest as both greater speed and accuracy in common, deeply ingrainedskills. A nervous system favored in its development by an increased plasticity function would distinguish itself from others in terms of performance in a manner similar to that in which a machine made of steel does from a similar one made of iron.

The high central value of pitch discrimination is perhaps the hardest to fit into this framework. However, it should be remembered that the process which allows us to judge one of any two tones as “lower” or “higher” is far more complex than it at first appears.

In any case, we only present the hypothesis outlined above with the greatest reservations; the data are by far not sufficient to allow us to answer this type of fundamental question conclusively. We are concerned mainly with the empirical facts that have been established in this work. If we have additionally allowed ourselves to express our preliminary speculations regarding their interpretation, the goal is not formulating a theory, but rather provoking further thought.

**VI. The main results**

I. A person’s abilities across very diverse domains (pitch discrimination, addition, filling in the gaps in texts, the speed of writing, reading, and counting) show a high and consistent correlation with each other. This correlation does not decrease noticeably if one of the abilities compared is tested by one researcher, the other a week later – without knowledge of the former results – by another (using the same method).

II. The quantitative relationships between these correlations seem to provide sufficient grounds for regarding them as effects of a *common “central factor*.”

III. Whenever one has found the correlations between any three abilities, one is then able to calculate the correlation of each of these abilities with the aforementioned theoretical central factor; we have called these correlations the “central values” of the respective abilities. As far as our data go, the central value for each ability seems, in fact, to remain *constant*, even when the ability is compared with completely different abilities by other researchers on other occasions.

IV. Several common-sense (an in part occasionally suggested) explanations of this central factor have turned out to be untenable. The central factor cannot be reduced to individual differences between subjects with respect to their motivation and momentary disposition, nor their ability to adapt to the experimental conditions or use additional aids, nor even to differing levels of attention.

V. The explanation seems to require a psycho-physiological approach, at least for the time being. The data collected so far possibly suggest that some nervous systems may have a globally increased *plasticity function* as compared to others. This functional proficiency would be a precondition for developing more fine-tuned and consistent complexes of skills, which would then manifest in a greater accuracy and speed of performance across diverse psychophysiological domains. However, this hypothesis is put forward (by us) only with the greatest reservations, mainly for stimulating further psychological and biological research.

VI. In order to obtain reliable correlation values in the first place, it is indispensable to test each case of the two attributes whose correlation is to be determined at least *twice*. Further, the resulting correlation between the two measurement series representing the same objective series of cases of the respective attribute has a general meaning: it serves as a “reliability coefficient” for the measurement method.

\_\_\_\_\_\_\_\_

**Appendix I. Example protocol**

Protocol.

Subject A. Researcher: Krueger.

1. Questions for the subject:

Age: *31; married*.

Music, passive: *moderate ability*.

active: *amateur singer*.

Mathematics, expertise: *Had an excellent aptitude at school.*

Mathematics, particular practice: —.

Health, general: *good, tendency towards anxiety*; current: *in a good disposition*.

Fatigue, physical: —; mental: —

Last period of sleep: *good*.

Distraction, emotional: —; intellectual: —

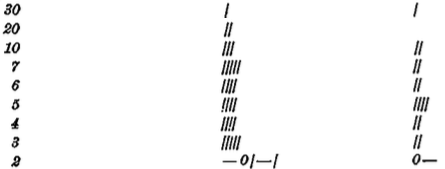
Time since finishing the last meal: *1.5 hours*.

Smoker: ~~heavy, light~~, non-smoker, }

Drinker, ~~abstinent~~: *very moderate*, } very balanced lifestyle

2. Pitch discrimination. Starting time: 3:15.

Difference in vibrations 1st tone is higher 2nd tone is higher



*Threshold = 2.5 vibrations*

Closing time: 3:30.

*Foreigner*

~~3. Combination method (training trial until task is clear, main trial 4 min).~~

4. Tactile thresholds. Starting time: 3:35.

Left cheekbone Right hand Right cheekbone

mm 2 tips 1 tip mm 2 tips 1 tip mm 2 tips 1 tip



*Threshold = 9.*

Closing time: 3:51

5. Addition (training trial, 1 min; 2 main trials, 3 min each).

6. Rote memorization (6, 8, 10, and 12 digits, three series per level).

7. Discussion.

Control question: whether the subject has participated in similar experiments, and when.

N.B.: No practice before the next session; no discussion of the experiments with future participants.

8. Notanda.

Date: *Jan 20 1904*.

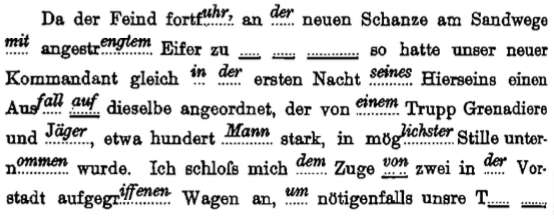
Room temperature: *15.5°C*; outdoor temperature: *-2°C*.

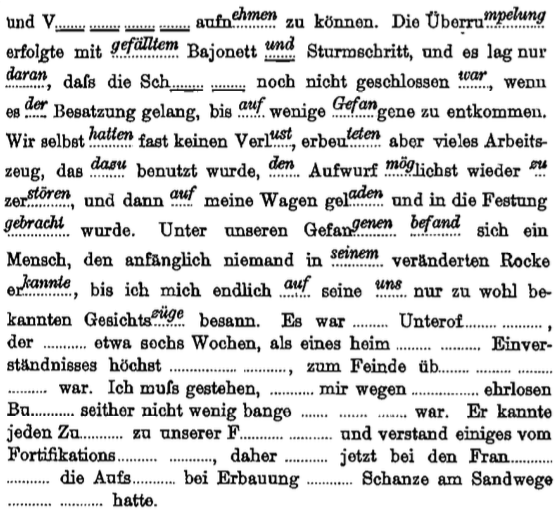
Humidity: *74%*.

Subject’s motivation: *high*.

**Appendix II. Example text for the combination method used by Ebbinghaus**

The siege of Kohlberg. 1807.

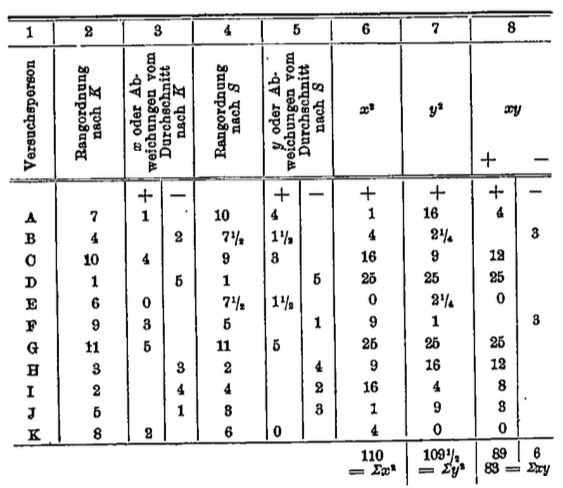




**Appendix III. Example calculation of Bravais’ correlation coefficient and the formula for its probable error**

This calculation can be carried for any two series of values that stand in a pairwise correspondence. In the example below, the two series are rankings; specifically, the rankings produced based on our subjects’ performance on the two addition tests; they are shown in columns 2 and 4. First, the mean of each column is taken, and all individual deviations from it are calculated; these deviations are referred to as *x* and *y* and presented in columns 3 and 5, respectively. We now calculate *x2*, *y2*, *xy*, and their sum (columns 6, 7, and 8, respectively). Then  holds, where *Σ* takes the sum and ***r*** is the sought correlation coefficient. So, in our example, this yields





The probable error of the correlation coefficient is calculated using the following formula:



so, in our example,

