THE CONCEPT OF CORRELATION IN ADULTS

JAN SMEDSLUND

Institute for Social Research, and University of Oslo, Norway

SMEDSLUND, J. The concept of correlation in adults. *Scand. J. Psychol.*, 1963, *4*, 165–173.—In the first of two experiments on the concept of correlation in adult subjects, the subjects' frequency estimates and inferences of relationship were studied relative to five different 2 × 2 distributions, each presented in a fixed sequence. In experiment II, the subjects' spontaneous strategies in subdividing and analyzing one 2 × 2 distribution were studied in a free situation. It is concluded that adult subjects with no statistical training apparently have no adequate concept of correlation (based on the ratio of the two pairs of diagonal frequencies), and that, in so far as they reason statistically at all, they tend to depend exclusively on the frequency of + + cases in judging relationship. The need for studies involving ordinal scale and fully quantified variates is stressed.

Two experiments will be reported, aimed at determining whether normal adults, with no training in statistics, have a concept of correlation. The concept of correlation in its elementary logical form is the ratio of the sum of two diagonal cell frequencies in a fourfold table, and the sum of the other two diagonal cell frequencies, or the total sum. It involves the categorization of events as conforming or not conforming to an hypothesis of equivalence, A = B, i.e. A and B are either both present or both absent. In order to combine the two diagonal frequencies f(A · B) and f(− A · − B), the subject must realize that they both support the hypothesis A = B, although in opposite ways (reciprocals). Likewise, the other two diagonal frequencies f(− A · B) and f(A · − B) are reciprocals, and support the hypothesis AwB (exclusive disjunction), which is the negation of A = B. AwB asserts that either A is present and B is absent, or A is absent and B is present. In the case of negative correlations the frequency of events supporting AwB is higher than the frequency of events supporting A = B.

If the subject, explicitly or implicitly, reasons according to the above mentioned logical structure, and if his concept of frequency entails a ratio scale, he will be said to have an elementary concept of correlation.

The presence vs. absence of a concept of correlation is reflected in the subject's strategies of data gathering, and in his inferences from the data. There are at least three main symptoms of the presence of a concept of correlation:

1. In situations requiring a selective strategy, the subject with a concept of correlation will order the data in four categories, and count or estimate the frequency in each.

2. In situations requiring a receptive strategy, the subject with a concept of correlation will attempt to distribute his attention equally over the four categories in order to estimate their respective frequencies. Two important indices of distribution of attention are time spent observing each event, and accuracy in judging the different event frequencies. The latter measure is employed in this study.

3. The subject with a concept of correlation will base his estimates on the ratio of the sum of two diagonal cell frequencies and the sum of the two other diagonal cell frequencies.

This will be reflected in accurate judgments of presence vs. absence of correlation, in accurate orderings of several materials according to degree of correlation, and, most importantly, in the subject's own explanations of his judgments. Symptom 3 is necessary and sufficient for inferring the presence of a concept of correlation, whereas the two first symptoms are necessary, to the extent that they can be reliably observed, but not sufficient.

EXPERIMENT I

In this experiment the receptive strategies of adult subjects were studied, as well as their inferences of degree of relationship. It was assumed that equal distribution of attention over the four event categories would result in approximately equal accuracy in judging the respective frequencies. Conversely, focusing on one category or ignorance of one category should result, respectively, in increased and decreased accuracy in judging the corresponding frequencies. The receptive strategy of a subject may also be reflected in his explanations and comments. Finally, the possibility of various systematic distortions in the frequency estimates was considered.

Materials

Five packs of one hundred cards each were used, the cards measuring $3\frac{7}{16}$ in. $\times 2\frac{1}{8}$ in. On top of each card was typed 'Patient No.' (from 1 to 100), and the cards were ordered according to these numbers. The upper part of each card contained the word 'Symptoms' and beneath it four different capital letters, taken from the group A, B, C, D, and E (on each card one of these letters was absent). The four letters B, C, D, and E were assigned randomly to the four positions. On the cards where A was present, the one of the other four letters that should have taken its place was suppressed. In all five packs A was present on 50 cards.

The lower part of each card contained the word 'Diagnosis' and beneath it four different capital letters, taken from the group F, G, H, I, and J (on each card one of these letters was absent). The four letters G, H, I, and J were assigned to the four positions in a randomized manner. On the cards where F was present, the one of the other four letters that should have taken its place was suppressed. The frequency distributions in the five packs are shown in Table 1.

Procedure

Sixty-seven student nurses at the University of Colorado Medical Center in Denver, and 19 student nurses from Ullevål Hospital, Oslo, served as subjects. They were seen individually in one 15-minute session. They were assigned to one of five groups differing in the pack of cards presented. Group I was given pack I, group II pack II, and so on. The Norwegian student nurses formed group V. Each subject was asked to read the following instructions:

'We are interested in learning how well beginner students of nursing are able to form an opinion about the practical usefulness of a symptom in diagnosis.

'The pack of cards represents excerpts from the files of 100 patients. The patients are numbered from 1 to 100 in the order in which they were admitted to the hospital.

'The upper part of the cards contains four letters, representing different specific symptoms of the patient. The lower part of the cards contains four letters representing specific diagnoses made by the hospital.

'You are to concentrate entirely on symptom A and diagnosis F. Your task is to look through the pack of cards once, and form an impression of the extent to which A is a useful symptom in the diagnosis of F. In other words, do you think A is a symptom one should pay attention to in trying to determine whether or not the patient is likely to be diagnosed as F?

TABLE 1. *The frequency distributions for the combinations of presence/absence of A and F.*

	Packs									
	I		II		III		IV		V	
	−A	A	−A	A	−A	A	−A	A	−A	A
F	35	35	15	15	15	35	35	15	25	25
−F	15	15	35	35	35	15	15	35	25	25

'Take the cards in your left hand, look at them one at a time and place them face down on the table. You are not allowed to go back to cards you have already seen.

'Read through this once more, and ask questions if anything is unclear.'

After the subject had looked through the pack of cards, the following questions were asked and the answers recorded.

1. Do you think there is a connection or relationship between symptom A and diagnosis F? Tell me about it! (In group V the subjects were, in addition, asked to indicate the strength of the relationship on a seven-point scale from 'perfect' to 'zero'.)
2. How many cards do you think there are with both A and F?
3. How many cards do you think there are without A but with F?
4. How many cards do you think there are with A but without F?
5. How many cards do you think there are without A and without F?
6. Consider yourself an experienced nurse, having reviewed these hundred cases. What would you tell another nurse, with no experience in this area, about the usefulness of symptom A in relation to the diagnosis F? (This question was omitted in group V.)

The following additional task was presented to the subjects in groups III and IV:

'Below are shown the number of the various combinations of presence and absence of A and F in eight other packs of cards.

'Your task is to look at the numbers for each pack and decide whether A and F are related (correlated) in this pack. A relationship may exist, although it may not be perfectly reliable.

'The first row shows the number of cards with the various combinations in Pack I. First, the number of cards with both A and F, then the number with A but not F, then the number with F but not A, and finally the number of cards with neither A nor F.'

	Both A and F	Only A	Only F	Neither of them
Pack 1	50	0	0	50
Pack 2	35	15	35	15
Pack 3	0	50	50	0
Pack 4	25	25	25	25
Pack 5	15	35	35	15
Pack 6	15	35	15	35
Pack 7	50	0	50	0
Pack 8	35	15	15	35

Finally, the subjects in group V were given the following written instructions:

'*a* to *g* below show the number of cards with the different combinations of presence and absence of A and F in seven other packs. Rank the seven materials according to the degree of relationship between A and F. Write the letter for the material with the strongest relationship first, then the letter for the material with the second strongest relationship, and so on. If you think two relationships have the same strength, then write them side by side.'

TABLE 2. *Real values (R), significance of tendency (T) to overestimate (+) or underestimate (−), and accuracy (σ^2 of real minus estimated values), for the cell frequencies in the five groups.*

Significance of over- and underestimations computed by means of the sign test. * and ** indicate significance at the 0.05 and 0.01 levels, respectively.

	Groups														
	I			II			III			IV			V		
	R	T	σ^2	R	T	σ^2	R	T	σ^2	R	T	σ^2	R	T	σ^2
AF	35	+	216.3	15	±	115.4	35	−*	230.1	15	−	158.2	25	+	233.7
A−F	15	+**	187.1	35	−	214.3	15	+**	232.3	35	−	103.0	25	−	354.5
−AF	35	−*	284.6	15	+	103.9	15	+*	478.0	35	−*	199.4	25	+	332.5
−A−F	15	−	100.0	35	−	303.1	35	−	276.1	15	+**	345.1	25	−	166.6

Example of a ranking (with other letters):

j. (strongest)
i.m. (equally strong)
l.
h.o. (equally strong)
p. (weakest)

	Number of cards with:			
	A and F	A but not F	F but not A	Neither A nor F
a.	40	4	16	40
b.	45	5	35	15
c.	70	30	0	0
d.	30	2	33	35
e.	55	15	30	0
f.	50	10	0	40
g.	60	20	5	15

At the end of the session the subjects in group V were also asked about their background with respect to the concept of correlation, and were asked to describe it in terms of one of the following alternatives. (1) Course(s) in statistics, where the concept of correlation was treated. (2) Read book(s) in statistics, where the concept of correlation was treated. (3) Read chapters or passages in other books, where the concept of correlation was treated. (4) Have encountered the concept, but have never seen it defined. (5) Have never heard the word correlation before.

Results

The data on tendencies to over- and underestimate cell frequencies and on the accuracy of estimates in the five groups are presented in Table 2. The accuracy scores show no unambiguous relationship with type of event category. There are several statistically significant tendencies to over- or underestimate frequencies, but they are not clearly linked with type of event category.

The judgments and explanations of the subjects may be roughly classified into the following categories:

TABLE 3. *Number of subjects in groups III and IV who thought a relationship was present for each of eight numerically presented distributions* ($N = 35$).

Distributions				
AF	A – F	– AF	– A – F	Number of subjects
50	0	0	50	34
35	15	35	15	28
0	50	50	0	4
25	25	25	25	21
15	35	35	15	16
15	35	15	35	14
50	0	50	0	31
35	15	15	35	34

Four or five of the subjects gave formulations which rather clearly expressed a *particularistic* concept of relationship, and several others probably also had this type of concept. Examples: 'In most cases there was no connection'; 'When neither symptom nor diagnosis was present, no relationship'; 'No relation in most of the cases.' These subjects appeared to think that a relationship was present in the AF cases, and absent in the others.

Closely related to the particularistic concept, but more statistical, is the notion that the strength of the relationship depends simply on f(AF). More than half of the subjects gave formulations clearly expressing this point of view. Examples: 'Counted after card 27 and found about 11 with both A and F. Think it is a fairly strong relation'; 'To a slight extent; I counted 14 out of 100 (AF's)'; 'There seems to be about ¼ with A and F'; 'No, not too many AF's.' Another concept, frequently hard to distinguish from the preceding, is that amount of relationship is the percentage or proportion of times F appears with A, or A with F. Examples: 'F has turned up about 24 per cent of the time (A was given)'; 'A accompanies F a third of the time, the rest of the time it isn't there'; 'A lot of times A is there but F isn't.' This category did not occur very frequently.

In addition to these categories of reasoning, there were indications of certain intellectual difficulties in coping with the task. Sometimes the subjects referred to the relationship as a causal one. 'I don't think there could be a causal relationship; there might be some other letter that with A could cause F, but not a direct relationship.' There were complaints about the task, like 'It's kind of abstract', and attempts to make it more concrete: 'I'll say symptom A is something that occurs in many illnesses, such as headache. Since it did occur sometimes, I think it's an important factor for diagnosis.'

Not a single subject gave any indication of having understood that the degree of relationship is determined by the ratio of the sum of the AF's and – A – F's and the sum of the – AF's and the A – F's.

Groups III and IV were asked to judge whether or not there was a relationship between A and F in eight numerically presented distributions. The results are given in Table 3.

The numbers in Table 3 have a certain ambiguity, since the individual subjects apparently set different limits to when a relationship was present. Some subjects apparently thought

TABLE 4. *The distribution of estimates of strength of relationship in deck V (N = 19).*

Perfect	0	Weak	3
Very strong	0	Very weak	2
Strong	3	Zero	1
Medium	10		

there was a relationship when f(AF) > 0, others when f(AF) was ⩾0.25, and still others when f(AF) was ⩾0.50. Finally, a few subjects seemed to think there was a relationship when f(AF) > f(A−F). In spite of this ambiguity there is a very high correlation in Table 3 between f(AF) and the number of subjects who think there is a relationship.

The subjects in group V judged the strength of the relationship between A and F in the given pack of cards. The results are presented in Table 4, indicating that the subjects generally had no understanding of the complete lack of relationship between A and F in pack 5. The single subject who thought there was no relationship, based this judgment on an estimated distribution of 50 AF, 20 A − F, 25 − AF, and 5 − A − F, but gave no intelligible explanation.

Finally, the subjects in group V were asked to rank-order seven numerical distributions, constructed in such a way that the rank orders corresponding to a genuine correlation concept, a concept based on f(AF), and a concept based on p(F/A) would be maximally different. The results turned out quite negative. The 19 subjects ranked the distributions in 19 different ways with no apparent clustering around common types. Only one subject had a rank order corresponding to a concept of relationship based on f(AF), and no subject had a rank order corresponding to a genuine correlation concept or to a concept based on p(F/A). Judging from the subjects' general behavior and comments, it appeared that the ranking procedure, in spite of repeated explanations and simplifications, exceeded their intellectual capacity. As to their previous knowledge of the concept of correlation, the results showed that ten subjects had never heard of the concept, six of them had encountered the word, but never had it explained, and only three had read chapters or passages in books containing a treatment of the concept. However, these three subjects did not perform better than the others.

EXPERIMENT II

In experiment I the subjects were given the essentially passive role of observing the cards once in a fixed order, and the arrangement of letters on the cards made it fairly exacting to register the presence and absence of A and F. These conditions put a heavy strain on the information processing and information storing capacity of the subjects, and may have prevented a maximal unfolding of their actual reasoning power.

Experiment II was intended to create optimal conditions for the occurrence of an understanding of the concept of correlation. The subjects were left free to arrange and sort the cards in any way they wanted to, they were allowed to look at them an unlimited number of times, and they were given pencil and paper to make notes. Also, the cards were simplified to contain only +A or −A and +F or −F.

Procedure

A pack of one hundred cards was used. On the upper part of each card was written in black ink either +A or −A, on the lower part either +F or −F. There were 37 cards with +A+F, 33 with −A+F, 17 with +A−F, and 13 with −A−F.

Twenty-eight nurses at Ullevål Hospital, Oslo, participated in one 10-minute session each. The following written instructions were presented:

'We are interested in learning how well students of nursing are able to form an opinion about the practical usefulness of a symptom in diagnosis.

'The pack of cards represents excerpts from the files of 100 patients. The upper part of the cards contains either a +A or a −A. +A means that the patient had symptom A, and −A means that the patient did not have symptom A. The lower part of the cards contains either a +F or a −F. +F means that the illness F was found, and −F means that the illness F was not found.

'Your task is to find out whether there is a relationship (connection) between the symptom A and the illness F. You may study and order the material in any way you wish, and you may make notes. Try to work as fast as possible. Read the instructions once more, and ask questions if anything is not clear.'

The experimenter recorded the subject's spontaneous behavior and his conclusions. If the subject did not spontaneously explain his conclusions, the experimenter asked questions of the type 'How do you know that?', 'How did you arrive at that?'

Results

The data may be divided into spontaneous procedure, conclusion, and justification. Despite repetitions, attempts at clarification, and concrete exemplification, the subjects generally found it hard to understand the task, and more than one third of them complained that they did not understand. Seventeen of the 28 subjects went through the material placing the cards together in one pile. Some of them asked whether one was supposed to sort the cards, but being told they were free to do anything they wished, they nevertheless placed the cards in one pile. Three subjects did not even look through the entire material, and one subject sorted the +A+F's in one pile, and the rest in another. Five subjects sorted the cards in four categories, one subject sorted them in five (two piles of −A−F's), and one sorted them in seven apparently unsystematic piles. No subject used the paper and pencil offered.

Twenty-four subjects thought there was a relationship between A and F, two subjects thought there was no relationship (correct), and two subjects gave up and did not answer this question.

The explanations may be roughly categorized according to whether they referred to frequencies or not. In the former group (18 subjects) fourteen said that there was a relationship because the number of +A+F's was the largest or was large. One subject said there was a relationship because there were more +F's than −A's, and another subject said there was no relationship because the −A+F's were the most frequent cards. Two subjects referred both to the +A+F and the −A−F categories. This may indicate a dawning understanding of correlation, but not very developed, since these two subjects sorted in respectively one and seven piles, and both were confident that there was a substantial relationship between A and F.

The ten subjects who did not refer to frequencies, either gave up without answering (three), said they felt it 'intuitively' (two), or gave vague explanations like 'Every symptom

must indicate an illness'; 'If the symptom is present, the illness will come'; 'If the illness was found, there had to be symptoms present.' There were also a few particularistic statements like 'There is no relationship, because when one finds "+" on one and "—" on the other, there can be no relationship.'

DISCUSSION

The data on the subjects' inferences, and on their selective strategies, all point to an absence of correlational reasoning. The complex data on frequency estimates offer no clear evidence concerning the subjects' receptive strategies. The lack of a clear relationship between the accuracy of frequency estimates and event category either means that accuracy is not, in this context, a good measure of attention distribution, or that no simple theory of attention distribution is adequate. However, the existence of pronounced variations in accuracy and in tendency to over- or underestimate means that there may exist important regularities to be discovered by more extensive and refined experimentation.

The apparent main finding of these experiments, then, is that normal adults with no training in statistics do not have a cognitive structure isomorphic with the concept of correlation. Their strategies and inferences typically reveal a particularistic, non-statistical approach, or an exclusive dependence on the frequency of + + instances.

This conclusion is supported by Wason's (1960) finding that very few psychology undergraduates spontaneously searched for negative evidence in a simple conceptual task. If an hypothesis repeatedly gave positive results, they tended to accept it without investigating whether alternative hypotheses would lead to negative results. This corresponds directly to the lack of understanding of the relevance of the − A − F cases in the present experiments. However, it should be noted that the absence of adequate reasoning may occur primarily when the situation is fairly complex. By drastically lowering the frequencies involved, by making the cases directly and simultaneously observable, and by selecting one's subjects with respect to training and intellectual capacity, it may be possible to elicit somewhat more frequent symptoms of an understanding of correlation. This is exemplified by a study of Inhelder & Piaget (1958). They report fairly frequent occurrences of the correlation concept in 14- to 15-year-old subjects, in very simple situations with small numbers of simultaneously visible faces with brown vs. fair hair, and brown vs. blue eyes.

The findings may be interpreted relative to Piaget & Inhelder's (1951) theory of the development of the concepts of chance and probability in children. According to these authors, children before 7–8 years have no concepts of chance and probability, since they have not yet differentiated the ideas of necessity and possibility. In the absence of logical structure, there are no compelling deductions of what has to be or cannot be the case, and consequently no discovery of the non-necessary, i.e. merely possible. In the widest connotation of the word chance, it is the realm of things whose occurrence is neither implied by, nor excluded by one's assumptions. From around 7–8 years the development of concrete reasoning introduces necessity, and thereby the complementary notions of possibility and chance. However, reasoning concerning the latter requires the manipulation of possibilities in a combinatorial system (formal reasoning), and only develops after 11–12 years. Probability estimates presuppose an ability to determine the structure of possibilities entailed by the given situation, and this lies outside the scope of concrete reasoning, which functions only

on the basis of actually observed events. For example, eight-year-old children may correctly estimate that a frequency of 3 out of 4 entails a higher probability than 2 out of 4, but may then proceed to argue that 3 out of 6 is more probable than 2 out of 3, thus revealing that their estimates are based on absolute rather than relative frequencies. The concept of correlation represents a rather late and complex development of formal reasoning.

The present data lead to the following interpretation of the Piaget-Inhelder theory: The analyses in this theory are seen as referring to the *optimal performance* at each developmental stage. Thus, a child capable of concrete reasoning in some situations may be functioning at a prelogical level in others, but will never reason at a formal level. An adult who is able to apply correlational reasoning in a highly simplified situation may regress to a particularistic or to an absolute frequency approach in many everyday life situations. Briefly, it is assumed that the developmental stages are descriptive of different levels of cognitive functioning, and that adults may, at various times, function at any of these levels, although under optimal conditions they are capable of formal, hypothetico-deductive reasoning. The empirical problem in each case is to determine under what range of conditions a given person will function at a given level. The present data support the hypothesis that this range is very restricted, or even non-existent, in most adults as far as the concept of correlation is concerned.

There are two important objections to this conclusion. One of them concerns the fact that nurses are thoroughly trained to adopt a strictly deterministic approach to symptoms and illnesses. Their comments frequently appear to reflect a reluctance to regard chance as an element in this connection. Some of them explicitly argue that one A − F and/or one −AF is enough to disprove the hypothesis of a one-to-one causal connection, and one or two even speculate about possible multiple causes. It is theoretically possible that this refusal to reason statistically reflects a preference rather than an incapacity. However, this argument is weakened by the fact that none of those who *did* accept a statistical approach seemed to have a concept of correlation.

Another objection to the conclusion that most adults completely lack a concept of correlation, stems from the restriction of the reported experiments to dichotomous variates. Although dichotomous presence/absence variation is elementary from a logical point of view, it may not be so psychologically. Indeed, the history of correlational analysis, as well as the customary textbook introductions of correlation by means of a scattergram, suggest that ordinal scale or fully quantified variates may yield the easiest access to an understanding of the concept. No data are available to settle this important question.

The first part of this study (experiment I, groups I–IV) was conducted during the author's stay as a Research Associate at the Behavior Research Laboratory, University of Colorado. It was supported by a grant from The National Institute of Mental Health (Project Code Number GN-9028 Clinical Inference and Cognitive Processes) and from the Council on Research and Creative Activity of the University of Colorado. The second part of the study was supported by the Norwegian Research Council for Science and the Humanities.

The author is indebted to Robert F. Schneider, Carolyn J. Hursch, N. John Castellan, Jr., and Hans Christian Rasmussen, for their assistance, and to Gudrun Eckblad for valuable comments.

REFERENCES

INHELDER, B. & PIAGET, J. (1958). *The growth of logical thinking from childhood to adolescence.* New York: Basic Books.

PIAGET, J. & INHELDER, B. (1951). *La genèse de l'idée de hasard chez l'enfant.* Paris: Presses Univ. France.

WASON, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quart. J. exp. Psychol.*, *12*, 129–140.