Biometrika (1961), **48**, 3 *and* 4, *p.* 303
Printed in Great Britain

Inconsistencies in a schedule of paired comparisons

BY PATRICK SLATER
Institute of Psychiatry, The Maudsley Hospital, University of London

1. THE METHOD, AND THE HYPOTHESES CONCERNING IT

The method of paired comparison has had a long and honourable history in psychological experiments, beginning with the researches of Witmer and Cohn, published in 1894. Titchener (1901) described it in detail in one of the earliest text-books on experimental psychology and Guilford (1954) devotes a chapter to it in the latest edition of his popular text-book. Theoretical investigations of the method, which has applications outside psychology, still continue to appear, e.g. by David (1959) in this Journal, in technical reports by Gulliksen & Tucker (1959) and in a thesis by the author (1960). The authoritative paper on the null hypothesis concerning it is the one by Kendall & Babington Smith which appeared here in 1939. With this I find myself in disagreement.

The experimental procedure is to show a set of m objects to an individual in pairs and ask him each time to choose one. It is always understood that the objects differ from one another, but there may be doubt whether the difference is discernible by the individual. The difference may be confined to one respect, e.g. a set of boxes may be used identical in appearance but differing in weight, and the observer's attention may be directed to that respect, e.g. by the instruction, 'Choose the heavier each time'. Or they may differ in several respects and the criterion of choice may be left to the individual, e.g. in Titchener's standard procedure the objects are coloured cards differing in hue and saturation and the individual is instructed to choose whichever he prefers. It is normally understood, but not always, cf. Myers (1925), that each of the $\frac{1}{2}m(m-1)$ possible pairs is presented once and once only.

We shall assume that the objects may differ in several respects and that the individual's attention has not been directed to any respect for which there is an independent criterion; also that he has been shown every possible pair once and is never permitted to evade the obligation to choose, e.g. by responding, 'Both alike'. The objects will be denoted $A, B, \ldots, M$.

Initially we may hope to show that the individual is aware of *one* dimension of preference, in accordance with which the objects can be arranged in an order from most preferred to least. The contrary, C_1, which must be disproved before any such hypothesis, H_1, need be conceded, is that the individual is unaware of any differences between the objects and that all his choices are made at random, independently of one another. It may be disproved if an unexpectedly large number of the choices are internally consistent, i.e. cohere with the same one out of all the $m!$ possible orders for m objects; for in the absence of any criterion all possible orders are equally admissible. The minimum number of inconsistent responses will be denoted by i, and an order with which there are only i inconsistent responses will be called a nearest adjoining order.

In some specimen schedules of responses the nearest adjoining order is not unique; there may be several orders, say j altogether, with only i inconsistencies. The numbers

i and j are always ascertainable, at least in theory, if the schedule is checked against the $m!$ possible orders. Consider, for instance, $A > B$, $A < C$, $B > C$, a specimen schedule for 3 objects (read $>$ as 'preferred to'). One of the responses is inconsistent with each of the three orders $A > B > C, B > C > A$ and $C > A > B$. The other possible orders, $A > C > B$, $B > A > C$ and $C > B > A$, can be omitted from consideration for two responses are inconsistent with each of them. The nearest adjoining order is not unique so the inconsistent response cannot be identified, but certainly $i = 1$ and $j = 3$.

The sample space or universe for $m = 3$ contains 8 distinct specimens of possible schedules of responses; and all are equiprobable on C_1. It is easy to verify that six have $i = 0$ and two have $i = 1$. So a schedule with $i = 0$ is not exceptional when $m = 3$ and C_1 is tenable for all the specimens in this universe. In general the universe for m objects contains $2^{\binom{m}{2}}$ equiprobable specimens. Let s_c be any specimen with a certain number of inconsistencies, i.e. with $i = c$, and let $f_m(c)$ be the frequency of occurrence of all such specimens in the universe for m. The question to be decided is whether C_1 is tenable concerning a particular s_c from this universe. If we make it our rule to reject C_1 when its probability is below 0·05 we can reach a decision if we know what is the limiting values of i, say $i = u$, for which

$$\sum_{i=0}^{u} f_m(i)/2^{\binom{m}{2}} < 0{\cdot}05.$$

Then we reject C_1 if $c \leqslant u$ but not otherwise.

Consider next the s_c with $c \leqslant u$ in different universes, for all of which H_1 must be admitted. As m and consequently u are allowed to increase, c can increase indefinitely without exceeding u. At some point it may begin to seem surprising that so many responses, which can be itemized if the nearest adjoining order is unique, are all consistent with one another, i.e. with one other ordering of the objects, which may be called the residual order; and we may feel tempted to consider the more elaborate hypothesis, H_2, that the individual is aware of *two* dimensions of preference. For the objects can be arrayed on a surface definable by two axes, falling in the nearest adjoining order along one and in the residual order along the other, so that every choice appears consistent with one or other of the two orders. The contrary we now encounter, C_2, which must be disproved before H_2 need be conceded, is that the choices inconsistent with the nearest adjoining order do not imply any awareness of a second dimension but are all made at random and independently of one another. The argument of § 3 below is that under the conditions of the experiment C_2 is always tenable, no matter how large m and i are.

The hypotheses under consideration all relate to the $\binom{m}{2}$ responses the individual is required to make. Each is the result of a single act of choice, potentially independent of every other such act and liable to bring the laws of probability into operation. So I regard the responses as the simple events from which the universe for m originates and conclude that the probability distribution for i, the number of inconsistent responses, is what needs to be examined when C_1 is under consideration—not the probability distribution for d, the number of circular triads in a schedule, which is the variable considered by Kendall & Babington Smith.

A triad is the set of responses relating to three objects. It may have $i = 0$ or 1, as already mentioned, and it is circular when $i = 1$. The authors only give reasons of simplicity and convenience for treating triads as units for enumeration. After describing them and larger

polyads within the complete configuration or m-ad which may be used to represent the schedule of responses, they remark 'it seems best to confine attention to circular triads, which, so to speak, constitute the inconsistent elements in the configuration, and to ignore the more ambiguous criteria associated with circular polyads of greater extent'. They do not mention the possibility of treating each response as a unit, nor do they offer any explicit definition of the null hypothesis to be considered.

Triads ought not to be treated as elements. They are compound events not conceivably independent of one another, for the total number of triads in a schedule exceeds the number of responses by a factor of $\frac{1}{3}(m-2)$ and each response features in $m-2$ triads. Moreover, there is no 1 : 1 relationship between i and d; schedules from the same m with the same i may differ in d, and vice versa. For instance,

$$\text{when } i = 1, \quad d \text{ ranges from 1 to } m-2,$$

$$\text{when } i = 2, \quad d \text{ ranges from 2 to } 2m-6;$$

and further evidence appears in Fig. 1 and Table 2. If the inconsistencies were subclassified d might be defined as a weighted summation of the frequencies in specified classes: that is to say, d may be viewed as a summation in which some inconsistencies receive more weight than others. But on the assumption that the inconsistent responses result from erratic acts of choice and occur at random there is no justification for subclassifying them. And the conditions of the experiment do not include any region where this assumption can be proved to have a negligible probability.

Inconsistent responses receive equal weights in Kendall's procedure for τ (1938, 1948), so there is a simple relationship between τ and i. A nearest adjoining order might be defined as any order which maximizes τ for the schedule under consideration, and the maximum value of τ is obtainable from i, given m, as $1-4i/m(m-1)$.

2. The first form of the null hypothesis, C_1

On C_1 when two of the objects, I and J, are presented to the individual, since he is unaware of any difference between them but obliged to make a choice, he is just as likely to choose I or J, and his choice will be not influenced by any choices he may have made on any previous occasion when he may have had I or J presented for comparison with any of the other objects.

The universe of different schedules of responses thus obtainable consists of $2^{\binom{m}{2}}$ specimens, all equiprobable. This total needs to be broken down into subtotals for specimens where $i = 0, 1, 2, \ldots$. Then to decide whether C_1 applies to a particular schedule for m objects we need to find the number of inconsistencies in it and see what proportion of the schedules in the universe contains no more than the same number of inconsistencies.

Table 1 shows the breakdowns for $m \leqslant 8$, and the cumulative proportions derived. If C_1 is considered acceptable at the 0·05 probability level but not below, it is tenable for all schedules where $m < 6$, but not when $i = 0, m \geqslant 6$, when $i = 1, m \geqslant 7$, or when $i = 2, m \geqslant 8$.

The frequency distribution for any m may be defined as the expansion

$$2^{\binom{m}{2}} = f_m(0) + f_m(1) + f_m(2) + \ldots.$$

A general algebraic definition of $f_m(i)$ would thus define the complete frequency

distribution for all m. Such an expression has not yet been found, but the expressions for values of i up to 3 can be given. They are

$$f_m(0) = m!,$$

$$f_m(1) = m!\,(3m^2 - 13m + 14)/6,$$

$$f_m(2) = m!\,(9m^4 - 78m^3 + 235m^2 - 438m + 680)/72,$$

$$f_m(3) = m!\,(135m^6 - 1{,}755m^5 + 8{,}685m^4 - 27{,}185m^3 + 77{,}820m^2 - 157{,}204m + 210{,}336)/6480.$$

The expression for $f_m(2)$ only applies when $m \geqslant 4$, and the expression for $f_m(3)$ only when $m \geqslant 6$. When $m = 5$, $f_5(3) = 24$.

These expressions give $\sum_{i=0}^{3} f_m(i)/2^{\binom{m}{2}} = 0{\cdot}009902$

for $m = 9$. So u is certainly not less than 3 when $m \geqslant 9$. It may even exceed 3, and appears to be increasing at an accelerating rate.

Table 1. *The frequency distribution of i for given values of m*

i	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$	$m=7$	$m=8$
				Part 1			
0	2	6	24	120	720	5,040	40,320
1	—	2	40	480	5,280	58,800	685,440
2	—	—	—	400	13,280	278,880	5,120,640
3	—	—	—	24	11,568	651,504	21,590,016
4	—	—	—	—	1,920	736,848	55,101,312
5	—	—	—	—	—	323,120	84,325,248
6	—	—	—	—	—	41,040	71,687,040
7	—	—	—	—	—	1,920	27,421,440
8	—	—	—	—	—	—	2,464,000
				Cumulative proportions			
				Part 2			
0	1·0	0·750	0·375	0·11719	0·02197	0·002403	0·000150
1	—	1·0	1·0	0·58594	0·18311	0·030441	0·002704
2	—	—	—	0·97656	0·58838	0·163422	0·021780
3	—	—	—	1·0	0·94141	0·474083	0·102209
4	—	—	—	—	1·0	0·825439	0·307477
5	—	—	—	—	—	0·979515	0·621613
6	—	—	—	—	—	0·999084	0·888668
7	—	—	—	—	—	1·0	0·990821
8	—	—	—	—	—	—	1·0

The data for $m \geqslant 6$ have been provided by the National Physical Laboratory using an electronic computing programme developed by G. G. Alway as a research project. An account of it will be published separately. Considerable expense would be incurred if the research were continued to obtain complete expansions of $2^{\binom{m}{2}}$ for larger m or expressions defining $f_m(i)$ for larger i, so that it seems desirable to publish the present results and to ascertain the consensus of expert opinion before proceeding. For most practical purposes it would be sufficient to know the values of u for $m \leqslant 15$. It is true that experiments have

been conducted with considerably more than 15 objects. Titchener (1901) regularly used one with $m = 27$ for his course in experimental psychology and Cattell, Maxwell, Light & Unger (1949) have described one with $m = 50$. Experiments with large m need not be difficult to conduct if the objects are suitably chosen and appropriate apparatus is constructed for presenting them in pairs and recording the responses automatically, but it is not often that any compelling reasons for conducting such experiments are encountered in practice.

3. The second form, C_2; and an argument that every inconsistency should be given an equal weight

The evidence of a single schedule of responses is never sufficient to make C_2 untenable. Take $A > B > \ldots > M$ arbitrarily as an order with which some of the individual's responses cohere. Then the remainder must all cohere with the opposite order $A < B < \ldots < M$. If none are consistent with the first all are consistent with the second. So the i responses inconsistent with the nearest adjoining order must *a fortiori* all be consistent with one another, and in general may be linked together in many different ways to form possible residual orders. In other words evidence in favour of H_2 is indistinguishable from evidence against it, so C_2 cannot be ruled out. If we wish to disprove C_2 we must adduce supplementary evidence from other sources, modify the conditions of the experiment in some way or advance some specific argument.

Thus the only alternatives to be considered when investigating the internal consistency of a single schedule of responses are

(i) C_1: the observer is unable to discriminate between the objects, or

(ii) H_1+C_2: he is aware of a single dimension of preference. Choices not made in accordance with it are produced by chance causes, i.e. causes operating independently on particular judgements, such as distractions or momentary lapses of attention, etc.

There is never any case for pressing on to consider alternatives such as might be denoted $H_1+H_2+C_3$, etc., without additional evidence.

We may argue from this that every inconsistency should be given an equal weight when C_1 is under consideration. If the order $A > B > \ldots > M$ is the dimension of preference characteristic of the individual, a cause operating accidentally is just as likely to produce the reversal $A < M$ as $A < D$, say, and as no causes other than accidental causes need be supposed, we ought not to assign more weight to one such reversal than another. A straight count, that is to say, an unweighted summation of the inconsistencies is therefore the index we should use in deciding whether H_1 or C_1 is to be preferred.

The proposition can be sustained, perhaps quite adequately, without reference to C_2. For *per contra* we cannot claim that $A < M$ should be given a greater weight than $A < D$ without postulating that A is further removed from M than from D on the scale of preference characteristic of the individual. But this is to concede a form of H_1, and we should not make any such concession before we have succeeded in disproving C_1. Moreover, the relative weights we assign to $A < D$ and $A < M$ must depend on the particular form of H_1 we choose to concede; but even after disposing of C_1 we may be left with $j > 1$, i.e. with several equally acceptable forms of H_1.

4. COUNTING i IN A SCHEDULE OF PAIRED COMPARISONS OBTAINED EXPERIMENTALLY

Mr Alway has kindly contributed the following practical notes:

No simple rule for obtaining i is known to be applicable in all cases, but in all practical applications encountered so far the following simple rules have sufficed.

First re-order the rows and columns of the preference matrix* so that the numbers of $+$'s in successive rows are in descending order. Next, examine each row to the right of the diagonal element, and proceeding element by element count separately the positive and negative ones. If at any stage the number of negative elements exceeds the number of positive ones the matrix may be transformed to decrease to total number of negative elements in the upper triangular half. For example, if one of the rows (starting with the diagonal element) is

$$. + + - + - - + - - + + +$$

then by placing this row and the corresponding column nine places further on, the total number of negative elements in the upper half is reduced by 1. The resultant matrix can be examined again in this way. The columns should also be examined in a corresponding fashion; this is the same as examining the rows starting with the diagonal element and counting backwards towards the first element. The process should then be repeated, and wherever the count of $-$'s equals the count of $+$'s the matrix should be transformed in a corresponding fashion. This change will not of itself reduce the total number of $-$'s, but it may alter the position of certain elements so that the number of $-$'s in some row or column exceeds the count of $+$'s, and then the total number may be reduced. This process has sufficed in all practical applications ($m \leqslant 10$) to reduce the number of inconsistencies to its lowest value, and also to give all the permutations for which this lowest value is attained.

Even when the first rule, to put the number of $+$'s in successive rows in descending order, is omitted, the simplest case in which the second rule, of counting $+$'s and $-$'s by row and column, fails by itself to give i is

$$\begin{array}{cccccccc}
. & + & + & - & + & - & + & + \\
- & . & + & + & + & + & + & + \\
- & - & . & + & + & - & + & - \\
+ & - & - & . & + & + & + & + \\
- & - & - & - & . & + & + & - \\
+ & - & + & - & - & . & + & + \\
- & - & - & - & - & - & . & + \\
- & - & + & - & + & - & - & .
\end{array}$$

for which $i = 4$, given by the permutation (24618357) of the rows and columns.

The calculation of the data mentioned in § 2 and the writing of this section have been carried out as part of the research programme of the National Physical Laboratory and they are published by permission of the Director of the Laboratory.

* A preference matrix for recording the responses in a schedule is an $m \times m$ table with a row and corresponding column for each object. The response $I > J$ is recorded as $+1$, or simply $+$, in row I column J, and as -1, or simply $-$, in row J column I.

5. Kendall's d and its relation to i

When one of three conjoined choices is inconsistent with the other two the triad formed is circular. Kendall & Babington Smith's procedure depends on finding the number, d, of such triads in a schedule. The simple computing method for finding d is a great advantage of their procedure. Counting the number of + 's in each row of the preference matrix **A** provides a column vector **a** which is a partition of $\frac{1}{2}m(m-1)$; and d can be obtained from

$$2d = m(m-1)(2m-1)/6 - \mathbf{a}'\mathbf{a}.$$

$\mathbf{A}_1$					$\mathbf{a}_1$	$\mathbf{A}_2$					$\mathbf{a}_2$	$\mathbf{A}_3$					$\mathbf{a}_3$
.	+	−	+	+	3	.	+	+	+	−	3	.	+	−	+	+	3
−	.	+	+	+	3	−	.	+	+	+	3	−	.	+	+	+	3
+	−	.	+	+	3	−	−	.	+	+	2	+	−	.	+	−	2
−	−	−	.	+	1	−	−	−	.	+	1	−	−	−	.	+	1
−	−	−	−	.	0	+	−	−	−	.	1	−	−	+	−	.	1

Fig. 1. Three preference matrices, and their partitions.

For example, the three matrices in Fig. 1 with partitions as shown have

	$\mathbf{a}'\mathbf{a}$	d	i	j
$\mathbf{A}_1$	28	1	1	3
$\mathbf{A}_2$	24	3	1	1
$\mathbf{A}_3$	24	3	2	5

It appears debatable whether the inconsistency in $\mathbf{A}_1$ should be given more or less weight than the one in $\mathbf{A}_2$. For instance, it might be argued from j that the one in $\mathbf{A}_2$ is the more reasonably attributable to some accidental cause, as it does not evoke any doubt about what order represents the individual's characteristic dimension of preference. My view, based on the argument in § 3, is that both inconsistencies should receive the same weight. Kendall's procedure weights the inconsistency in $\mathbf{A}_2$ three times as heavily as the one in $\mathbf{A}_1$.

Moreover, Fig. 1 shows that no simple relationship exists between i and d: $\mathbf{A}_1$ and $\mathbf{A}_2$ have the same i but a different d, $\mathbf{A}_2$ and $\mathbf{A}_3$ have the same d but a different i. Table 2 shows the relationship between i and d in the universes for $m \leqslant 8$. The two quantities are quite closely correlated, viz.

In the universe for	r is
$m = 4$	0·9317
5	·9087
6	·9031
7	·8969
8	·8927

It is not surprising that r diminishes as m increases. Increasing m provides more freedom for preference matrices with the same partition to vary in the internal arrangements for their + 's and − 's.

The correlation is not close enough to prevent different results being obtained when d and i are used to test C_1 with reference to a single schedule of responses. When $m = 7$ d leads to the rejection of some A's at the 5 % significance level where $i = 1$ and the acceptance of others where $i = 2$; and when $m = 8$ to the rejection of some where $i = 2$ and the acceptance of others where $i = 3$ or even 4. The 5·85 million possible schedules acceptable

Table 2. *The association between i and d shown by the frequencies of schedules with given i and d in the universes for m = 4, 5, 6, 7 and 8*

Case m = 4

d \ i	0	1	Total
2	—	24	24
1	—	16	16
0	24	—	24
Total	24	40	64

Case m = 5

d \ i	0	1	2	3	Total
5	—	—	—	24	24
4	—	—	280	—	280
3	—	120	120	—	240
2	—	240	—	—	240
1	—	120	—	—	120
0	120	—	—	—	120
Total	120	480	400	24	1,024

Case m = 6

d \ i	0	1	2	3	4	Total
8	—	—	—	720	1,920	2,640
7	—	—	—	4,800	—	4,800
6	—	—	2,880	5,760	—	8,640
5	—	—	3,360	288	—	3,648
4	—	720	5,520	—	—	6,240
3	—	1,440	1,440	—	—	2,880
2	—	2,160	80	—	—	2,240
1	—	960	—	—	—	960
0	720	—	—	—	—	720
Total	720	5,280	13,280	11,568	1,920	32,768

Case m = 7

d \ i	0	1	2	3	4	5	6	7	Total
14	—	—	—	—	—	—	720	1,920	2,640
13	—	—	—	—	—	31,920	40,320	—	72,240
12	—	—	—	—	52,080	181,440	—	—	233,520
11	—	—	—	5,040	139,440	99,680	—	—	244,160
10	—	—	—	50,400	323,568	10,080	—	—	384,048
9	—	—	—	131,040	149,520	—	—	—	280,560
8	—	—	20,160	184,800	72,240	—	—	—	277,200
7	—	—	30,240	157,920	—	—	—	—	188,160
6	—	—	70,560	109,200	—	—	—	—	179,760
5	—	5,040	72,240	13,104	—	—	—	—	90,384
4	—	10,080	65,520	—	—	—	—	—	75,600
3	—	15,120	18,480	—	—	—	—	—	33,600
2	—	20,160	1,680	—	—	—	—	—	21,840
1	—	8,400	—	—	—	—	—	—	8,400
0	5,040	—	—	—	—	—	—	—	5,040
Total	5,040	58,800	278,880	651,504	736,848	323,120	41,040	1,920	2,097,152

Table 2 (*cont.*)

Case $m = 8$

d \ i	0	1	2	3	4	5	6	7	8	Total
20	—	—	—	—	—	—	40,320	1,612,800	1,576,960	3,230,080
19	—	—	—	—	—	—	1,236,480	8,279,040	887,040	10,402,560
18	—	—	—	—	—	483,840	11,544,960	12,284,160	—	24,312,960
17	—	—	—	—	—	3,010,560	17,310,720	4,166,400	—	24,487,680
16	—	—	—	—	483,840	11,329,920	24,326,400	1,048,320	—	37,188,480
15	—	—	—	—	1,317,120	16,251,648	11,719,680	—	—	29,288,448
14	—	—	—	40,320	5,362,560	24,465,280	4,863,360	30,720	—	34,762,240
13	—	—	—	403,200	8,682,240	14,784,000	645,120	—	—	24,514,560
12	—	—	—	1,290,240	12,566,400	10,631,040	—	—	—	24,487,680
11	—	—	—	1,908,480	9,639,168	3,207,680	—	—	—	14,755,328
10	—	—	161,280	4,179,840	11,319,168	161,280	—	—	—	15,821,568
9	—	—	241,920	4,058,880	4,032,000	—	—	—	—	8,322,800
8	—	—	645,120	4,587,520	1,693,440	—	—	—	—	6,926,080
7	—	—	779,520	3,091,200	—	—	—	—	—	3,870,720
6	—	40,320	1,249,920	1,747,200	5,376	—	—	—	—	3,042,816
5	—	80,640	967,680	256,256	—	—	—	—	—	1,304,576
4	—	120,960	806,400	26,880	—	—	—	—	—	954,240
3	—	161,280	241,920	—	—	—	—	—	—	403,200
2	—	201,600	26,880	—	—	—	—	—	—	228,480
1	—	80,640	—	—	—	—	—	—	—	80,640
0	40,320	—	—	—	—	—	—	—	—	40,320
Total	40,320	685,440	5,120,640	21,590,016	55,101,312	84,325,248	71,697,040	27,421,440	2,464,000	268,435,456

in accordance with i and the 9·92 million acceptable in accordance with d when $m = 8$ include 4·80 million in common. There is disagreement about the remainder. Part of this disagreement arises because i and d are discrete variables, so that the tails of their probability distributions cannot be cut off exactly at the 0·05 level. In percentages, 2·18 pass the test on i, 3·70 pass on d, 1·79 pass on both.

When several individuals, say n altogether, are asked to compare the same m objects in pairs and little evidence of agreement is found between them, the question may arise whether the absence of agreement reflects differences in taste or lack of discernment. It should be possible to extend the use of i or d to consider problems of this kind, and the correlation between them should be sufficient to lead to convergent conclusions when n is not too small. For n above a certain limit the advantage of easy computation might tell decisively in favour of d.

I would like to emphasize the importance of distinguishing between problems of discernment and problems of agreement in this context. Comparison in pairs is specially appropriate for problems of discernment; it provides more evidence of internal consistency, or the lack of it, than comparison in sets of more than two at a time. But for investigating problems of agreement it does not appear to have any advantages over other methods of multiple comparison, of which ranking is administratively the most convenient.

REFERENCES

CATTELL, R. B., MAXWELL, E. F., LIGHT, B. H. & UNGER, M. P. (1949). The objective measurement of attitudes. *Brit. J. Psychol.* **40**, 81–90.

COHN, JONAS (1894). Experimentelle Untersuchungen über die Gefühlsbetonungen der Farben, Helligkeiten und ihrer Combinationen. *Philos. Stud. Leipz.* **10**, 562–603.

DAVID, H. A. (1959). Tournaments and paired comparisons. *Biometrika*, **46**, 139–49.

GUILFORD, J. P. (1954). *Psychometric Methods.* New York: McGraw Hill Book Co. Inc.

GULLIKSEN, H. & TUCKER, L. R. (1959). A general procedure for obtaining paired comparisons from multiple rank orders. Princeton University, NR 150–088.

KENDALL, M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**, 81–93.

KENDALL, M. G. (1948). *Rank Correlation Methods.* London: Charles Griffin and Co. Ltd.

KENDALL, M. G. & BABINGTON SMITH, B. (1939). On the method of paired comparisons. *Biometrika*, **31**, 324–45.

MYERS, C. S. (1925). *A Text Book of Experimental Psychology.* Cambridge University Press.

SLATER, P. (1960). The reliability of some methods of multiple comparison in psychological experiments. London University thesis.

TITCHENER, E. B. (1901). *Experimental Psychology. I. Qualitative.* London: Macmillan and Co. Ltd.

WITMER, L. (1894). Zur experimentellen Aesthetick einfacher räumliche Formverhältnisse. *Philos. Stud. Leipz.* **9**, 96–144 and 209–63.