

epa
Cart
call
Pro
the
ult
The
sid
ity
lig
th
need
ned
iss
can
dot
on
the
is
th
to
in
of
be
wh
1.

EDITORIAL COMMENT

On learning that an extensive report of his telepathic experiments was to be published by W. Whately Carington in the *Proceedings* of the Society for Psychical Research, we asked whether he and the Editor of the *Proceedings* would be so good as to allow us to bring out the same article in the United States approximately simultaneously. He and the Editor very generously agreed. The methodological importance of the paper is very considerable, especially in view of the apparent probability that the method can be repeated in full.

A brief study by Bugelski is published for the light it throws on an hypothesis of Leuba published in this Journal.¹

A reply is offered by Stuart to a paper by Kennedy published in our last issue. It is hoped that Kennedy, or another competent scholar interested in the issue, will continue the discussion till everything that can be done by exchange of analytical comments has been done.

A brief note by Hoffmann is offered in comment on the Reiser article in our last issue.

Papers by Rhine, and by Martin and Stribic of the University of Colorado, are in hand for the December issue.

As we feared at the time of going to press on the last issue, our publication funds are reduced almost to nothing. The extreme economies which we are practicing are forced.

As always, we are deeply grateful to the Board of Review, whose comments on the Carington article will be of great value to the reader and to the investigator who first repeats the procedure.

Gardner Murphy
Bernard F. Riess

1. Leuba, C., An experiment to test the role of chance in ESP research, *Jour. Parapsychol.*, 1938, 2, 217-221.

①
M
S
V

R

de
ex
sc
se
fo
th
wh
su
ha
ki
to
we
to
sh
en
fy

pr
le
ot
a
an

Te
ful
or
ri
mus
mus

1.

General Library
Psychology
Direct
10-24-46
56149
v. 4, no. 1-2

EXPERIMENTS ON THE PARANORMAL COGNITION OF DRAWINGS

W. Whately Carington

Section I

INTRODUCTORY

1. Genesis of the Experiments: The experiments described below arose almost inevitably out of my re-examination of various earlier researches, which I described in a paper read to the Society for Psychical Research in June, 1938, and were summarised in the *Journal* for December of that year. These studies convinced me that, despite the machinations of the malevolent hoodoo which apparently dominates the subject, the case for supposing that significant and genuine positive results had been obtained in the past from experiments of this kind was very strong. On the other hand, there seemed to be a general rule that the more carefully experiments were carried out, the less 'successful' they were likely to be. Still, if the main conclusion were correct, there should be no reason inherent in the nature of the phenomena why, given a modicum of good fortune, results satisfying the necessary criteria should not be obtained again.

I accordingly decided to undertake as soon as practicable a new series of experiments which should at least be free from the weaknesses noted by myself and others in earlier work, and to press the attack home to a point at which it would be possible to give a definite answer one way or the other.

2. Objects of the Experiments and Choice of Technique: There are four criteria to which any successful experiment in this field must conform. In approximate order of importance these are: (1) The conditions must be rigid, (2) The scoring must be unbiased, (3) The results must be statistically significant, (4) The experiment must be repeatable.¹

1. It might well be argued that, since the repeatable experiment is the very foundation of science, the last criterion is the

In all cases the general situation is essentially the same. That is to say, someone or other is required to display knowledge of some object or event, which he could not have obtained by normal sensory means or by rational inference from normally perceived facts; the accuracy or frequency of the knowledge is then assessed by some method of marking, scoring, judging or the like; and finally an estimate is made, by one means or another, of whether the extent of the knowledge shown is greater than can reasonably be attributed to chance or luck.

It is evident that if the knowledge displayed *could*, by any reasonable stretch of imagination, have been obtained by the normal processes of perception and inference, the experiment has been just so much waste of time, at least so far as establishing the reality of the phenomena is concerned; for those who are unwilling to accept them will, not unnaturally, maintain that a known cause, even if intrinsically unlikely, is to be preferred to an unknown. And it is difficult to say at what point this attitude becomes unreasonable.

Similarly, if the process of assessment is in any way subject to the prepossessions of the assessor, the results, though not necessarily invalid, are bound to be suspect; indeed, a very conscientious judge might well bias the outcome against his own views, just as a wishful enthusiast might bias it in favour of them. It is in fact essential that the judge should be as incapable of influencing the outcome in either direction, however strong his prejudices, as the percipients of obtaining knowledge of the material by normal means.²

Consequently, although it would be an overstatement to say that watertightness of conditions was the primary *object* of the experiments, it was certainly my chief preoccupation, in the sense that I was completely determined from the start that, whatever else might

(footnote continued) most important of all; but the question of repeatability does not arise till the other criteria have been satisfied, so I make no excuse for putting it last.

2. In absolute strictness, this ideal is not necessarily possible of attainment in all cases; for it is often possible for a judge to 'sabotage' an experiment by a suitable display of irrationality. Such extremes need not be considered here; for the remedy is simple - discard the judge - while, under a proper procedure, it is never possible for a judge to generate a falsely significant effect.

happen, there should be no room for argument as to whether the percipients could have obtained by normal means any knowledge of the material they might display. It has all too often happened that someone has carried out experiments with an enthusiasm and diligence amounting to devotion, and has obtained results which he reasonably believed to be convincing, only to have some critic point out that there might have been a leakage of information through some normal, if unlikely, channel which he had omitted to block. I viewed with great distaste the prospect of having months of work nullified in some such way as this, and with not less the alternative of being dragged into interminable and inconclusive arguments as to whether the observed results might or might not be due to involuntary whispering, olfactory hyperaesthesia, the subconscious interpretation of subaudible pencil scratchings, the purloining of cards by corrupted housemaids, and all the rest of it.

I accordingly laid it down as a first principle that, until after a percipient had done his work and had duly handed in his efforts, he should in no case be in the same room as the material he was required to 'guess'. In the experiments here discussed this condition has been rigidly adhered to throughout. With the exception of the second experiment, *q.v.*, no percipient, to the best of my knowledge and belief, has so much as been in the same building during the preparation or use of any of the material; and even in the 'Individual Experiments', which are not discussed here but will be repeated later, percipients never entered the experimenter's room, or even that adjoining it, until after the experiment was over.

I was well aware that, in insisting on this extreme rigidity of conditions, I would be, in the opinion of some students, running the risk of inhibiting altogether the effects I was interested in establishing. But against this two considerations weighed. In the first place, the whole contextual evidence of the subject suggests that such factors as distance and brick walls are no impediment to the occurrence of the phenomena in question, if they are to occur at all; in the second, I think I would much prefer to do a null-resulting experiment under rigid conditions than one which yielded exciting results under conditions which were sloppy - one would not merely have the satisfaction of knowing that one had done the job properly, but would be spared the indignities of controversial disputation.

Having settled this point, two further decisions had to be made, namely as to the type of percipient and the kind of material to be used.

As regards the first, beggars cannot be choosers; but there is always the question of whether to attempt to find and concentrate on a few specially gifted performers, or to rely on the 'random samples' likely to be met with in the course of mass experiments. The first plan has obvious labour-saving potentialities and may well prove the best when it comes to examining the effect of particular changes of conditions, etc., after the reality of the phenomena has been established, provided always that specially gifted performers exist; but at the present stage, when that reality is still gravely in question, I have no doubt at all that the second is not only preferable but almost obligatory. Certainly from the point of view of repeatability there can be no two opinions about it; for history indicates that so soon as an investigator has obtained magnificently successful results from some specially gifted sensitive, she either develops moral scruples or some similarly fell disease, or 'disappears into the Middle West', or gets married, or just loses her powers, and we are left to speculate as to whether the results were genuine or whether, after all, the experimenter had left some loophole unguarded. It seemed to me, therefore, that if I could get significant results from more or less randomly selected people under rigid conditions, I should be putting the phenomena on a very much wider and firmer basis than by obtaining perhaps more spectacular performances from one or two special sensitives who might easily become inaccessible for future work. It will be seen from what follows that at the time of writing (Dec., 1939) about 250 different percipients have taken part in the experiments, none of them having, so far as I am aware, any special antecedent claims to being abnormally 'psychic'.

As regards the second question, the decision is not so easy to make. Broadly, the choice is between 'restricted' and 'free' types of material. In the first case the percipient knows in advance that the object he is required to 'guess' will be one of a limited range, such as a two figure number, or a card from a particular sort of pack; in the second, the problem is more in the nature of 'What?' than of 'Which?', for he may be asked to record his impressions of what is in the experimenter's mind, or what is the subject matter of a book or a drawing, which leaves a field of almost boundless extent open to him.

It goes without saying that the restricted type of material is, on the face of it, very much the easier to deal with from the point of view of assessment; in fact relatively few attempts have been made to treat free material in a quantitative manner at all.³ But this seems to be in reality something of a delusion, for the method of assessment described below appears to be, in some respects, actually easier to apply than the very error-labile scoring of successes with cards or the like.

As a matter of fact, my first intention was to use standard 5 x 5 packs of Zener cards. I had the idea of placing ten such shuffled packs face downwards on some suitable shelf in my study every evening, asking my per- cipients to guess all ten 'down through' at a distance for, say, ten successive evenings in each case, and to go on doing this until either a statistically significant over-all result emerged or we all gave up in despair.

Fortunately, however, it was strongly represented to me (as, indeed, I had realised for myself) that this would be a procedure singularly lacking in human interest - in fact, about as tedious as could be devised; while it is not unreasonable to suppose that, if para-cognitive phenomena occur at all, they may well be to some considerable extent influenced by the degree of interest, emotion or the like associated with the material used. I was accordingly very ready to adopt the view that it would be better to use a type of material a trifle livelier than the somewhat arid austerity of Dr. Zener's symbols.

The use of drawings in one form or another at once suggested itself, partly on account of the considerable degree of success which, by inspection and assuming conditions to have been as rigid as described, appears to have attended many attempts on these lines⁴; partly because I felt that, in any event, a good deal more fun would be had by all than with the dreary task of guessing and scoring cards; while, if the worst were to come to the very worst, I should at least salvage from failure a good collection of free drawings which could hardly fail to be of some general psychological interest.

Moreover, this sort of material promised to allow subjects a degree of latitude impossible with the

3. But see such contributions as J. G. Pratt, *Jour. Parapsychol.*, 1937 I, 248 ff., or Saltmarsh and Soal, *Proc. S.P.R.*, XXXIX, 266 ff.

4. E.g., Mr. and Mrs. Upton Sinclair, Warcollier, and others.

restricted type, which might, it seemed to me, prove of decisive importance in enabling any real effect there might be to show itself. It might well happen, I reflected, that, with many people if not with all, an impression might have difficulty in reaching consciousness in its original and undistorted form but might readily appear in a disguised or symbolic shape. If this were so, it would presumably lead to null results in the case of cards, particularly perhaps Zener cards with which guesses cannot be partially right, whereas with drawings some element might be recognizable, or the reproduction might represent something more or less closely associated with the original; thus, given a method of assessment which could take account of such modifications without abandoning impartiality, it might be possible to detect genuine cognitions which would otherwise escape notice.

Finally, the decision to insist on the utmost severity of conditions as regards the possibility of 'leakage', by not having the percipients on the premises at all, in most cases, enabled me to introduce what I believe may have been a valuable relaxation by allowing them to make their attempts in the familiar surroundings of their homes, instead of coming to a laboratory or to a stranger's house, and to do so, within fairly wide limits², whenever they felt most inclined.

Thus the final plan, which was followed in all experiments except the second (*q.v.*), may be summarized as follows:

Each experiment lasted ten nights; on each night a fresh drawing, made by either myself or my wife, was put up at 7:00 p.m. in my suitably curtained, etc., study, and left there till 9:30 a.m. the next morning; the subjects of the drawings were determined immediately before their production by a substantially random method; the percipients were allowed to make their attempts at 'reproduction' wherever they happened to be and at any convenient time within the indicated limits.

This procedure, which appears to combine the maximum of freedom for the percipient with the utmost rigidity of experimental conditions has yielded results which I cannot regard as other than extremely satisfactory.

3. Arrangement of Discussion: In dealing with these experiments, I shall first give, in the next main

5. *Viz*: between 7:00 p.m. and 9:30 a.m. Cf. below.

Section, a detailed account of the experimental procedure with especial reference to the precautions taken to ensure that no knowledge of the drawings could be obtained by normal means. In Section III, I shall discuss the question of assessment generally, and shall describe the three methods, of which the first two were abandoned after trial, which were actually used. In Section IV, I shall first present the results of the Main Calculations, by which the success or otherwise of the experiments must be chiefly judged; and shall then go on to discuss such other points of interest and importance as have been investigated at the time of writing (Dec. 1939). In Section V, I shall give some account of a kind of dummy experiment intended to act as a control to the outcome of the experiments proper. Next, I propose to try to save time and trouble all round by anticipating some of the more obvious criticisms; and I shall conclude by discussing in a tentative manner the conclusions which it seems legitimate to draw from the results so far obtained.

4. Two Categorical Statements: Remembering how often I have been distressed by the omission of other workers to give essential facts, and the allegations of improper selection that have not infrequently been made against them, I wish to make the two categorical affirmations following:

(A) The whole of the work of all the percipients who took part in Experiments I to IVB here discussed has been included, except (a) two sets eliminated by Dr. Thouless from Experiment I, because they contain clues likely to give guidance to the judge, and (b) the masterpieces of two persons who sent in ten completely blank sheets each in Experiment IV. As none of my original drawings was intended to represent the Bellman's Chart, I had no scruples about consigning these, unhonoured and unsung, to my wastepaper basket. The Private Experiments, not reported here, consisted of eight ten-drawing tests, each conducted in a single evening, with five percipients selected on account of their success in one or other of the main experiments. My wife (six times) and I (twice) also acted as percipients in these, as a matter of interest, and the results will be discussed in due course. In addition, a fortuitous visitor was on one occasion invited to participate; but his results will not be counted. Thus we have a maximum of 170 drawings not here discussed, as compared with 2,194 which are. This is a negligible proportion anyway; but as a matter of fact, the

170 are, by inspection⁶, quite up to the general standard, if not better.

(B) To the best of my knowledge and belief (and, as will be seen, it would be very strange if I were in error) only two persons other than my wife and myself ever entered by study while an original drawing was displayed. One was the occupant of the flat below; he was present for a few seconds and declares (I have no doubt correctly) that he did not notice the drawing at all. The other was the lady who attends to our domestic needs; she, on two occasions during Experiment IV, came into the room before I had taken the drawing down in the morning. It is not, of course, mathematically excluded from possibility that she may represent an extensive system of espionage, and Dr. Goebbels would have us believe scarcely less fantastic propositions; but I do not think the suggestion need be taken very seriously.

5. Acknowledgements: The reader will soon see for himself how deep and extensive is my indebtedness to others. I believe that I have duly acknowledged in the appropriate passages of the text the many particular acts of help received; if anywhere I have inadvertently omitted to do so, I can only extend here my regretful apologies to whoever may have been so neglected.

But in addition to these specific acknowledgements, I must express my more general though not less sincere gratitude to the members of the Cambridge Committee, Professor Broad, Dr. Thouless and Mr. Oliver Gatty, who have aided, abetted and supervised my labours; to the first named in particular I am especially indebted, not merely for the several long and tedious tasks he has undertaken, but for much personal as well as official encouragement and support. I am also very deeply obliged to Dr. J. O. Irwin, who kindly consented to join the Committee in October, 1939, for invaluable guidance and help, which it has been impossible adequately to acknowledge in the text, in the matter of statistical treatment. I am also much indebted to Dr. E. J. Dingwall not only for constant and stimulating criticism but for moral support and practical help at many stages of the work. To Mr. M. T. Hindson, whose perspicacious judging has been a factor of prime importance in the success of the work, I must pay a very special tribute; and similar acknowledgements are due to Mr. H. F. Saltmarsh who undertook

6. I.e., without applying formal tests of significance. The words 'by inspection' are used in this sense throughout.

the equally laborious 'control marking'.

Last, but scarcely least, my thanks are due to the percipients, without whose disinterested cooperation nothing could have been done at all. Theirs has been the somewhat thankless task of doing a not very interesting job for the benefit of someone whom they usually had never so much as heard of, and without even having, in most cases, the trifling satisfaction of knowing what degree of success had attended their efforts. I hope that any of them who may read this will realize that it has been quite impracticable to render individual reports, or even thanks, except in a very few special cases, but that their good offices have been none the less appreciated.

Section II

EXPERIMENTAL PROCEDURE

PRELIMINARY: I have found it a good deal more difficult to write this Section in anything like a coherent form than might be expected. It is of course easy enough to describe factually what I did, and what precautions I took to avoid this pitfall or that; but it is not nearly so easy to make clear why I did it, without entering upon a variety of digressions which are not irrelevant and out of date.

The trouble is that, in order to keep things moving at a reasonable speed, it has usually been necessary to plan experiments in advance before the outcome of those already performed was fully known or adequately digested. If I had not done this, which involved a certain amount of would-be intelligent anticipation, I should probably still be trying to squeeze the last drop of juice from Expt. I, and still not quite sure that there was any juice to squeeze.

On the other hand, the procedure had the drawback that probably no experiment would have been planned quite as I did plan it, if I had known at the time what I knew later. Consequently, although certain features, notably that of preserving completely watertight conditions, are common to all, the policy behind them has had to be to some extent modified as the work progressed and as the nature of what appeared to be happening was more fully appreciated.

In particular, I originally intended to assess my results by the method of matching the drawings made by the percipients in any experiment against the originals used in that experiment. This method, which is discussed at some length below, is based on the assumption that if a percipient, as the result of a 'paranormal' cognitive process obtains a correct impression of an original which he cannot see, he will do so (or at least is most likely to do so) on the same occasion as that on which the said original is displayed. This assumption is now known to be untrue, at any rate so far as the material and percipients involved in these experiments are concerned; but realisation of this was delayed by the fact

that the first ten sets of drawings examined were exceptional in showing a significant positive result⁷ when this method was used. Naturally enough, therefore, the second and third experiments were designed primarily with a view to continuing with this technique; and both parts of the fourth were arranged so as to permit of it, though the experiment as a whole was intended to be based on a comparison between the two parts, each taken *en bloc*, rather than on the matching of drawings against originals within the parts.

In these circumstances, I have decided that the best plan will be as follows: In the present Section I shall describe the experiments, mainly from the point of view of procedure and precautions against leakage, adding only a few notes on the outcome by way of a kind of running commentary. In Section III I shall deal with the three methods of assessment which were tried out at different stages, giving fairly full results for the first two, as a matter of interest. Only after this shall I embark on a discussion of the results obtained by the third method, which alone are submitted as offering positive evidence for the occurrence of paranormal cognition.

EXPERIMENT I; 1. General: So far as the points of procedure and technique here considered are concerned, this first experiment is so much the prototype of all others except the second that it will be necessary and sufficient to describe it in detail, indicating only where necessary the details in which Expts. III, IVA and IVB differ from it.

The general procedure has already been outlined and may be summarised as follows:

On each of ten successive evenings, beginning on that of Feb. 1st, 1939, a different simple drawing, made by either my wife or myself, was exposed at 7:00 p.m. in my study at 5 Fitzwilliam Road, Cambridge, England, and was left in position till 9:30 a.m. the next morning. Percipients were asked to draw, in books provided for the purpose and at any time within the period of exposure convenient to themselves, the best 'reproduction' they could manage of what they thought each drawing represented, or (which probably came to much the same thing) whatever came into their minds when they made the attempts.

⁷ Cf. p. 18 below.

Inasmuch as it was at this time intended to assess the degree of success achieved by the method of Direct Matching⁸, it was vitally important that the judge, in this case myself, should have no normal knowledge of which drawing by any percipient was intended to represent which original. Accordingly, the books were handed in by the percipients at the end of the experimental period to Dr. Thouless, who detached the sheets from them, assigned to each sheet a suitable 'code' number, and shuffled them before passing them to me.

The results of the matching, which started by appearing very promising but ended by giving no significant result, will be discussed in the next section. At present we are concerned only with those points of experimental technique which are of fundamental importance irrespective of what method of assessing the results is finally used. The following details should be carefully noted and should serve to make all clear.

2. Percipient's Books and Instructions: Each percipient was provided with a specially printed book consisting of a cover and ten pages measuring 13" by about 8½". These pages were perforated about ¼" from the left margin, so that when they were torn off at the perforation they were just foolscap size, viz, 13" x 8". Suitable spaces were provided at the top of each sheet for the percipient's name, the hour of making the attempt, the code number to be inserted by Dr. Thouless, and for notes by the percipient as to his degree of confidence and the occurrence of visual imagery. Another space for general notes and impressions was ruled off at the bottom, leaving an area of about 7" x 8" for the actual drawing. The ordinal number of each sheet was printed on it to the left of the perforations. Full instructions for the percipient were printed on the cover; these are reproduced in Appendix I. To guard against the possibility that some percipients might press so heavily on the paper as to indent the sheet below, and thus give some clue to the order in which the attempts were made which would vitiate the process of matching, a sheet of stiff card, about .5mm thick, was provided with each book, and percipients were instructed to insert this between the sheet they were using and the next. No such indentations were in fact noted, but I think the precaution was worth taking.

8. Cf. Section III, 2.

3. Photograph of the 'setting': In order to give the percipients some idea of the setting of the experiment, and to form some kind of a link between them and the location of the drawing to be reproduced, photographs were also distributed showing the relevant parts of my study with a blank sheet of paper pinned in the position to be occupied by the drawings in the course of the experiment. I am much indebted to my friend Mrs. Ramsey, of Ramsey and Muspratt, for the trouble she took in this matter. In the later experiments, it was not always possible, owing to irregularities of supply, to ensure that every percipient had a photograph all to himself; but most had, and I think there were very few, if any, who had no opportunity of looking at one.

4. Percipients: A total of 37 percipients took part in this experiment, and the number was made up as follows:

a. The bulk of the group was formed by 27 students from Dr. Thouless' lecture class, all of whom were training for the teaching profession. Of these 19 were women and 8 were men, and their ages ranged from about 22 to 25, though no exact data were sought; the other participants were appreciably older than this.

I am particularly obliged to Dr. Thouless for setting the ball rolling by obtaining the co-operation of these percipients, as well as for undertaking the work of randomization already mentioned.

b. The S.P.R. was represented by Mr. and Mrs. Salter, Dr. and Mrs. Thouless, Mr. and Mrs. Tyrrell, Professor Broad, and Dr. E. J. Dingwall.

c. In addition, two ladies resident in Cambridge, friends of my own, were induced to take part.

5. Production of the Originals: To determine the subject matter of each drawing I opened a copy of Chambers' Mathematical Tables at random, noted the last digits of the first three or four entries encountered, turned to the corresponding page of Webster's Dictionary, and took as the subject for the drawing the first reasonably drawable word found on or after that page.

9. 'Three or four' because the terminal digits of the first four entries met with might form a number greater than the number of pages in Webster.

This method was by no means perfect; in particular, it led to the use of certain originals which I later came to regard as unsuitable for the purpose, on account of their vagueness or unfamiliarity. But it served the purpose for which it was intended, namely that of ensuring that no percipient could possibly forecast by rational means what the nature of the drawing would be, and of guarding against the effects of possible coincidental thinking prompted by contemporary events.

In general, the drawings were made between 6:30 p.m. and 7:00 p.m. on the evening each was to be used; but on one occasion (9.11.39) the necessary absence of both my wife and myself¹⁰ during the afternoon and evening led to the drawing being prepared at about 1:30 p.m. The room concerned was, of course, carefully curtained and locked (see below) from that time onward.

6. Nature of the Originals: The pages in Webster found by the above described procedure and the words selected for illustration, together with notes on the choosing, are given below:

1. p. 323: 'Bracken sickness' was rejected, a BRACKET, illustrated in the dictionary, was drawn.

2. p. 2886: The first word on the page is Water Ox or Water Buffalo; a picture of a horned bovine animal (not conspicuously of the genus *Bubalus*) was drawn and labelled BUFFALO (*etc*).

3. p. 385: Various words such as Embalm, Embark, Embank, Embarrass, and Embassy were rejected, and an illustration of an EMBATTLED FESS (heraldic) was drawn.

4. p. 1496: A considerable number of words such as Maniac, and others beginning with Mani- were rejected; but Manicure suggested hands, and a left HAND was drawn with fingers spread.

5. p. 632: A great number of compound words beginning with Cross was passed over, and an illustration of CROSS STITCH was copied.

6. p. 2811: Vacillator, vacillatory, vacoa, and a number of words beginning with Vacu- were rejected; the illustration given of a Vacuum Bottle was judged too complicated for the purpose, and an ordinary BOTTLE was drawn, with a label marked VINO, by way of preserving the V.

7. p. 969: Flite (Miss) and various words in Flit- were passed over. FLITTERMOUSE was illustrated by a Bat with outstretched wings.

10. Hereinafter referred to as H.S.C. and W.W.C., respectively, when convenient.

8. p. 1644: The first word was Net Blotch. H.S.C. decided to illustrate a NET and, to make it more interesting, drew a sketch of a man pulling a net with fish in it out of water.

(This subsequently proved to be a somewhat unfortunate original, for opinion was considerably divided as to whether the man, the fish, the waves, the net or the beach should be regarded as the principal feature.)

9. p. 1519: Two earlier pages were rejected as providing nothing suitable. On this page, the illustration of a Meal Moth was rejected as being too like the Bat already drawn, and finally an illustration of a Meal Worm BEETLE was approximately copied.

10. p. 97: The word ANCHOR was the first suitable for representation, and a picture of an Anchor was drawn.

The drawings illustrating these words were drawn by W.W.C. and H.S.C. alternately, the former starting with No. 1. They were done on sheets of white paper substantially identical with those issued to the percipients and were of a size to fill the 7" x 8" space more or less completely. All except the last were line drawings in Indian ink made with a broad nib; in the case of the Anchor, the outlines were filled in with ink, thus producing the effect of a silhouette with shaft and limbs about 10 mm broad.

Whatever objections may be made either to the objects finally selected or to the method of selecting them, and I have already said that I do not consider either to have been ideal, I trust that all will agree that there was no possible means whereby the percipients could possibly have inferred what the originals were, or even what class of object they were most likely to represent. This is all that matters from the evidential point of view, and such questions as whether the process of selection was truly 'random' in a strict mathematical sense are irrelevant from this point of view.

Personally, I do not now consider the method to be a very good one. I think it would be much better to use an artificial dictionary consisting of words specially picked for their suitability, and to apply, perhaps, a more convenient and more truly random method in selecting from it. In Experiment VI, for Example, which will be reported later, I have used a list of 216 'suitable' words, arranged in six blocks, six rows and six columns, so that the selection can be made by throwing three dice.

As regards 'suitability': Speaking quite provisionally and by inspection only, I am now fairly sure

that there is usually no question at all of percipients in any sense *copying* the original; and it seems as if it is the 'idea' rather than the form of the drawing that is cognized - though admittedly the word 'idea' is unpleasantly vague. If this is so, then the first criterion of suitability is that the idea should be reasonably familiar, for otherwise it will not be recognised and cannot be reproduced; while the second, I think, is that it should be unambiguous. But this is a digression.

7. Location, etc., of the Originals: The room in which the originals were exposed is a kind of study-bed-sitting-room on the first floor of a house looking south over a relatively unfrequented road. There are no houses immediately opposite, and the nearest that could be said to 'overlook' the room is at a distance, I suppose, of about half a mile. In addition to the ordinary defences of the house, the outer door of which is always locked in winter, a Yale type lock was specially fitted to the door of the study at the beginning of the experimental period, and the keys did not leave the possession of my wife and myself. The precaution was highly supererogatory, but I made a practice of always locking the door whenever I left the room for more than a very few minutes, and even after retiring for the night. The chance of anyone making an unauthorised entry during any period of exposure may be regarded as completely negligible.

The curtains with which the room is normally provided are sufficiently opaque, as I have tested by careful observation, to prevent anyone seeing from the road outside so much as whether there is a sheet of paper in the position where the originals were placed - let alone distinguishing any design there might be on it. As a further precaution, however, an additional thickness of fairly heavy rep was drawing-pinned over the lower half of the window, through which alone a glimpse of the drawing might be supposed to be obtained, during the exposure periods. This was always put in position, and the ordinary curtains drawn and secured over it, before the original was displayed.

Each original was pinned in turn to the centre of the top shelf of a bookcase which stood against the wall to the west of the window; this brought the upper edge of the paper to a height of about 5' 8" above the floor. The room was not prepared in any way, except by taking down two or three portraits from the wall near the bookcase and removing (subject to the one exception noted below) a calendar which normally hung just below the

position selected¹¹: The portraits were not removed for the later experiments, though the calendar was. In the case of this experiment only, intending percipients from Dr. Thouless' class were invited to visit the room before the experiment started; but only two did so.

At the end of each exposure period, namely at 9:30 a.m. on the morning after exposure, or a little later, I was careful to remove the drawing and lock it in a steel box before any 'outsider' had access to the room or the curtains were drawn back. With the trivial exceptions noted in the Introduction above, I can say with complete confidence that no one except my wife and myself saw, or could have seen, any of the originals during the course of the experiment or before the great bulk of the books had been received by Dr. Thouless¹².

Finally it may be just worth noting that the lights were not left on in the study if no one was in it.

I need hardly say that both my wife and I were scrupulously careful not to mention or hint at the nature of the originals to anyone at all during the course of the experiment, or to make any remark that might give a clue to their nature. This was easy enough, for we were not at that time acquainted with any of Dr. Thouless' students, while our contacts with the other percipients were very slight.

8. Concentration, etc: No special effort was made by us to 'concentrate' on the originals during exposure, nor were any instructions to this effect given to the percipients. But the original for each evening would naturally be more or less in the minds of my wife and myself, while the position selected ensured that I at least would be fairly often reminded of it.

-
11. It is not uninteresting to note that on one occasion I rehung this calendar for reference during the day and inadvertently omitted to remove it in the evening. On the night in question, and on no other, an excellent picture of a calendar of the same general type and proportions as mine was drawn by one of the percipients.
 12. The slight reservation implied in the last sentence is necessitated by the fact that three sets of drawings came in very late, and a few people, not in touch with the percipients concerned so far as I am aware, were shown the originals before these sets arrived. I mention the point only for the sake of meticulousness, for it certainly has no bearing on the results.

While on this topic, I may record the wholly provisional and personal impression that attempts at 'concentration' by percipients are likely to do more harm than good, except in so far as they denote no more than trying to free the mind from thoughts of which the origin can be identified. In the fourth experiment, for example, more than one percipient reported in such terms as "Complete blank, even after fifteen minutes Intense Concentration". Such prodigies, I feel, are better qualified for Cabinet Rank, or at least as Ministers of Information, than as participants in experiments of this kind.

9. Results: As I have already indicated, and shall be obliged to emphasise almost *ad nauseam* later, significant positive results have only been obtained by inter- as opposed to intra- experiment comparisons. These will be fully described and discussed in due course, while some account of the results of the matching technique will be given in the next Section. There is accordingly not very much to be said here, particularly as we cannot say whether an experiment of this type is successful or not, taken as a whole, until we have something to compare it with. But a few first impressions may be worth recording, if only as a matter of historical interest.

Speaking personally, I shall not easily forget the thrill I received when I opened the very first set of the first batch of randomised drawings passed to me by Dr. Thouless and found a fine sketch of a Hand (original No. 4.) backed up by an unmistakable Fisherman and a kind of slender Jug, which were pretty good shots, in the circumstances, at Net and Bottle respectively; and the next set contained a battlemented archway, which was by no means bad for Embattled Fess. Nor were these mere isolated examples, for throughout the 37 sets of the experiment H.S.C. and I kept on finding unmistakable 'winners' of one kind and another, which soon made us feel that there was something going on which 'mere chance' would be unlikely to account for. Of course we knew very well that these impressions were purely subjective, quite likely to be due to wishful thinking, and in no sense evidential. It was only a matter of personal judgement that made us think it unlikely that people would have drawn as many Hands, Cows (for Buffalo) and so forth as we found, if there had been no Hand or Buffalo among the originals; but it is perhaps just worth recording the fact that our immediate reaction to the drawings, before we in any way 'knew the answer' was that the experiment had been very definitely a success. This

feeling was confirmed and enhanced when we learned the result of our matching of the first ten sets; it dwindled to little more than a conviction based on a belief in the soundness of our own judgement as that result was gradually whittled away to nullity by the outcome of subsequent matchings; but it has been more than justified by the application of the method of 'palpable hits' and inter-experiment comparison.

EXPERIMENT II; 1. General: At a very early stage of the investigation it had been suggested that percipients might be able to match their drawings for themselves against the originals they were intended to represent, or which had in some measure influenced their production, a good deal better than any outsider could do. This seemed reasonable enough, for it might very well happen that the person who produced the drawing might recognise some element in the original which had formed part of his impression, but had been imperfectly portrayed; or he might have personal associations, unknown to the experimenter, which would lead him to connect one of his drawings with an original in a way which no one else could do. Since at this time (Feb. 1939) I was still proposing to rely on the method of matching for the assessment of results, I was naturally anxious to see whether this was the case.

It was accordingly arranged, by the courtesy of Professor Bartlett, to carry out an experiment on Feb. 11th, in the Cambridge Psychological Laboratory, with the contemporary 'Part II' Psychology class. I am very particularly indebted to Mr. Grindley for organising this experiment, as well as to others mentioned below for the parts they played.

2. Procedure: Although, in this experiment, the percipients were in the same building as the experimenter during the production and 'exposure' of the drawings, the conditions were not less rigid than those for the other experiments. The class, numbering seven women and thirteen men¹³, was assembled in the practical class room on the second floor of the building, and Professor Broad and Dr. Banister kindly consented to invigilate it. The originals were produced by me in a small room on the ground floor and the opposite side of the building (which, it may be noted, is unusually well insulated acoustically) and were done under the supervision of Dr. Rawdon Smith and Mr. R. C. Oldfield. Communication between the

13. This included one percipient who had taken part in the first experiment.

two rooms was maintained by Mr. Grindley with the aid of a buzzer and telephone, and the relevant corridors were patrolled by two of the laboratory assistants. When the time came for the originals to be taken upstairs to be matched by the percipients, as explained below, this was done by Mr. J. C. W. Craik, who had not been in either room during the progress of the experiment. Even Mr. Grindley, who informed the invigilators, and through them the percipients, when to start and stop attempts to reproduce each original, was not actually in the same room as myself, but in a partitioned recess, and could not see what I drew. So the possibility of leakage from experimenter to the percipients may be regarded as completely excluded.

Each percipient was supplied with a pad of ten 6" x 4" cards, numbered 1 to 10. Each card had spaces in which the percipient was to enter his name, the identifying letter of the original to which he would eventually match his drawing, and the degree of confidence of the matching.

I began by briefly explaining the nature of the experiment to the class, and at once retired to the room downstairs in which the originals were to be produced. Since at this time I had no idea what these originals would represent, it was impossible for me to indicate their nature to the percipients, even inadvertently.

The subjects for the originals were determined as follows: A set of cards, numbered 1 to 10, was randomised by Mr. Oldfield and Dr. Rawdon Smith, using log tables, and was then inserted by Mr. Grindley, face downwards and in a haphazard manner, between the pages of a copy of the Concise Oxford Dictionary, which was also held face downwards. Care was taken to ensure that Mr. Grindley did not see between what pages the cards were inserted, and, as already mentioned, he was not in a position to see what I was drawing or had drawn until after I had finished all drawings. Moreover, to make assurance doubly sure, I was careful not to mention the subject of any original aloud or to make any remarks concerning it, though I occasionally wrote down some explanatory comment and showed it silently to Dr. Rawdon Smith.

To select the subject for the various originals, I turned the dictionary right way up, opened it successively at the cards numbered 1, 2, 3 . . etc, and took the first reasonably drawable or 'illustrable' word that came in each case.

The words found and drawings made to illustrate them were as follows:

1. SPINNING: I drew a spinning top.
2. PARNASSUS: Illustrated by a roughly outlined mountain at the foot of which was drawn a Greek temple with pillars, steps and pediment.
3. JENNET or small Spanish horse: A horse was drawn.
4. EXFOLIATE: I drew some leaves.
5. BRIM: Illustrated by a sort of Goblet or chalice with a heavy brim or lip.
6. SHOOTING: I thought of drawing a field gun on wheels, and made a note to this effect on the bottom of the sheet; but finally drew a sporting gun with indications of a puff of smoke.
7. ANCESTOR: Illustrated by an old man, with a bald head and long beard, leaning on a stick.
8. PRAWN: I drew as good a prawn as I could, but it might equally well have been a shrimp.
9. STANDARD: I tried to draw the Royal Standard, and produced an unmistakeable flag; but the lion rampant was more suggestive of a demented monkey.
10. THRONE: I drew a kind of wooden arm-chair with upright back and a very bad indication of a crown on the top.

These originals were then randomised by Mr. Grindley and Dr. Rawdon Smith, were lettered A to I for purposes of identification, and were taken up to the classroom by Mr. Craik. They were there pinned up in view of the percipients, who were asked to write on each of their cards the letter of the original which they thought it most closely resembled and to grade this assignment, , , or , according to the degree of confidence they felt in it. When they had done this, they noted their matchings on a separate slip of paper and handed in their pads to Professor Broad. Meanwhile, Dr. Rawdon Smith, Mr. Oldfield, Mr. Grindley, and I remained *incommunicados* in the downstairs room; but so soon as we had been informed, by telephone, that all pads had been duly handed in, we went up and the true order in which the originals had been produced was disclosed.

3. Results: From the point of view of assessment by matching, and particularly as regards the question of whether the percipients could match their own

drawings better than an outsider could, this experiment was a most discouraging failure; for the percipients only contrived to make 17 correct matchings between them, which is three below expectation.¹⁴

On the other hand, when I come to view this experiment in retrospect, it is clear to me that it was of vital importance, and marked the turning point of the whole investigation.

To start with, of course, I consoled myself for the apparent failure by reflecting that the conditions under which the percipients - and, indeed, the experimenter - worked were so different from those of the first experiment as amply to account for any deterioration of results.¹⁵ In the first place, the elaborate precautions taken to exclude leakage, admirable and necessary as they were, inevitably produced an atmosphere of bustle and strain likely to be much more inimical to success than that prevailing in percipients' own homes; in the second, a number of persons forming a group might not behave in the same way as they would if they were functioning independently; in the third, there might be a big difference between producing a drawing at one's own time and inclination, and doing it to order whether one felt like it or not; finally, whereas the trials of the first experiment were spaced at intervals of a day, those of the second were separated by no more than some five minutes. Any of these factors, or perhaps others not enumerated, might easily be held to have militated against success.

But as soon as I began to examine the drawings themselves at all closely, I received the impression that the appearance of failure was illusory and due to the breakdown of the matching technique rather than to the percipients not receiving satisfactory impressions. So far as superficial indications went, indeed, they seemed to have done rather better than those of the first experiment. As a rough and ready test, we find that they themselves assessed thirty-one drawings as 'alpha' grade resemblances to originals - an average of 1.5 per set - which was confirmed by H.S.C. working independently, who gave 30; whereas, in the first experiment, H.S.C. and I

14. For slightly greater detail, see Appendix II, p. 118.

15. As it happened, the first ten sets of experiment I, which gave a significant result, were matched and the outcome was known, before the second experiment was carried out; thus the drop to a null result appeared a sad contrast to what had gone before

working together gave thirty-nine 'alphas' in thirty-seven sets, an average of only just over 1.0 per set. Since the assignment of 'alphas' depends solely on the personal opinion of the judge, this is far from being conclusive; but there is certainly a strong suggestion that, so far as intrinsic resemblance to the originals was concerned, the percipients of the second experiment were not inferior to those of the first.

But there was very much more to it than that, for inspection left little doubt that the drawings which 'caught the eye' as manifest successes in the second experiment were not of objects commonly drawn in the first, and *vice versa*. Among the second experiment drawings, for example, was a really beautiful spinning top - almost a point to point replica of my original - but there was nothing of the kind in those of Experiment I. Again, there were some five or six guns of sorts, or mention thereof, in the second experiment drawings, but nothing at all definite in the first. Again, whereas the first experiment drawings showed about as many hands and cows as there were guns in those of the second, the latter produced no more than a glove and a finger-nail in the one case and two cows in the other.

I am deliberately using indefinite language here, corresponding to the qualitative impressions obtained from inspection; but the conclusion was very strongly indicated that, *if the order of the drawings were ignored and the various sets considered as wholes regardless of the positions of their constituents, the drawings of the second experiment resembled the originals of the second experiment, which they were intended to resemble, much more closely than they resembled the originals of the first experiment, which they were NOT intended to resemble; and that the drawings of the first experiment resembled their own originals much more closely than they did those of the second.*

This was an enormous and vital step forward, though it took me some little time to realise its full implications; for it involved jettisoning altogether the natural assumption that success would necessarily take the form of good resemblances being produced synchronously¹⁶ with the originals which they resemble.

16. That is to say, within the same period of time as that during which the original concerned was displayed; or alternatively, showing a one to one correspondence as to order with the original concerned.

If I had not performed this experiment and been impressed by the spinning top and the relative profusion of 'Shootings', coupled with the paucity of hands, and similar observations, I might well have gone on for an indefinite period vainly trying to repeat the remarkable 'flash in the pan' matching success obtained with the first ten sets of the first experiment.¹⁷ But as soon as I realised that the basis of enquiry had to be shifted from comparisons within an individual experiment to comparisons between different experiments, progress became possible; and as soon as I began to study seriously the displacement of drawings from their expected positions, results of major importance began to emerge.

EXPERIMENT III: 1, General: This was the smallest of the 'mass' experiments so far conducted and, as will be seen later, conspicuously the least successful. Only eleven percipients (four women, seven men) took part. These were members of Mr. O. L. Zangwill's Workers Educational Association psychology class, to whom he was so kind as to introduce me. I was glad to take an opportunity of having a few minutes talk with this class on the evening before the experiment started, for I thought that the personal contact so obtained might increase the chances of success; the results suggest, however, that its effects, if any, were in the opposite sense.

2. Procedure: The procedure adopted was precisely that of the first experiment, with the trifling exception that H.S.C. did the odd-numbered drawings in this case, and I the even, instead of the other way round. That is to say, ten originals were exposed in my study, in the same position as before, on each of ten successive evenings from 7:00 p.m. till 9:30 the next morning, starting with that of Wednesday, March 8th.¹⁸ Exactly the same precautions as regards curtains and door-locking were taken as before. The books used were some that had been left over from the first experiment, so that I need not describe them again; and they carried, of course, the same instructions of which only the dates needed alteration. At the end of the experiment they were handed in to Mr. Zangwill, who very kindly carried out the necessary randomisation and numbering of the sheets before

17. Cf. p. 18.

18. In the interests of strict accuracy it may be noted that these periods of exposure were twice interrupted for an hour or so while I carried out two 'individual' experiments. Percipients were warned not to make attempts during the periods of these interruptions.

passing them on to me.

3. Originals: The originals were selected by precisely the same process, using Chambers Tables and Webster's Dictionary, as was used in experiment I. The subjects thus chosen and illustrated were:

1. VIOLIN;
2. A BIRD (supposed to be a 'Corn Bunting', but actually of a non-specific passerine appearance);
3. A FISH;
4. CLEOPATRA'S NEEDLE;
5. A SKULL;
6. The Planet SATURN;
7. A FROG;
8. FLEUR-de-LYS;
9. COTTON APHID (this was blacked in by H.S.C.);
10. A pair of SPECTACLES.

4. Results: The matching was done by H.S.C. and myself, jointly, as before. I had hoped that we might be able to repeat the success of the early sets of the first experiment; but, as will be seen on reference to p.32, this hope was sharply disappointed.

None the less, the drawings sent in were far from lacking in interest. For example, there was one out and out success for the spectacles, whereas I had found only one pair of spectacles in the fifty-seven sets of the first and second experiments and the twelve sets which, up to this time had been produced in the individual experiments not reported here. Again, the association between Saturn and rings is, in my mind at least, extraordinarily close; and in these eleven sets there were three instances of rings being drawn, though not on the right days, two being of the jeweler's variety and one a very solid-looking annulus, as compared with only one - and that as a quite secondary ornament on one of the hands - in all other sets; that is to say, rings were about nineteen times as frequent in these sets as in those among the originals for which Saturn did not figure. Further, there were three pyramids, which may be regarded as pretty close associations to Cleopatra's Needle,

as compared with one rather doubtful example in the other sixty-nine sets mentioned. No one knows better than I that this sort of thing is not coercive evidence; but it served to support to some extent the view to which the outcome of the second experiment was driving me.

In view of this, and of the fact that three of these percipients showed great promise (one was of outstanding interest) when subsequently tried out under the conditions of the Individual Experiments, it is not at all easy to understand why this experiment should have been so very unsuccessful, not only absolutely but relatively, as will be seen in due course. So far as I can judge, the percipients were above rather than below the average in intelligence and good will, while my personal contact with them, if this be supposed potentially deleterious, was scarcely greater than that with the class of experiment II. On the whole, while content to leave the point for the present as a minor mystery, I am inclined to suspect over-conscientiousness operating by way of a kind of 'reversed effort'.

EXPERIMENT IV, A and B: 1. General: The results of these first three experiments left certain quite clear and definite impressions on my mind. These were, first, that although I was not yet in a position to prove it formally, successes were being obtained which could not plausibly be attributed to chance; second, that the matching technique was no good for the purpose, because most of the best 'hits' were made on the wrong occasion; third, that the only way to bring out the effect I was sure was there was to adopt some quantitative plan for comparing experiments as wholes, so as to see whether there was, as seemed plainly visible to me, a significant tendency for drawings to resemble their own rather than other originals, when sets of the one and series of the other were considered as wholes.

I also had to face the difficulty, which is likely always to be embarrassing, that I could not decently rely to an indefinite extent on the good offices of other people for scoring results. I accordingly planned a large scale experiment in two parts, A and B, each forming as it were an experiment in itself, and arranged it in such a way that, after suitable randomisation, I could score each set of drawings sent in against both series of originals used, without knowing to which it was supposed to correspond. The plan did not work very well, because, although I used different percipients in the two parts (only one worked in both), which I imagined would

keep them satisfactorily distinct, displacement took place *between* the two parts just about as freely as it had previously done *within* the previous separate experiments, though the double experiment as a *whole* was highly successful. But we shall consider all this at a later stage and need say no more about it here.

2. Percipients: This experiment was on a very much bigger scale than any that had preceded it. About 370 books were sent out to various people whom I thought might be willing to participate, and 183 came back.¹⁹ These 183 percipients were made up as shown in the following table:

Table I

Part A

Group	Women	Men	Total
1. Cambridge Students	0	8	8
2. Cambridge S. P. R.	7	3	10
3. Cambridge Residents	0	2	2
4. Edinburgh Residents	3	2	5
5. Dutch S. P. R.	34	34	68
6. Duke University Group	6	6	12
Totals	50	55	105

Part B

1. Cambridge Students	4	9	13
2. Cambridge S. P. R.	1	3	4
3. S. P. R. Members	13	4	17
4. Edinburgh Students	12	14	26
5. Edinburgh Residents	2	2	4
6. Dutch Spiritualists	2	12	14
Totals	34	44	78
GRAND TOTALS	84	99	183

This table is probably somewhat inaccurate as regards the distribution between women and men. This is because a considerable number of percipients, despite explicit instructions, omitted to write their full names on the cover of the book; consequently, as I had not asked them to state their sex in so many words, I was

19. This is not counting the two books of ten blank sheets already referred to.

often left in doubt on this point. Provisionally, I have counted all such delinquents as men, on the assumption - which has at least the merit of gallantry - that no woman would be so negligent as not to read the instructions.

My thanks are due to Dr. Thouless in the matter of coopting most of the Cambridge students, to Mr. Fraser Nicol for the Edinburgh residents, and especially to Dr. Mary Collins for the handsome contribution of twenty-six Edinburgh students.

The outstanding features are, of course, the splendid effort of the Dutch participants and the Duke University Group mobilised by Dr. Rhine. I am particularly grateful to Mr. J. J. Poortman for the immense amount of trouble he took over the Dutch S. P. R. percipients, not merely in circularising his members and in translating my instructions into Dutch, but in translating all the key words in the 'Notes and Impressions' subjoined to the various drawings, so that there should be the minimum of difficulty in judging. I am also much indebted to Mr. Nout, of the Nederlandsch Spiritualistisch Genootschap, for analogous good offices in respect to his Association.

3. Books, etc: The drawing books used in this experiment were substantially identical in form and arrangement with those used in experiment I. A few trivial alterations were made in the instructions with a view to extra clarity, but the only important difference lay in shifting the spaces for Name and Age to the outer cover instead of the inner sheets, so that I should have no clues from these sources as to whether the book I was scoring was an 'A' or a 'B'. For the same reason I resisted the temptation to use two different colours for the covers, for I feared that in the process of trimming or the like particles of cover might be transferred to the pages and thus give an indication of which kind of book it was.

4. Originals: The subjects for the originals were selected by the same substantially random method as was used in the first and third experiments. They were:

For Part A:

1. DODO.
2. FLAG (This was drawn as a black Flag with a conspicuous *white* Latin Cross)

and
ext
tak
ly
BOC

(f
of

3. CASTLE (very 'schematic' and toyshop-like).
4. MOUSTACHE.
5. STOP-COCK or water tap.
6. BUTTERFLY.
7. BOOT (The dictionary word was shoe, but this was illustrated by what is commonly known as a boot in England, so a boot was drawn.)
8. FAN.
9. BALANCE (A chemical balance was drawn).
10. SCISSORS.

For Part B:

1. TREE (Illustrated by a solitary tree of indeterminate species).
2. FISH-SPEAR (Commonly known and thought of as TRIDENT).
3. BENCH (A kind of garden seat was drawn).
4. GEISSLER POTASH BULBS (Commonly known as BULBS, *tout court*).
5. HAMMER.
6. EWER, (Illustrated by a single-handled jug with a constricted neck).
7. Boat (The dictionary word was SHIP, illustrated by a drawing of a full-rigged ship; but H.S.C., whose turn it was, thought this was too difficult to draw and drew a fore-and-aft rigged sailing boat instead. This was shown in silhouette).
8. WINDMILL.
9. ARROW.
10. SHELL (A more or less Whelk-like Shell was drawn).

These originals were drawn, as before, by W. W. C. and H. S. C. in turn, the former starting, except that external circumstances made it impossible for H. S. C. to take her turn on the sixth occasion; W. W. C. accordingly drew both STOP-COCK and BUTTERFLY, and H. S. C. did BOOT and FAN.

These twenty originals were drawn and put up (from 7:00 p.m. till 9:30 a.m. as before) in two sequences of ten consecutive evenings each, from Wednesday, April

26th to Friday, May 5th, and Wednesday, May 10th to Friday, May 19th, 1939, inclusive, respectively. I hope I need hardly say, except as a matter of form, that the same precautions were maintained, as regards locking of doors and curtaining of windows, as were used in the previous experiments.

5. Randomisation: Despite the failure of the Method of Direct Matching used in experiments I, II, and III, it was decided to persevere with this, in the hope that H. S. C. or I might recapture her or my initial virtuosity. It was accordingly necessary to effect a double randomisation, first of the books themselves, so that I should not be able to tell which was intended for A and which for B: and second of the sheets in each book, as was done for the books of experiments I and III.

This very considerable labour was most kindly undertaken by Professor Broad, to whom I am greatly indebted for carrying out the work. Each book was identified by means of a pair of letters, such as *Fm* or *Pt*, taken from a suitable key table, in which the A or B character of the book was entered. These letters were written on the original cover of the book, which was detached, and on a temporary paper folder in which the torn out sheets were placed. Another key table, giving a large choice of random sequences of the numerals 1 to 10, suitably prepared from logarithm tables, was also supplied, and the sheets of each book were numbered in accordance with one of these sequences before they were torn out; they were then re-arranged in the order 1, 2, 3, . . . 10, which constitutes effective randomisation for this purpose. Identification was ensured by writing the appropriate two letters in the margin of the table, and the serial number of the sequence used on the cover of the book. As a final precaution against the matching judge being able to draw any rational inferences as to the most probable true order of the sheets from a study of the distribution of digits in the logarithm tables, approximately half the books were numbered backwards instead of forwards, suitable indications being given in the key table.

Since none of this is relevant to the results later presented as evidence, though some of the matching work may be usefully informative, I shall not go into further detail here.

6. Results: It seems worth recording that during the period of this large double experiment, extending

over more than three weeks, both H.S.C. and I were extremely 'stale' and - not to put too fine a point on it more than a little bored with experimenting in general and with the necessity of being on the spot to produce an original every evening in particular. Several times we nearly forgot; once we were twenty minutes late (WIND-MILL), and I fear that our thoughts throughout were more on an approaching trip abroad than on paranormal cognition. I accordingly half expected that the experiment would be a failure, or at least comparatively so; and if success or failure depended on the conscious mind of the experimenter, I do not see how it could have failed to be affected, even though I was no less anxious than before that it should succeed. But, as will be seen, it was at least as successful as the others, at any rate in the sense that it played its full share in producing the high significant result obtained from the work as a whole.

Probably the most notable features by inspection were: A fine crop of scissors (nine in the whole experiment, and none previously); five balances (including scales and steelyards) also unprecedented; five boots (only two or three slippers before this); eleven shells (counting snails) as compared with two in earlier experiments; and a relatively great number of trees, boats and jugs (ewers), though all these had been fairly common.

These features will, of course, be incorporated in the calculation of the main over-all results and there is nothing to be added by gloating over them here, striking as they were when first observed. On the other hand, there was another feature, of a different kind, which appears to me of such outstanding general interest as to deserve a section to itself.

7. Special Note on Dr. Rhine's Group: In due course I hope to make a complete comparative study of the various groups who have engaged in these experiments, but I have not been able to undertake this as yet. I feel it incumbent on me, however, to put on record here the extremely striking results obtained by the group of twelve percipients from Duke University.

By inspection this group is outstandingly good; in fact I have no doubt at all that it will be found to be easily the best of all which have taken part. At the moment I can say definitely:

1. When tested in a 2x2 table against all percipients who did not take part in experiment IVA, *viz*, the percipients of experiments I, II, III, and IVB, it yields an *intrinsically significant* positive result, for we have

	Duke Group	Others	Totals
Hits on Originals of IVA	18	94	112
Hits on other Originals	60	729	789
Totals	78	823	901

whence χ^2 is found to be 7.853 with P less than .01.

2. When tested against all other percipients who did take part in experiment IVA, it yields a *significant-ly better* result, for we here have

	Duke Group	Others	Totals
Hits on Originals of IVA	18	40	58
Hits on other Originals	60	268	328
Totals	78	308	386

giving χ^2 4.204 with P less than .05.

(N.B. In so far as the above calculations are not self-explanatory, the reader must wait for full enlightenment till we have discussed Methods of Assessment and the calculation of results; they may, however, be taken as correct pending explanation of these matters.)

3. It is better, though not significantly better ($P \sim .09$) than the other percipients of experiment IVA at discriminating between the Part A and Part B originals.

In view of the great amount and intensity of criticism to which Dr. Rhine's work has been subjected, I think it is only fair that these very remarkable facts should be noted at the earliest practicable moment. They in no way invalidate, of course, the considerable legitimate criticisms which might be, but usually have not been, brought against the work in question; still less do they guarantee that all the results reported by Dr. Rhine and his colleagues or followers, or even any particular examples thereof, are veridical. But they do go a very long way towards substantiating Dr. Rhine's main contentions in a general fashion. It will be seen later that the results of my experiments taken as a whole are very significantly positive; and in so far as they may be accepted they indicate the occurrence of a 'paranormal' mode of cognition. But this might perfectly well be true

and Dr. Rhine's results still due to a mixture of carelessness, practical joking and so forth. On the other hand, the fact that members of the Duke Group did so well in this experiment suggests to the point of demonstration that they possess in good measure the ability revealed by the investigation as a whole; and if this is so it seems to provide circumstantial evidence for supposing that they may also possess the presumably closely allied abilities claimed for some of them by Dr. Rhine.

SUMMARY: The important points to note in this section are not the trivial details, which I record only for the sake of completeness, but the following:

1. In all experiments, the possibility of any percipient seeing any original, or of his hearing anything, such as pencil scratchings, involuntary whisperings or the like, was completely excluded.
2. In no experiment was it possible for any percipient to infer or forecast the nature of any original by any process of rational inference.
3. The process of random selection which assures the second point also assures the elimination of coincidental thinking prompted by contemporary events.
4. If anyone can show that, despite the precautions described, there was, in fact, scope for the operation of sensory clues, rational inference, or coincidental thinking, to any appreciable extent, the experiments are invalidated. *But, if not, then not;* and, in such circumstances, critics must transfer their attentions either to the Method of Assessment employed or to the statistical treatment.

Section III

METHODS OF ASSESSMENT

1. General: From various comments I have made in the preceding pages it will have been gathered that I had substantially no difficulty in obtaining what appeared to be genuine positive results when judged by inspection and assessed by the persuasive if intangible criteria of common sense. Fortunately or unfortunately, however, the findings of 'common sense', like the sayings of the immortal parrot, are not evidence; and I found it by no means so easy to devise a satisfactory method of assessment which should not only be completely free from any suspicion of bias and proof against all manner of wish-thinking, but also sufficiently sensitive and capable of doing justice to the material.

Somewhat correspondingly, I find that whereas most people with whom I have discussed the matter have little difficulty in understanding the necessity for the precautions taken against leakage, and their efficacy, they tend to go astray so soon as questions of judging and scoring arise. I shall deal faithfully with some of the commoner pitfalls and fallacies at a later stage, but it will be desirable here as a preliminary to what follows in this section, to consider the problem in general terms.

The difficulty arises both technically and in the mind of enquirers, with the transition from the restricted to the free type of material. Almost anyone can understand that there is just one chance in fifty-two of guessing correctly a playing card drawn from a normal shuffled pack; or that, if a pack of twenty-five Zener cards contains five specimens of each of five varieties, the chance of any guess being right is one fifth, and five the most probable number of successes in twenty-five trials. And most people can at least grasp the idea that it is possible to calculate the probability of any given number of successes arising, in these circumstances, by 'chance alone' or 'pure guesswork', even though they may be incapable of performing or even following the requisite calculations themselves. But so soon as we begin to deal with drawings, which may be of anything under the sun, where there are no such convenient antecedent

probabilities to guide us, and where opinions may very well differ widely as to what constitutes a 'success', they seem to imagine that the whole business must necessarily degenerate into a mere matter of opinion to which the application of precise methods is impossible.

This is very far from being the case, and the various methods described below, though of widely differing practical value, are all as logically impeccable as any that have been used in the assessment of restricted material. Indeed, the boot is, if anything, rather on the other leg; for the antecedent probabilities mentioned above are not, as is commonly but erroneously supposed, god-given *a priori* certainties: they are hypotheses based on assumptions (usually justifiable in the circumstances) to the effect that the packs of cards, etc., which are used approximate very closely to ideal packs and suffer negligibly from defects, such as differences of stickiness, which might cause their behaviour to diverge from the ideal pattern. I think I am right in saying that all cases where these facts do not, in principle, need to be taken into account are special cases - e.g. that of guessing all the cards in a pack of known composition.

In practice, however, scoring situations are pretty sharply divided into two types, in the first of which there can be no doubt as to what constitutes a success, while in the second there may be. In the first, the scoring is truly objective, so that it does not matter whether the judge or scorer 'knows the answer', provided he is accurate and honest; whereas in the second it is imperative, as a rule, that he should not know, lest his personal prepossessions should bias the outcome. All experiments with playing cards or Zener cards fall into the first class, and all experiments with free drawings into the second. There can, for example, be no doubt that a success has been scored if the Queen of Hearts is guessed when the Queen of Hearts is drawn, or that a guess is a failure if the percipient says "square" when the card in question showed a star. Even if we take partial successes into account, giving so many points for rightness of suit, colour, number, and so forth, the procedure remains perfectly objective, for there can be no reasonable doubt that a percipient who guesses Nine of Diamonds is thinking of a red card, so that there can be no question, once the system of scoring has been settled, as to how many marks should be given in any particular case.

But the moment we turn to unrestricted material, such as drawings where the percipient is not asked to say *which* of a number of known things the object to be cognised is, but *what* it is, the situation becomes very different, and is still more so if the percipient is required to draw what he thinks it is instead of stating its nature in a word or two. For in these circumstances the question of whether a success has been scored or not may easily be a matter of opinion or even of acute controversy. Is it to be counted a success, for example, if the original is an arch bridge and the drawing is a suspension bridge; or if the original is a monkey and the drawing a gorilla, or a lemur or a baboon? Sceptical purists might say No, wishful enthusiasts would say Yes; and the only thing certain is that, regardless of which might be the more reasonable view, no assured conclusion could be reached.

It is accordingly clear that, unless our efforts are to be stultified before we start, we must base our enquiry on a somewhat wider conception than that of the simple 'right or wrong' antithesis applicable to restricted material. This, of course, most emphatically does *not* mean that we are to allow any one-sided relaxation of standards such as would allow the wishful to claim mediocre resemblances as successes and to discard others as failures, just as it suited their purpose. On the other hand, it would evidently be absurd to demand an exact point to point correspondence between original and drawing before conceding a 'hit', for this would merely ensure that we should never record a 'hit' at all, either in the right place or the wrong.

Clearly, our proper plan will be to allow whatever latitude we see fit as regards what shall constitute a resemblance, or 'hit', but to arrange our procedure in such a manner that this will cut equally both ways and be as likely, if chance alone is operative, to increase the number of hits in the wrong places as in the right. The first is necessary in order to secure any material to work on; the second is not only necessary but also *sufficient* to guard against warping the outcome in one direction or the other.

It is important to get this last point clear, for one of the commonest delusions in the subject is that lowering the standard of acceptance (i.e. the closeness of resemblance demanded before a hit is scored) must *ipso facto* favour the production of spurious positive results. I do not know what the origin of this belief

may be, but I suspect that it is based partly on a false analogy with experiments using restricted material, and partly on a failure to realise that hits in the wrong place as well as in the right are, and must be, taken into account in the assessment of any material of the type we are discussing. Obviously, for example, in an experiment with playing cards, we should soon generate spurious results if we counted a guess of knave or queen as a full success when the card drawn was a king, on the ground that there were court cards and 'very like' the king. No sane person would do this, of course, without taking into account the fact that the antecedent probability of success has been materially increased; though it would be perfectly legitimate to do so if appropriate adjustment were made. I suspect that some people cannot get away from the idea that analogous antecedent probabilities must be used in assessing free material and that lowering the standard must be equivalent to the kind of thing just indicated without the possibility of applying the corresponding correction; but this, as will be seen, is not the case. Alternatively or additionally, persons unfamiliar with this class of work are apt to think only of one side of the situation and to forget that although lowering the standard is likely to increase the number of hits in the right places, it is likely correspondingly to increase the number in the wrong, just as raising it will diminish the number in the wrong as well as in the right.

Perhaps the point at issue may best be clarified by reflecting that the relation between the original and the drawing, or the card and the guess, etc., is always of a two-fold character; there is not only a relation of likeness but also a relation of position. That is to say, we not only demand of a 'successful' guess or drawing that it shall be 'like' some card or original used by the experimenter; we require also that it shall occur in a position related in some definite way to that in which its prototype occurs. This may, at first sight, appear too vague; actually it is no more than stating our second requirement in accurate if general terms. Usually, of course, we require that the positional relationship shall be one of identity; that is to say, just as we demand that the percipient shall guess square and not circle when a square is concerned, so we insist that he shall do so on the tenth occasion, say if that is the occasion on which a square is drawn, and not on the eleventh or the ninth or any other. But there is no kind of theoretical necessity for doing this, as regards either part of the relationship. In practice we usually do it

because it seems to us more likely on common sense grounds (not always an infallible guide) that if paranormal cognition occurs at all it is most likely to do so in a certain particular form, namely that which will produce a strong positive relation of likeness and an identity of ordinal position between the 'prototype' (original, card, etc.) and the 'reproduction' (drawing, guess, etc.). But this is based on a judgement, not necessarily correct, as to what is likely to happen, and we are perfectly entitled to modify either part of the relationship, or both, in any way that seems good to us, as by enquiring whether there is a significant tendency to guess black cards when red are drawn, and *vice versa*, or whether correct guesses tend to occur seven places later than, or within a range of three places before or after, the prototype. I am not suggesting that it would often be wise or profitable to go out of our way to hunt for peculiar 'configurations', to adopt a convenient term, such as these or others more fantastic; but it would, in principle, be perfectly legitimate to investigate the frequency of their occurrence, if there seemed any object in doing so, provided always that we make the necessary allowances and apply the appropriate safeguards in estimating the probability of their happening by chance.

My purpose in the preceding remarks was to introduce and emphasise the notion that there are more places (position) than one in which a 'hit' may occur. Colloquially speaking, there are 'right' places and 'wrong' places; the right places are those specified by the positional part of the compound relationship defining the configuration we are considering, such as 'the same ordinal position', or 'within the same experiment', and all other places are wrong. Once this idea is grasped, it is easy to see that what interests us in the general case is not the absolute number of hits scored, which may vary greatly according to the intrinsic popularity, so to say, of the prototype and the strictness or otherwise of the scoring, but the question of whether relatively²⁰ more hits are scored in the 'right' places than in the 'wrong'. Further, it should be clear that, provided the process of scoring or marking is applied

20. The word 'relatively' is necessary because, as a rule, there will be more wrong places available than right; so that if chance only is at work, the absolute number of hits in the wrong places will be greater than that in the right. What we want to know is whether the right places get more than their chance indicated share of the hits recorded.

impartially, that is to say without any systematic bias in one sense or the other, any increase or diminution in the total number of hits scored, as produced by lowering or raising the standard, will correspondingly inflate or diminish the number of hits in the right and wrong places indiscriminately.²¹

In dealing with restricted material there is, of course, no difficulty about ensuring impartiality, because there can be no difference of opinion about whether a hit has been scored or not; but as soon as we begin to consider drawings, where acute and perfectly legitimate differences may easily arise, the position is quite otherwise. Cases about which there can really be no two opinions are the exception rather than the rule, and it would be no more than a time-wasting engenderment of controversy if we were to conduct our process of assessment in such a way that any position results which might emerge, could legitimately be attributed to bias on the part of the judge.

There is only one way of ensuring absolute impartiality in these circumstances, and that is by arranging that the judge is wholly ignorant of which places are 'right' and which are 'wrong' so that he cannot possibly favour the one category at the expense of the other in his allocation of hits. Under these conditions, the most that wickedness or stupidity can achieve is a voluntary or involuntary sabotaging of the experiment, either by recording so many hits as to swamp any real effects there may be, or so few as to prevent its emergence.

For this reason, no less care has been taken throughout this work to prevent leakage of sensory and inferential clues to the percipients.

In the following sections I describe three methods of assessment which have been tried. Only the third has proved satisfactory, but I have thought it worth while to give some account of the other two, partly for the sake of completeness, and partly in order to emphasise the importance of not deciding too rigidly in advance the form which a real effect, if any, 'must' take.

²¹ In what follows, I shall frequently use the term 'winners' for hits in the 'right' places, and 'losers' for hits in the 'wrong', the words 'right' and 'wrong' having the meaning given above; I shall reserve the word 'success' for the particular case in which the ordinal position of the reproduction in the series considered is the same as that of its prototype.

The third method is neither more nor less theoretically valid than the first, but it revealed the significance of previously unsuspected facts indicated by inspection of the material; the first was just as good a way of tackling the job in the light of antecedent ignorance, but since the effect it was designed to bring out was not detectably present, it led to null results.

2. The Method of Forced Matching: The matching technique discussed below is not new to psychological practice, for it has been fairly freely used in cases such as the connection of character with hand-writing where ordinary quantitative methods cannot readily be applied. The principles involved are likely to be more easily grasped by some readers if we begin by considering an illustrative example of this kind.

We will suppose that we want to know whether a graphologist can form, from specimens of handwriting, estimates of the writers' characters which cannot plausibly be discounted as no more than lucky shots; and we will suppose that we are rightly anxious to eliminate any bias due to our own preconceptions as to the possibility of doing so. We select, say, ten subjects, A, B, C . . . I, J, for experiment and obtain from each a specimen of handwriting which we submit to the graphologist, asking him to make out the best character sketch he can of each person concerned and to give us the ten sketches shuffled up so that we cannot tell which is meant for which person but so numbered that they can later be identified. At our end, we either rely on our personal knowledge of Messrs A, B, C, etc, or we induce some competent observer acquainted with them all to prepare ten other character sketches based on ordinary experience; these may either be randomised or not, but we will suppose they are not. When the graphologist's versions arrive, we will compare them with those of the competent observer or with our own knowledge, and try to pair the two sets off - graphologist's A against observer's A, graphologist's B against observer's B, and so forth; but of course, as I have just said, we do not know which of the graphologist's sketches is meant for A and which for B, so that we are forced to rely exclusively on the resemblances, if any, between his version and that of the other observer. When we have paired off the two sets to the best of our ability, we ask the graphologist for the key to his numbering, and thus ascertain how many correct pairings we have made. We might find this sort of thing for example:

Graphologist's estimate of:- J C H D E A G B I F
is paired with

Observer's estimate of:- A B C D E F G H I J

Here there are four correct pairings, namely those of D, E, G and I, which is a good deal more impressive than it looks, for we should obtain four or more coincidences only once in fifty such trials, on the average, if chance alone were operative - if, for example, we had simply drawn the two sets of sketches at random out of two hats.

The point to grasp here is that, if character and handwriting (strictly speaking as estimated by the particular observer and graphologist concerned) have no systematic connection with each other, and the graphologist has no other source of information than the specimens submitted to him, then the matching procedure is precisely equivalent to the hat drawing ritual so far as the chances and expectations of finding coincidences are concerned. For, *ex hypothesi*, there is no more than a random relation between the graphologist's only guide and the characters he is trying to assess; and even if the observer's estimates were perfect, the relation of the graphologist's versions to these could still be no more than random.

It is very important to note here that we are not concerned in a situation of this type with the question of how good the graphologist's estimates are, in any absolute sense, but only with whether they are good enough to enable us to identify each as *more* like some other one than it is like the rest.

Now let us replace the known estimates of character, as made by the observer or ourselves, by the originals in one of our experiments, and the graphologist's efforts by the drawings of one of the percipients suitably shuffled and code numbered so that we do not know which is which. Then the situation as regards matching the drawings to the originals is formally the same as that of matching the graphologist's character sketches to the observer's estimates. Most people seem to have no difficulty in understanding the graphological parallel, but many come to grief when they try to think of the drawing-original situation. This appears to be mainly due to an insistent confusion between the probability of the judge correctly matching the n th drawing to the n th original by chance alone, when he has nothing but the intrinsic resemblance to rely on, and the probability of the percipient drawing an 'X', or an 'X-like' object, on

the same occasion as an 'X' is represented by the original.

However, since the method is well established, and I am not relying on its results for evidential purposes here, there is no need to spend more time on demonstrating its validity.

The procedure itself is simple enough. The judge (in these cases usually H.S.C. or W.W.C.), with the originals of the experiment concerned before him and knowing, as a rule, the order in which they were used, receives the drawings of each percipient arranged in a random order by some third party (*vide supra*) and code-numbered for subsequent identification. He accordingly has no external clue to guide him in deciding which drawing was intended for which original and must rely solely on such intrinsic resemblances as he can detect. His task, of course, in respect of every percipient, is to assign each of the (normally) ten drawings to whichever of the originals he considers it most closely resembles, or is least unlike. Sometimes this is easy enough, as with the hands and cows of experiment I, which could be assigned to hand and buffalo without hesitation, or with the tops and guns of experiment II and scissors and balances of IVA; but more often it is necessary to look for some more or less recondite similarity of form, or for associations of greater or less remoteness, in order to come to a decision; for example, in experiment I we frequently assigned drawings of boats to the original anchor, on the grounds that both were 'nautical', or vases, etc., to bottle, as being 'containers'. The judge is obliged, it will be understood, to assign every drawing to some original (hence the name 'Forced' Matching) and it not infrequently happens that the utmost ingenuity fails to discern any plausible resemblance at all, so that one or more drawings are finally placed by elimination or simply at random. After all the assignments have been made, on one basis or another, and duly noted, the occasion on which each drawing was actually produced is ascertained by reference to the key held by the third party aforesaid, and the number of correct matchings made in the case of each percipient is counted. It is then only a matter of some not very advanced mathematics to determine the probability of the observed number in each case being due to chance alone. I need not inflict these on the reader here, but the following table, for which I am indebted to Dr. Thouless, shows the probabilities of making the various possible numbers of correct matchings for series of ten originals and ten drawings,

such
R gi
O to
ber
R a
of
shot
is

nat
to
it
mul
tha
to
for
and
equ
the

exp
fro
are
val
The
are

such as we are concerned with here. The column headed R gives the number correct, which may have any value from 0 to 10 inclusive, except 9. Under N is shown the number of ways in which 10 drawings can be arranged so that R are rightly matched²²; column P gives the probability of getting exactly R right by chance alone, while P' shows that of obtaining R or more right by chance, which is what interests us here.

Table II

R	N	P	P'
0	1,334,961	.367,9	1.0
1	1,334,960	.367,9	.632,1
2	667,485	.183,9	.264,2
3	222,480	.061,31	.080,3
4	55,650	.015,34	.019,0
5	11,088	.003,056	.003,66
6	1,890	.000,520,8	.000,600
7	240	.000,066,1	.000,079
8	45	.000,012,4	.000,013
9	0	0.00	
10	1	.000,000,276	.000,000,3
Total:	3,628,800		

These figures apply to any single set of ten matchings; when, as in any of these experiments, we wish to combine the results from a plurality of percipients, it is more convenient to apply the Stevens matching formula (Cf. p. 57 below), from which it is easily seen that the expected number of correct matchings is equal to the number of percipients, with variance the same; for in each case the expectation is 1 with variance 1; and the expectation and variance for the whole group are equal respectively to the sum of the expectations and the sum of the variances for each constituent.

The actual results of matching in the first three experiments in this way are summarised in Appendix II, from which it will be seen that, when all the results are taken together, no conclusion of serious evidential value in favour of a cognitive effect can be drawn. There are, however, one or two points of interest which are worth mentioning.

22. These necessarily add up to 3,628,800, or 'factorial ten', which is the total number of different ways in which ten things can be arranged.

Chief of these is the very remarkable score of no fewer than six correct matchings obtained with one of the percipients of the first experiment. Reference to the table just given will show that such a result, or better, could be expected, on the average, only about six times in 10,000 such trials, or once in about 1,600 attempts. This, of course, is highly significant, and the probability of its occurring by chance even as the best among the sixty-eight sets of the first three experiments is no more than 1 in 25, which would itself be tolerably significant in normal contexts.

But the attendant circumstances render this result even more remarkable than it appears at first sight. The percipient was the fourth whose drawings were matched by H.S.C. and myself, and it so happened that we completed a batch of ten and ascertained the number of matchings made correctly before going on to deal with the remaining sets. In this first batch of ten, we scored as many as twenty successes, which is an excess of ten over expectation with standard error 3.162. Such an excess or greater would occur only once in about 660 such groups of ten by chance alone, and no more than once in almost exactly 100 times as the best of 6.8 such groups (i.e., in the whole of the material so far dealt with by matching). It is quite legitimate to single out this group for special consideration, because it is, so to say, isolated from the rest by considerations other than its high score: it was the first we matched; we stopped when we had done it; and we did not resume the work till after we were aware of the success of our efforts; thus the group is sharply distinguished from all others on both chronological and psychological grounds. The success achieved *may*, of course, have been no more than a fluke, either in its entirety or as regards the six-success set only, and if anyone wishes to write it off as such on the ground that we failed to repeat it, I cannot reasonably object; but it will be his loss, not mine. I might have felt constrained to do so myself if the Method of Palpable Hits had not yielded the very much more significant results reported below. As it is, I think it much more probable that it is genuine, and represents an unrepeated (though I hope not unrepeatable) display of what I can only call "insight" by H.S.C., who was almost wholly responsible for the matching at this stage. Certainly it was not due to these ten percipients having drawn things more clearly like the originals than others did; in particular, the hero of the six successes produced no unmistakable resemblance at all, and H.S.C.'s assignments in this case appeared to me (I must confess)

to border on the far-fetched and unconvincing. This, however, does not alter the facts or the probabilities, which clearly suggest that the achievement was more in the nature of a *tour de force* than a stroke of luck. If this is so, the implications are of very great psychological interest from two points of view: in the first place it is suggested that a genuine cognitive process may be subject to such distortion as leads to expression in almost unrecognisable form; in the second, it would appear that to detect the relation between drawing and original in such cases calls for something more, or other, than rational perspicacity can provide; presumably it involves a kind of intuitional process at the same mental level as that at which the distortion or transformation took place.

As a matter of fact, we shall find evidence for supposing (p. 70) that drawings having no more than a very feeble resemblance to their originals are yet to some extent determined by them, which tends to support this view; and it is greatly to be hoped that H.S.C. may succeed in recapturing her initial virtuosity, the loss of which (if real) is probably accounted for by the feeling that the success achieved settled the question at issue beyond need of further effort.

The other point that requires mentioning is the remarkable tendency we noticed for the best and most convincing resemblances to occur on the wrong occasions. We made it a practice to grade our matchings as α , β or γ according to the degree of confidence we felt in them, which very approximately corresponded to the degree of resemblance discernible, and the table below shows the numbers of α 's given in each of the first three experiments and how many of them were displaced early, late, or not at all.

Table III

Displacement

	Early	Zero	Late	Total
Experiment I	9	2	28	39
Experiment II	11	3	17	31
Experiment III	9	1	9	19
Total	29	6	54	89

We should of course expect about nine zero displacements out of eighty-nine awards if chance only were operative, and the difference is not significant. On the other hand

the excess of Late over Early displacements is quite definitely so, both for the first experiment ($P < .01$) and for the Totals ($P < .02$). This might be due, on the chance-only hypothesis, to the more popular originals (i.e., those more commonly drawn under truly random conditions) happening to have come early in the series. I did a considerable amount of work, by control matching and otherwise, to investigate this possibility; but concluded that though the effect was probably a real one, it was to some extent due to this cause, while the data available were insufficient to settle the matter decisively. But we shall be dealing with the whole question of Displacement, which is extremely important, at a later stage and by rigid methods, so that I need not go into details of these early explorations here, beyond remarking that these observations went far to support the belief engendered by the success of the first batch of matchings that something other than chance was at work.

For similar reasons I need not describe here the investigations I undertook to test sundry hypotheses, of varying degrees of far-fetchedness, advanced by assorted critics to account for our initial success on normal grounds; for example, the suggestion that percipients might tend to do their drawings later and later as the experiment progressed, and that we might have been unwittingly guided in our assignments by the time data recorded on the sheets. All these yielded null results.

On the whole, I think there is very little to be said for the Method of Forced Matching for this purpose, or probably for any other. Apart from the fact that it is liable to be completely wrecked by the phenomenon of displacement, it does not seem to yield any information that would not be given by a suitable system of marking, while it automatically precludes the possibility of giving recognition to the influence of more than one original on the same drawing, and it is hopeless for dealing with multiple or composite drawings. The chief advantage gained is that the assessment of each experiment, and indeed of each percipient, is self-contained, and does not require reference to any other drawings; and it would be of outstanding value for cases, if ever they occur, in which very faint resemblances, such as would probably be ignored in any ordinary system of marking, preponderantly appear on the correct occasions.

3. The Method of Decimal Scoring: The results of the first experiment suggested, and those of the second and third confirmed, that the Method of Forced Matching was not likely to be successful.

This was not because the drawings produced by the percipients bore no discernible resemblance to the originals at which the set as a whole was aimed, or resemblances so infrequent or so feeble as to be of insignificant importance: on the contrary, so far as common sense inspection and estimation of probabilities could tell, they seemed to be due to chance alone - for example, by the appearance of hands and cows in experiment I and of shootings and tops in experiment II. The trouble, as already indicated was that the most successful reproductions tended to come in the wrong place, so that the hand shown as an original on Saturday, say, would appear as a percipient's drawing on Sunday or Monday instead of on its proper day. Since the whole Method of Matching depends on the assumption that there will be preponderantly a one to one correspondence of occasion between an original and the drawing most closely resembling it, it is pretty well bound to be defeated by displacements of this kind; at the very least, its utility is likely to be seriously impaired. On the other hand, it is far from unreasonable to suppose that an impression subconsciously received might remain latent for some little time until either an internal process of gestation, or the incidence of some extraneous stimulus, brought it into the conscious field; alternatively, a process of gestation in the mind of the agent might be necessary prior to transmission, if the cognition were dependent on telepathic factors.

I accordingly sought for a method which should enable us to answer the broader question of whether the sets of drawings produced by the percipients, taking each set as a whole, significantly tended to resemble the series of originals at which they were aimed (i.e., the series of originals used in the experiment in which those percipients were concerned) more closely than they resembled series of originals at which they were not aimed. The idea, in other words, was to match each set of drawings, as a whole, against two or more series of originals, as wholes (one of them being that at which the drawings were aimed) instead of matching individual drawings against individual originals.

There is, of course, no theoretical reason why this should not be done in the literal manner suggested by the works used above. We might, that is to say, present the judge with the originals of perhaps half a dozen experiments, divided into the six series in which they were actually used, together with all the sets of drawings produced in those experiments, suitably

randomised and coded, and ask him to assign each set to the series which, on the whole, he thinks it most closely resembles, working by inspection and quality of judgement alone. But the practical difficulties of such a procedure are great, as was discovered when Prof. Broad kindly attempted to apply this plan to six sets of drawings and originals produced in the course of the Private or Individual experiments. Apart from the difficulty of bearing a large number of originals in mind at a time, there is that presented by the conflicting claims of drawings in the same set, of which some may resemble an original in one series, while others resemble originals in other series; it may become necessary to decide whether the strong resemblance of a certain drawing to originals in series A should or should not outweigh the fainter resemblances of two or more other drawings to originals in series B. Considerations of this kind necessitated the introduction of some quantitative method of expressing estimated degrees of resemblance, as opposed to making a purely qualitative decision to the effect that one drawing, or set, was more like one original, or series, than the others with which it was compared.

The plan I thought most promising was as follows: Judges were to be given a collection of shuffled and coded sets of drawings taken from two (or possibly more) experiments together with the originals, also shuffled, belonging to those experiments, and were to be asked to assign to each drawing as many points, from 0 to 10²³ inclusive, as they thought it deserved for its resemblance to any original or originals. In cases where two or more objects appeared in the same drawing the marks which would have been given to any of them if shown singly were to be reduced to an extent corresponding to the relative importance of the object in question in the drawing in which it appeared: To quote from the instructions "If the relevant object shares the drawing more or less equally with one, two, three, etc., other objects of apparently equal importance, it should be given a half, a third, a quarter, etc., of the points it would have been given if it had appeared alone". This regulation was probably unwise, and was discarded in the Method of Palpable Hits finally adopted; for further reflection suggested that the important point is not likely to be that of whether the idea depicted in the original is the only impression received by or present to the mind of the percipient, but of whether the drawing provides good evidence for supposing that the idea was prominently

23. Hence the name given to the method.

pres
lott
cern
whic
the
grea
we m
own

set
of e

Not:
thi
whi
bet
in
res
aim
II,
ces
ser
if
the
ser
who

pro
had
mar
fra
exe
ly
ner
osy
the

present in his mind at all. If we add up the points allotted in this way to the originals of each series concerned, for each set so scored, we shall obtain totals which will, in general, be unequal; and if the total for the series at which any particular set was aimed is greater than for that or those at which it was not aimed, we may say that the set has been correctly matched to its own series.

Thus, to take an imaginary example, a particular set from experiment I, say, scored against the originals of experiments I and II, might yield results like this:

Drawing No.	Original	Points
1.	Parnassus	3
2.	Hand	10
3.	Horse	5
4.	Buffalo	8
5.	Net	6
6.	. .	.
7.	Throne	7
8.	Bottle	10
9.	. .	.
10.	Bat	3

Noting that hand, buffalo, net, bottle and bat, scoring thirty-seven between them, are originals of experiment I, while Parnassus, horse, and throne, scoring only fifteen between them, belong to experiment II, we conclude that in the opinion of the judge the set, taken as a whole, resembles the originals of experiment I, at which it was aimed, more closely than it resembles those of experiment II, at which it was not aimed. It is clear that the process of allotting points to the originals of the various series concerned according to the degree of resemblance, if any, that each drawing of a set is judged to show to them, and then adding the scores so obtained by each series, is only a roundabout way of 'matching' sets as wholes, which is what we set out to do.

A rough and ready trial of the method yielded promising results; but exploration soon showed that it had serious disadvantages. The chief of these may be summarised by saying that it proved extremely difficult to frame instructions elastic enough to give room for the exercise of common sense and at the same time sufficiently precise to avoid leaving the result too much at the mercy of the individual caprices of the judge. Such idiosyncracies could not, of course, systematically falsify the outcome, nor would they be likely altogether to

obscure a genuine positive result in the long run; but by introducing factors other than the straightforward resemblance or close association of drawing to original, they in effect increase the variance, and may thus inordinately lengthen the 'run' necessary to secure a definite answer to the questions studied. As an example of the kind of thing I have in mind, one potential judge of whom I had great hopes, insisted, in flagrant violation of the instructions, on giving the maximum of ten marks to hand for a drawing expressly described by the percipient as representing a fingernail; two others seemed to think that they were engaged in a puzzle picture competition, on the lines of 'In this drawing of a bird's nest find five hidden pirates'; while in another instance it was seriously contended that a drawing of a girl on a surfboard should be given points for incubator, on the ground that the words 'poule' in French and 'chicken' in American are colloquially used to refer to the female young of the human species. This kind of thing might be of interest in pursuing the more abstruse ramifications of the subject, but so long as we are in doubt as to whether any paranormal cognition takes place at all it can only obscure the issue.

Another, and sometimes more serious, trouble is that if the objects depicted in one series of originals are appreciably more popular, in the sense of being more frequently drawn under chance conditions, than those of the other series they will tend to attract, as it were, an undue proportion of points, so that only sets resembling the other series strongly as to outweigh this difference will escape the attraction and have a chance of being allocated to the less popular series. This effect is very marked in the case of experiment IVB, which contains some highly popular originals, notably tree and boat; reference to Appendix III will show how it ruins my scoring of eighty sets from experiment IV.

Finally, if the method is applied, as I applied it, to drawings and originals taken from only two experiments at a time, it is liable to be extremely insensitive, because so high a proportion as fifty per cent correct assignments might be expected to result from guesswork alone; while, on the other hand, the task of attempting to decide what fraction of a maximum of ten points should be allotted to which, or perhaps each of several, of more than twenty originals is too difficult to expect any ordinary person to attempt with success.

In view of these handicaps it is hardly surprising that application of the method was next door to being a failure. Mr. and Mrs. Oliver Gatty, to whom I am very much indebted in this matter, working on the suitably grouped, randomised and camouflaged drawings of the first three experiments, obtained a result slightly beyond the .07 level of significance, which looked reasonably promising, though hardly a rich reward for the examination of over 650 drawings. But even this was reduced to no better than .12 when my own scoring of eighty sets from experiment IV, A and B, were included, and was only raised to a level a little better than .03 by the incorporation of the sixteen sets of the Individual experiments. Details are given in Appendix III.

These results were disappointing, especially in view of the amount of labour involved, and by contrast with the degree of success apparently visible to the naked eye. It seemed clear to me that the method was at once too delicate and too cumbersome, too elaborate and too insensitive, to be of permanent utility, and that something considerably better would have to be devised.

4. The Method of Palpable Hits.

General: This, the third and (so far) last method adopted was designed to overcome the difficulty of Displacement which stultified the first, while avoiding the defects of vagueness and insensitivity which, so far as I could judge, were chiefly responsible for the poor results obtained from the second.

It must be realised that all the time the collection of material and attempts at assessment described above were going on I was continually impressed by the plain, commonsense evidence that the experiments seemed to thrust upon my notice. For example; The originals of the first experiment contained a hand, and the thirty-seven sets of the first experiment showed at least six obvious drawings of hands, apart from several with plausible claims as partial successes; in contrast, the thirty-one sets of experiments II and III produced no more than one glove and one finger-nail, which might perhaps count as one success for hand between them. For buffalo, I found five obvious cows in the first experiment, but only two in the second and third together. Similarly, I found five unmistakable guns, cannon, etc., for shooting among the drawings of the twenty sets of experiment II, and only one very doubtful candidate in the forty-eight sets of experiments I and III. Again, when I

began to examine the drawings of experiment IV, I found a very satisfactory number of scissors and balances, neither of which had appeared at all in the drawings of the first three experiments, together with an apparently undue proportion of ewers (or similar vessels) and isolated trees. It seemed to me almost incredible, on common sense grounds, that these effects were solely due either to chance or to the wishful selectivity of my own mind, or that, if they were genuine, it should be impracticable to demonstrate them significantly by some tolerably simple method.

I accordingly determined to abandon, for the time being at any rate, my ideas on giving full scope to possible associations, distortions of form, symbolisms, etc., and to concentrate on the simplest and most straight-forward method of assessment possible. I similarly resolved to discard the niceties of gradation afforded by giving anything from 0 to 10 points to a resemblance, together with the system of scoring in proportion to the importance of the object in the drawing, and to base the assessment as nearly as possible on the simple principle "Do this drawing and that original plainly and unmistakably portray the same thing? If they do, give one mark; if they do not, give nothing." In other words, only 'palpable hits' were to be counted.

This sounds simple enough on paper, but in practice it is not nearly so easy as might be supposed, and I found it quite impracticable (somewhat fortunately, as it proved) to induce others concerned to adopt and maintain the high standard I originally envisaged, and extremely difficult to adhere to it myself.

In the first place, I realised that it would be necessary, even before I started, to modify the basic rule, so as to cover composite drawings containing more than one object, by adding "or does the drawing provide plain and unmistakable evidence that the object portrayed in the original was prominently in the mind of the percipient when he made the drawing", or words to that effect; and the introduction of the word 'prominently' at once opens the door to a certain amount of difference of opinion and ambiguity; yet without this modification a number of drawings which common sense would indicate as clearly 'hits', such as horses in carts, or cows standing under trees, might be rejected by too literal a follower of instructions.

In the second place, even the words "the same object" cannot be wholly unambiguous; for example, a drawing of a child might be held to represent ancestor just as well as the old man I actually drew, for very many children have descendants and all ancestors were young once. In cases such as this it seems necessary to keep pretty closely to the actual drawing; whereas with an original like net (experiment I,8) it is more important to consider what was in the agent's mind (namely, illustrating a net) rather than all of what was drawn (namely, a man pulling a net with little fish in it out of the sea) for it would be obviously absurd, or so I think, to give a mark to every drawing which shows a man, a fish or the sea, without any indication of the net. On points such as these opinions may well differ, and it does not very much matter what we decide to do, provided our decision is not manifestly nonsensical and that we hold to it with reasonable consistency; indeed, failure to conform to even these conditions would not necessarily prove fatal. It will be understood, or so I hope, that no convention of a general kind can possibly produce a spurious positive result; this could only be done by forcing upon the marker some arbitrary specific convention based on a preliminary study of the material. For example, I might notice that there was a large number of keys drawn among the sets of experiment IV as was the case, and then suggest or insist that keys should be given a point for, let us say, castle; this, if followed would lead to a number of unjustifiable winning hits being recorded; but a general instruction such as 'clearly identifiable parts of objects should be treated as if the whole object were drawn' could do no harm; for there is evidently no reason why such a rule should increase the proportion of hits registered on originals in the series at which percipients were aiming (i.e., 'winning hits' or 'winners') compared with those on series at which they were not aiming ('losing hits' or 'losers'), and it is by the relative frequencies of these, not by the total number of hits of all kinds, that the issue must be judged..

On the other hand, it behooves us to frame our rules with intelligence, because, if we made them too lax, we shall tend to blur whatever pattern the data might otherwise reveal, while, if we make them too stringent, we shall risk having no pattern to blur. Thus, if we gave a mark to arrow for every drawing with a straight line in it ('straight as an arrow'), and one to shell for every drawing with a curly line in it, we should be likely so to over-weight these two originals with marks

as completely to swamp out the kind of thing we are looking for; while if we insisted on a drawing being an exact point to point counterpart of the original before giving it a mark, we should ensure a perfectly null result by being unable to award any marks at all.

Our business, clearly, is to steer between the two extremes by framing the simplest set of common-sense rules we can, and to leave the doubtful cases to the discretion of the marker aided by whatever suggestion of a general and non-tendentious character we can supply. I shall refer again shortly to the instructions and guidance actually used, but the underlying principle remained substantially that given above, namely: 'Stick to the obvious; avoid the recondite; give one mark for a palpable hit, nothing for a miss, and half a point if there is real doubt'. Thus three degrees of resemblance were recognised - hit, miss and doubtful - compared with the eleven permitted in the method of decimal scoring.

In the interests of sensitivity I also decided that all drawings must be marked against all originals, within any self-contained group of experiments²⁴, as opposed to working by pairs of series of originals as I did with the preceding method. This involved trying to bear fifty originals in mind at once, or else adopting the terribly tedious procedure of comparing each drawing with each original in turn; but on the credit side there was no longer any need to think about exact gradings or to look for associations and the like, so that it was not quite so formidable as it sounds, and the experience showed that it could be done without undue difficulty.

I have already emphasised the necessity for basing our enquiry on an empirical study of the frequency with which people do in fact draw the objects depicted in the originals, on occasions when these originals are not exposed, but a few remarks in amplification may not be out of place here. It is no use congratulating ourselves on the fact that hands and cows, guns and spinning tops, scissors and balances turn up where they should and not where they shouldn't, if at the same time, unremarked by us, other originals are scoring a profusion of hits in winning and losing positions indiscriminately. If our

24. I shall discuss elsewhere the question of what is to be done when we are dealing with isolated experiments, or small groups; one cannot go on marking all drawings against all originals indefinitely, nor can one be sure of always having a suitably sized group within which to work.

originals consisted preponderantly of objects which, like trees and boats, are very commonly drawn, it might quite well happen that a few lucky hits on the less popular originals, in the right place and not repeated elsewhere, would look extremely impressive to the expectant eye, whereas they would really be no more than some of those insignificant aberrations with which any large collection of random data is almost certain to be embellished. We have no right to pick out the plums, however conspicuous they may be, and to use them alone in a test of significance; though we may quite legitimately test them mentally, so to speak, in the light of our general experience and common sense, with a view to forming a qualitative judgement as to what is going on. For a statistical test we must take account of *all* hits, whether 'winners' or 'losers', within the group concerned, before we can decide on the likelihood of the observed proportion of winners being the result of chance alone.

Mathematical Treatment. The statistical treatment necessitated by the method does not, fortunately, involve any very recondite process; indeed, the first stage is a matter of no more than common sense arithmetic. We will suppose that an unbiassed judge has given us the number of hits scored by all the percipients of each experiment on the originals of that experiment and of all others in the group considered; that is to say, the number of hits scored by all drawings on all originals suitably subdivided under experiments. A certain number of these will be 'winners', namely those scored by the percipients of any particular experiment or the originals used in that experiment, and the remainder will be 'losers'. Our task is first to calculate, from the empirical data provided, how many winners we should expect, and then to apply a test of significance to show how likely it is that the observed number differs from the expected number as much as it actually does (or more) if this difference were the result of chance alone.

In dealing with the first of these stages we ignore 'misses' altogether; for these are of no more interest to us than would be calls of 'spade' or 'club' by a percipient trying to guess cards in a pack consisting (unknown to him) solely of hearts and diamonds. All we are interested in is the question of whether percipients tend to draw hands, horses, violins, scissors, trees, etc., significantly more often when these objects are represented among the originals used in the course of an experiment than when they are not, and the frequency with which they draw armadillos, razors, kite-balloons,

or other objects which have not been presented at all, can throw no light on the problem.²⁵

Now suppose, for the sake of illustration, that our judge has recorded a grand total of 1,000 hits by all drawings on all originals, and that these include twenty-five boats; and suppose also that 400 hits of one kind or another were made by the percipients aiming at the series of originals in which boat appeared. It is now no more than a simple proportion sum to ascertain how many of those 400 hits may be expected to be boats, if chance alone is operative. For we have just found that boats make up twenty-five thousandths of all hits, or 2.5%, and on the chance hypothesis there is no reason for supposing that this proportion will be greater or less in any particular sample we may happen to choose; so we should expect to find 2.5% of 400, i.e., ten boats, appearing among the drawings of the percipients in question. We may repeat the process for all the originals in turn, or, more simply, may write a_j for the total number of hits scored by the drawings of the j th experiment on all originals, b_j for the total number of hits scored on the originals of the j th experiment by all drawings, and N for the total number of hits, when $a_j b_j / N$ gives us the expected number of hits by the drawings of the j th experiment on the originals of that experiment; that is to say, the expected number of winners for the experiment considered. Then the total expectation of winners will be $S(a_j b_j) / N$, where S denotes the process of summation over all experiments of the group. If we find that the observed number of winners is considerably in excess of the expected number thus calculated, we shall have grounds for suspecting that the chance hypothesis is insufficient to account for the facts observed, and that the frequency with which drawings representing the various originals appear is not independent of whether those originals occurred in the series at which the percipients were aiming. But we cannot form an accurate judgement until we have determined the 'variance' of the expected number, that is to say, a quantity which serves as a measure of the extent to which the theoretically

25. Readers who prefer a more mathematical form of argument on this point may consult Mr. W. L. Stevens' paper on *Tests of Significance for Extra Sensory Perception Data*: Psychological Review, Vol. 46, No. 2, March, 1939. A few moments study will show that the situation in which we are interested is formally equivalent to that of card matching there discussed, and that to ignore the misses is merely to say that certain types of card in the target pack have zero frequency.

calculated expectation is likely to vary under chance conditions and accordingly affords a means of judging whether any actually observed deviation therefrom is likely to be due to chance alone.

This is not such a simple matter as that of determining the expectation, and the interested reader should refer to Mr. Stevens' paper, *Tests of Significance for Extra Sensory Perception Data*; Psychological Review, Vol. 46, No. 2, March, 1939, where it is shown that the value of the variance, σ^2 , is given by

$$\frac{1}{N^2(N-1)} \left\{ S^2(a_j b_j) + N^2 \cdot S(a_j b_j) - N \cdot S(a_j b_j)(a_j + b_j) \right\}$$

which is not, in practice, nearly so formidable to deal with as it appears. From this quantity we derive the Standard Error (σ) of the expectation, by taking the square root; dividing this into the observed excess of winners we obtain the value of the 'normal deviate', whence, by reference to the appropriate tables in the usual way, we find the probability of such an excess, or a greater, arising as a result of chance alone. An example will be shown worked in full when I come to discuss the outcome of the experiments as a whole. A discussion of the applicability of Mr. Stevens' method to the present problem, by Dr. Irwin and Mr. Gatty, together with a physical analogy due to Professor Broad, will be found in Appendix V.

Marking the Drawings. Meanwhile we must consider the very important matter of how the drawings were in fact marked, and the precautions taken to ensure that the data to be used in subsequent calculations were wholly or substantially unbiased.

To start with, all drawings were marked against all fifty originals by myself. This was done partly as a matter of exploration, in order to see whether the method seemed promising, and partly to find out at first hand what sort of difficulties or ambiguities were likely to be encountered. I did this work as carefully and conscientiously as I could; but, since I knew which drawings were aimed at which originals, the highly significant result I obtained cannot be accepted as reliable evidence. I do not think I erred in the direction of giving winning points where none were deserved - that, I found, was fairly easy to avoid; but it was only natural that, knowing a given batch of drawings to have been aimed at a particular series of originals, I should have other series less prominently in mind than this,

and should thus tend to miss a certain number of 'losers' which a wholly uninformed judge might find. On the other hand, since I had drawn thirty of the fifty originals myself, and had assisted H.S.C. in the preparation of the remainder, it is reasonable to suppose that I was in a better position than any outsider to know what the agent actually had in mind at the time, and thus whether any particular drawing successfully reproduced it; and this might well be of importance from certain points of view. Be this as it may, however, my own markings must evidently be ruled *hors concours*: they are given below as a matter of comparative interest only and no conclusions are based on them.

For the next, and all-important, stage I was so exceptionally fortunate as to secure the help of Mr. M. T. Hindson, the value of whose contribution to the work cannot be overestimated. Mr. Hindson not only had the leisure and critical interest necessary for undertaking the task of marking, but had pursued until recently the career of Bank Inspector - a vocation well calculated, one may suppose to promote just those qualities of objective judgement and indifference to tedium which are most necessary here. To say that I am very deeply indebted to him is seriously to understate the case; for without his aid the work might well have come virtually to a standstill for lack of a suitable judge. As it was, his marking, carried out under carefully planned and maintained conditions of ignorance, yielded a highly significant result of the highest importance.

The fifty originals were given to Mr. Hindson arranged in alphabetical order, that is to say *randomised* or *shuffled* from the point of view of deciding which belonged to which experiment; and I was careful to remove completely, by cutting off the tops of the sheets, such numbers or letters as had been used at earlier stages, which might possibly have identified certain originals as belonging to the same experiment.

Instructions were contained in a 'Guide to Scoring Hits', which seems of sufficient importance, in the circumstances, to be worth reproducing as Appendix IV. This 'Guide' was supplemented by a number of 'Notes on Originals', which was intended to draw attention to various points I had noticed in the course of my own marking. The idea, which Mr. Hindson adopted, was that the judge should go through the Notes first, before tackling the work of marking, and make up his mind what policy he would adopt with regard to each of the points

indicated, before actually meeting with them, so as to reduce the risk of varying his standard in the course of the work.

In preparing these notes, I was careful always to use either an interrogative or an imperative form, in drawing attention to the points raised, in order to avoid giving any suggestion as to how I thought they ought to be dealt with. Examples are:

"BOTTLE: This was based on the dictionary word VACUUM BOTTLE, which was rejected as too difficult. What will you give to carefes, medicine bottles, rubber hot-water bottles, jars, vases, etc.?"

"EWER: In the agent's mind the distinguishing features were the handle and the constricted neck. Should full or half hits be given for two-handled vases? For watering cans, teapots, etc., with handles but also spouts? For teacups, etc., with handles but no constriction? For saucepans, etc., with a different sort of handle? For vases, etc., with constriction but no handle?"

"FLAG: It will be necessary to distinguish this from Royal STANDARD. The flag drawn is a black flag with a strongly marked White Latin Cross. Consider black flags (without crosses), plain flags, striped flags, Union Jacks; Latin crosses without flags; other sorts of cross; also pennants, burgees, etc."

It will presumably be agreed that notes such as these, while they may serve a useful purpose in directing the attention of the judge to the kind of problems he will have to solve, contain no suggestion as to which solution he should adopt in any instance, and still less (if possible) as to which solution would be favourable to the experiment. The notes are given in full in Appendix IV.

I hope I need hardly add that, when I discussed the matter with Mr. Hindson, I was most scrupulously careful not to say anything that could possibly be interpreted as a specific recommendation, in respect of any original or type of drawing, one way or the other.

I must add that Mr. Hindson specifically authorises me to say, which is important, that he had no clue, at any time during the marking, as to which originals corresponded to the various batches of drawings submitted to him, and that neither my written instructions nor my verbal comments influenced him in any way towards assigning marks to drawings in respect of one original rather than another.

In other words, he did the marking completely 'blind' in every relevant sense, so that his results may be accepted as wholly without bias, and I venture to assert without much fear of contradiction that data obtained under these conditions cannot plausibly be assailed even by the most exacting criticism. I accordingly take Mr. Hindson's figures²⁶ as fundamental in the whole of the investigation of this group of experiments that follows. Only these figures, be it noted, and such others as may be derived from them by rigidly objective processes, are eligible for consideration from the strictly evidential point of view; and only results obtained from them will be claimed as evidence: though, as we shall see later, figures derived from other judgments, into which subjective factors might theoretically be supposed to have entered to some extent, if in practice they probably or apparently did not, may be of considerable interest from the standpoint of information as opposed to evidence.

Alternative Markings: The next points I wish to discuss will be best understood if I describe the circumstances which gave rise to them.

When the drawings of the first experiment with their assigned points, were returned to me by Mr. Hindson, I found that he was marking on a very much more generous scale than I myself had done; for on these thirty-seven sets he had given a total of 272.5 points where I had given only 93.0 - a ratio of almost exactly three to one. This meant that he had been using a very much lower standard, or 'level of acceptance', than I had used; but it did not, of course, mean that his marking was necessarily in any sense 'worse' (or 'better') than mine, for the optimum standard to adopt for detecting a real effect, if there is one, can clearly only be found by experiment. It might quite well happen, indeed - and apparently does - that many resemblances much less noticeable than those which I myself would reckon as 'palpable hits' are yet to some extent inspired, so to say, by the originals and, if this were so, a standard of marking which counted them would be preferable to one that did not. But at the time in question I was full, to the point of obsession, with the idea that it was very important to keep the standard as high as possible; this was partly in reaction from the disappointing results of the Method of Decimal Scoring, in which marks so low as one tenth of the maximum could be given for

26. Hereinafter known for short as the 'M. T. H.' figures.

faint resemblances and associations, and partly because I knew from my own marking that the use of a high standard would lead to significant results.²⁷ It was not unnatural, therefore, that I became apprehensive lest the adoption of so much lower a standard might result in a serious and possibly fatal dilution of the effect I had good reason to suppose was there.

I accordingly asked Mr. Hindson to raise his standard somewhat in future, and at the same time I cast about in my mind for some method of eliminating the feebler resemblances, if need be, on as objective a basis as possible.

As regards the first, I may say that Mr. Hindson adjusted himself to my demands with great success; in marking experiments II to IVB, he gave 576.0 points as compared with my 466.0, which is a ratio of only about 1.25 to 1.

As for the second, I finally decided on the plan of picking out all the drawings on the marking of which Mr. Hindson and I did not exactly agree²⁸ and submitting them to independent arbitration. This seemed to be the best practicable expedient in the circumstances, which were intrinsically somewhat difficult (August, 1939), even though it was not absolutely rigid evidentially, for reasons which I shall discuss in a moment. I wanted, first of all, to be quite satisfied in my own mind that the results of the marking I had done were not due to pathological aberrations on my part; so that in the event of Mr. Hindson's marking yielding null results I could be reasonably sure that this was due to his employing too generous a standard and not to there being nothing to detect. If, on the other hand, Mr. Hindson's unarbitrated marking were to lead to a significant positive result (as it did) no harm would be done; on the contrary, a very approximately rigid alternative assessment of this

27. I suppose I must point out here that the fact of my own marking being inadmissible as strict evidence in favour of the occurrence of a real effect does *not* render it useless as a guide; on the contrary, the fact that it was very carefully and conscientiously done by the person who knew more than anyone else about the originals might well compensate for the possibilities of unwitting bias already mentioned. I knew very well that though I could not claim it as evidentially rigid, there was uncommonly little wrong with it. W.W.C.

28. I.e., including those to which one of us gave a full mark and the other a half, as well as those to which one gave a full or half mark and the other nothing.

kind, using an appreciably higher standard than Mr. Hindson's, could hardly fail to be of considerable interest from the comparative point of view²⁹.

There are two reasons why these arbitrated figures may not be perfectly rigid. First: It might be suggested that I myself, wittingly or unwittingly, had marked potential winners with generosity and potential losers with meanness. If this were so and if, as was the case, M.T.H. marked much more generously all round, the result would be that relatively few potential winners would come up for arbitration, because he would have passed most of my selections; but there would be many potential losers, because I should have disagreed in advance with his losing selections. If then, as was also the case, the arbitrators were urged to adopt a high standard, these losers would be extensively rejected. As a result, the proportion of winners would be unfairly inflated. This supposition is contra-indicated by the evidence. M.T.H.'s original entries contain 280 or 23.16% potential winners, while the 934 entries submitted for arbitration contain 213, or 22.81; the difference is trivial by inspection.

Second: Professor C. D. Broad and Professor H. H. Price very kindly consented to do the arbitrating. Of these, the former had considerable knowledge of the originals, though the latter had not, and it is not altogether impossible that this knowledge might have introduced some slight measure of unconscious bias into some of the judgements. But this also lacks material support. If such a tendency were appreciably operative, we should expect to find Professor Broad giving considerably more points to winners and fewer to losers as compared with Professor Price. There is a very slight tendency to this effect to be observed in the 586 cases in which independent judgements from both are available³⁰; but when the figures are tested for significance by the ordinary 2 x 2 method, we find that a discrepancy as large as that observed, or larger, would occur a shade more often than not by the operation of chance alone.

-
29. External circumstances, I may remark, made it impracticable to postpone action until after all Mr. Hindson's figures had been received, collated and computed.
30. Professor Price was unfortunately unable to complete the whole of the work. For the latter part of experiment IVA and the whole of IVB, the onus of arbitration fell on Professor Broad alone.

arbi
are
purp
rath
for
figu

a ge
too
mome
opti
tag
cons
whet
in
ably

star
resu
the
is
fect
ed
cha
eff
int
In
full
in
to
will
from
tra
abl

by
and
con
cau
jud
pol
aft
the
ing

31.

We may accordingly conclude that, although the arbitrated figures are not evidentially impeccable, they are none the less entirely satisfactory for all practical purposes, and may be safely used wherever information rather than the highest grade of evidence is in question; for the latter, we may rely on the unarbitrated M.T.H. figures and their derivatives.

Derived Figures: I have already pointed out that a genuine effect, if present, might be obscured by either too generous or too strict a standard of marking, and a moment's reflection will show that there must be some optimum standard which will bring out to the best advantage whatever effect there may be. It is clearly of considerable practical interest to ascertain, if we can, whether we are approximating to this optimum, or whether, in future work, we should do better to employ appreciably higher or lower standards.

This can, of course, only be done by varying the standard in both directions and seeing whether the best result (after making due allowance for differences in the size of the sample) lies between these limits. It is fortunately possible to do this very easily, and perfectly objectively, in the case of any set of data scored on the full mark, half mark, or zero method; for changing all half marks into full marks will give us the effect of using a lower standard, while changing them into zeros will be equivalent to raising the standard. In the one case we treat all doubtful assessments as full-fledged winners, which may well be over-generous; in the other we treat them all as misses, which is likely to err on the side of strictness.³¹ These figures, which will be discussed in due course, were easily obtained from the original M.T.H. data, but to obtain the arbitrated figures and their derivatives called for considerable care and labour in tabulation and checking.

Tabulation, etc.: I began, as already implied, by preparing lists of all drawings about which M.T.H. and I did not exactly agree; but to these I added a considerable number about which we did agree, either because I was not altogether confident that even our agreed judgement was sound, or because it seemed to me that our policy had not been wholly consistent. From these lists, after the arbitrators had recorded their judgements on them, I prepared 'Summaries of Arbitration' for the drawings of each experiment in turn. Each such summary

31. *Mutatis mutandis*, a similar procedure may be applied to all data obtained from a graduated system of marking.

showed, for the drawings of the experiment concerned as compared with each of the fifty originals, (a) the number of hits on the original as agreed by M.T.H. and myself and not submitted to arbitration, (b) the number of drawings submitted which were unanimously discarded by the arbitrators, (c) in detail, the mark given by each arbitrator to those drawings they did not unanimously discard, and (d) the lower of these two marks, if different. This last was because it was agreed at the time that, 'in the interests of conservatism' and to maintain a high standard, the lower of the two arbitrations should be taken; but there was, of course, nothing binding or sacrosanct about this decision, and I shall also give the results obtained by taking the higher of the two: note that, since there was only one arbitrator for the latter part of IVA and the whole of IVB, the 'high' and 'low' arbitrations are identical in this region. From these Summaries it was easy to extract the requisite information, namely the number of hits scored by the drawings of each experiment on the originals of their own and of the other experiments.

When I had done this, I retabulated the whole of the data *de novo*, on a larger and more elaborate scale, in a kind of Master Table arranged by originals in alphabetical order instead of by experiments in chronological order, and showing every entry made by M.T.H. or myself and the judgments of both arbitrators whether unanimous or not. By collecting the same figures from this table as from the summaries and comparing the two, I was able to detect any discrepancies there might be and to track them to their origin. At the same time I checked through all the arbitrators's decisions (and a great many of M.T.H.'s and my own) with a view to detecting inconsistencies of marking. A few instances were found, and submitted to Professor Broad for a ruling, which was then applied to all relevant cases. A few more drawings were discovered which had somehow managed to slip through the net, e.g., had inadvertently been omitted from arbitration; there were not more than ten or a dozen of these, and nearly all could be confidently marked by analogy with others of the same kind on which the arbitrators had given decisions; in only some three or four cases was I obliged to rely on my own judgment, and in these I was careful, so far as was compatible with common sense, to weight the marking against rather than in favour of the effect looked for.

Thus the arbitrated figures, while lacking the perfect evidential rigidity of the M.T.H. unarbitrated,

repr
scie
usin
part
bias
cert
the
thei
sing

ple
givi
orig
ther
to w
his
plet
call
good

represent the product of an extremely careful and conscientious attempt to form considered composite judgments, using a somewhat higher standard than M.T.H., on the part of judges who, where they had knowledge that might bias the result, did their best to discount it. For certain purposes, such as determining the magnitude of the effect at the level of acceptance adopted, despite their lack of complete rigidity, are preferable to any single judgment however 'blind' it may have been.

To Sum Up: The method is based on the very simple plan of marking all drawings against all originals, giving a full mark if a drawing plainly resembles any original, nothing if it does not, and a half mark if there is real doubt. Mr. Hindson had no kind of clue as to which drawings were aimed at which originals, so that his figures and their derivatives may be taken as completely rigid. The arbitrated figures, though theoretically less perfect, are practically reliable and amply good enough for informative purposes.

Section IV

RESULTS

A. Main Results

1.a. Calculation of the Main Results from the M.T.H. Figures: We are now in a position to approach the central and crucial point of the whole enquiry, namely the question of whether or not percipients tend, to a significant extent to score relatively more 'hits' on the originals of the experiment in which they are engaged than they do on those of experiments in which they are not engaged.

The simple mathematical technique necessary for this purpose has been discussed above, and we will start by applying it to Mr. Hindson's unarbitrated and unmodified figures. These are shown tabulated in the appropriate manner at the top of Example I, where the whole calculation is given in full.

This table should be interpreted as follows; When the originals were removed from the alphabetical sequence in which they had been given to Mr. Hindson and were regrouped according to the experiments in which they were used, it was found that the drawings done by the percipients of experiment I had been credited with 51.5 hits on the originals used in experiment I, with 77.5 on those used in experiment II, with 36.5 on those of experiment III, with 25.0 on those of experiment IVA, and with 82.0 on those of experiment IVB, making a total of 272.5 assorted hits by these percipients. Similarly the percipients of experiment II scored 7.0 hits on the originals of experiment I, 18.0 on those of their own experiment. 12.5 on the originals of experiment III, and so on throughout the table, which shows a grand total of 848.5 hits scored on all originals by all percipients put together. Alternatively, we may read along the lines instead of down the columns to the effect that the originals of experiment I had 51.5 hits scored on them by the percipients of experiment I, 7.0 by the percipients of experiment II, 5.5 by those of experiment III, and so forth, showing a total of 127.5 scored on them altogether. Either way, it is clear that the figures in the diagonally placed cells, set in heavy type, are the 'winners' in

whic
ers'
find
this
winn
of h
of o
the
vary

127.
the
tion
ther
ient
inal
the
hits

we s
scor
ents
hits
atio
expe
appr
and
we a
of w
ment
and
sum
ampl
conc
we f
163.
whic
D, t
is t
cipi

which we are interested, while all the others are 'losers'. Summing the figures in the diagonal cells, we find 201.5 as the total number of winners scored; but this is of no use to us unless we also know (a) how many winners we should expect to find, given the known number of hits by each group of percipients and on each series of originals, if chance alone were operative, and (b) the extent to which this expected number is likely to vary under these conditions.

Recapitulating an earlier passage, we know that 127.5 hits out of a grand total of 848.5 are scored on the originals of experiment I; that is to say, the frac-

tion $\frac{127.5}{848.5}$ of all hits. Since, on the null hypothesis, there is no reason to suppose that one group of percipients would score relatively more or less on these originals than any other, we should expect this fraction of the total hits scored by any group of percipients to be hits on the originals of experiment I. In particular,

we should expect $\frac{127.5}{848.5}$ of 272.5 to be the number of hits scored on the originals of experiment I by the percipients of experiment I. This comes to about forty-one hits expected, so it is clear that, in this case, expectation has been handsomely exceeded. Thus, to find the expected number of winners in any case, we multiply the appropriate column total by the corresponding row total and divide the product by the grand total.³² But since we are only interested here in the total expected number of winners, and not in the expectation in each experiment separately, it is quicker to multiply the column and row totals in pairs, sum the products, and divide the sum by 848.5. This is shown in the first step of the example, where it will be seen that the sum of the products concerned amounts to 138,545.25; dividing this by 848.5 we find the total expected number of winners, E_w , to be 163.28. Subtracting this from O_w , the observed number, which we have already found to be 201.50, we obtain for D , the difference between them, a value of 38.22. That is to say, according to Mr. Hindson's marking, the percipients have succeeded in obtaining between them some

³². It should be noted that this procedure automatically allows for the varying 'popularity' of the originals of the different experiments; for this is measured empirically by the total number of hits scored on the originals concerned by all the percipients taken together, i.e., by the marginal totals 127.5, 181.0, etc., and the expected numbers calculated are proportional to these. *Mutatis mutandis*, the same is true of the differing numbers, or activity, of the percipients in the various experiments

thirty-eight more hits, or 23%, on the originals at which they were aiming than we should expect them to do if chance alone were responsible. This sounds very fairly impressive, but whether it really is so, or whether it is the kind of fluctuation we might reasonably expect under chance conditions, cannot be decided until we have found the appropriate measure of chance fluctuation, namely the variance, in the manner already indicated.

We start by summing the corresponding column and row totals in pairs and then multiplying each of these sums, $a + b$, by the previously determined product ab , thus obtaining a new series of products $ab(a+b)$, which are listed in the example and have a total value of 57,863,437.000. The work of calculating the variance is then more or less self-explanatory. We write down the square of the sum of the ab products and add to it the result of multiplying the sum of those products by the square of the total number of hits; from this we subtract the sum of the compound products $ab(a+b)$ multiplied by the total number of hits. Dividing this by $N^2(N-1)$ gives us the variance, σ^2 , which we find comes to 114.468. Taking the square root of this gives the standard error, σ , of 10.699, and when we divide this into the value of 38.22 already determined for D , we obtain a 'normal deviate', D/σ , of 3.572; that is to say, the observed deviation of 38.22 hits in excess of expectation is more than three and a half times its standard error. Reference to the appropriate tables, and interpolation, shows that a deviation of this magnitude or greater may be expected less than once in a thousand such cases, or about once in 2,944, as the result of chance alone.

The result is highly significant, and the chance hypothesis would be regarded as disproved in any normal context.

1.b. Main Results from M.T.H. 'derived' figures:

We may next apply precisely the same method to the figures derived from Mr. Hindson's original data (a) by turning all his half points into full points, that is, by treating all his entries as of equal value, (b) by counting all his half points as zeros (i.e., 'misses') and using only those entries to which he gave a full point in the first instance. As already observed, these figures are every whit as rigid evidentially as the unmodified version, but correspond to a lowering and raising of the standard in the two cases respectively. For comparison, we will also work two results of some interest from the half points taken by themselves and treated simply as 'entries'. There is no need to tabulate these

Table IV

Example I, Main Result for M.T.H. Figures

Table of Points Awarded:

On the Originals of	Hits by the Drawings of Experiments					Total
	I	II	III	IVA	IVB	
I	51.5	7.0	5.5	40.5	23.0	127.5
II	77.5	18.0	7.5	51.0	27.0	181.0
III	36.5	12.5	6.5	45.0	36.5	137.0
IVA	25.0	7.0	4.0	41.0	24.5	101.5
IVB	82.0	19.0	10.0	106.0	84.5	301.5
Total:	272.5	63.5	33.5	283.5	195.5	848.5

Let $a_1 \dots a_5$, $b_1 \dots b_5$ be the row and column totals, taken in order, respectively (or vice versa). Then

$a_1 b_1$	34,743.75;	$a_1 + b_1$	400.0;	$a_1 b_1 (a_1 + b_1)$	13,897,500.000
$a_1 b_2$	11,493.50	$a_2 + b_2$	244.5;	$a_2 b_2 (a_2 + b_2)$	2,810,160.750
$a_1 b_3$	4,589.50	$a_3 + b_3$	170.5;	$a_3 b_3 (a_3 + b_3)$	782,509.750
$a_1 b_4$	28,775.25	$a_4 + b_4$	385.0;	$a_4 b_4 (a_4 + b_4)$	11,078,471.250
$a_1 b_5$	58,943.25	$a_5 + b_5$	497.0;	$a_5 b_5 (a_5 + b_5)$	29,294,795.250
Sums:	138,545.25		1,697.0;		57,863,437.000

E_w 163.28 O_w 201.50; D 38.22

N 848.50 N^2 719,952.25; $N^2(N-1)$ 610,159,531.875

Whence $S^2(ab)$ 19,194,786,297.5625
 $N^2 \cdot S(ab)$ 99,745,964,464.3125

Sum 118,940,750,761.8750
 $N \cdot S(ab)(a+b)$ 49,097,126,294.5000

Subtracting 69,843,624,467.3750

Dividing by $N^2(N-1)$ gives σ^2 114.468

Whence σ 10.699

D/σ 3.572

$P < .001$ or $1/2,944$ v.n.

data in full, or to reproduce the calculations, which exactly follow the procedure of Example I.

The results are of considerable interest, particularly when compared with that already obtained, which is included in table V below for ease of reference. The successive lines of the table show: The total number of points given, the expected number of 'winners', the observed number, the difference between these, the value of the normal deviate, the probability of such a difference (or a greater) arising by chance alone, and the difference expressed as a percentage of the total number of points, the use of which I shall explain later.

It will be seen that all the first four results are handsomely significant, and that the effect of lowering the standard is to push the value of P beyond the one in ten thousand point. No one can reasonably ask more from a test of significance than this, and the supposition that the observed excess of winning over losing hits is due to chance may be dismissed with considerable assurance.

Table Y

	Half and Full Points			Half Points Only	
	Halves as Wholes	as Marked	Halves as Misses	All Halves	Rejected Halves
N	1209	848.5	488	721	546
E _w	227.01	163.28	100.62	128.90	93.76
O _w	280	201.5	123	157	110
D	52.99	38.22	22.38	28.10	16.24
D/σ	4.166	3.572	2.772	2.924	2.043
P <	.000,1	.001	.01	.01	.05
100D/N	4.383	4.504	4.586	3.897	2.974

I hope no one will imagine that the foregoing treatment of the material is no more than an illegitimate monkeying with the data undertaken with a view to extracting a more impressive significance from them than they properly contain. My object is, rather, to throw light on the method of assessment adopted, both intrinsically and as handled by Mr. Hindson; and this, I think, the figures given succeed in doing with some success.

It is at once evident from an inspection of the table that I maligned Mr. Hindson in suspecting that he

was using too low a standard, and that I was wrong in supposing that a very high standard was necessary or even desirable; on the contrary, it will be seen, the improbability of the result being due to chance alone is raised more than ten-fold by ignoring the difference between half and whole points and treating all entries as of equal importance, as is done in the first column of the table. If, on the other hand, we throw the half points out altogether, as under 'Halves as Misses', the significance is greatly diminished. So it is clear that the halves make a very useful contribution to the result instead of being, as I had feared, a dangerous diluent. Moreover, if we separate out the half points and treat them simply as 'entries', by themselves, as is done under 'Half Points Only: All Halves', we find that they yield a strongly significant result. Finally, even those half points which were rejected by myself and at least one arbitrator as not up to so much as 'doubtful' by our standard ('Rejected Halves' in the table) give a result just better than the conventional level of significance.

We are accordingly led towards the very interesting conclusion that the process of cognition involved is not an all or none affair, but that, on the contrary, the more remote resemblances may often represent just as genuine an effect as the closer, even if it be more difficult to detect. But before accepting this conclusion, even provisionally, there are two possibilities which must be considered.

First: May it be that the drawings receiving half points are largely those which are badly drawn? I am fairly sure that this is not the case, though I do not see how I could easily justify my belief. But it is scarcely more than a matter of common sense to realise that the veriest scrawl of a tree, a boat, a jug, an anchor, a balance, a cow, a butterfly, and so forth, will be unmistakably *recognisable*, which is all that is necessary to gain a full point. Moreover, the instructions, (q.v.) were explicit as to disregarding lack of skill in draughtsmanship and, I am confident, were duly followed. The situations which led to the giving of half points were rather those in which it was necessary to decide whether the clearly depicted object was near enough to the original in nature to justify a point being given. For examples: Is the outline of a Latin Cross to be given anything for flag, in which a cross is the most conspicuous feature? Should a snail have a point for shell? Should a steamer, or a rowing boat, or a gondola, have one for boat? What sort of leaves are like enough

to those in exfoliate to justify a point? In very many cases such as these the judge would end by giving a half point; and it is they together with a certain number of vague, as opposed to ill-executed, drawings which are mainly responsible for the large number of half-point entries.

Second, there is the possibility that Mr. Hindson may have been exceptionally chary of giving full points at all, and may have diffidently given half points where a less cautious marker would have given wholes. I think I am prepared to do so far as to say this is quite definitely not the case. Mr. Hindson gives a total of 488 full points; I appear to have given 453, so that he is slightly more lavish than I am - not less - though only by about 7½%. On the other hand, only 175 of his 721 half points survive arbitration at all, and only five are converted by the arbitrators into full points; so he clearly has not marked a large number of 'palpable hits' too low. What he has done is to be appreciably more generous in the matter of full points than I, and to give no fewer than 721 half points to my 212. This is in no sense a reflection on his marking, which, in my opinion, would be difficult to improve upon. On the contrary, because he was quite free from bias, as no one who 'knew the answer' could possibly be, and could approach the work with a free and open mind devoid of guess-founded obsessions, he was able to recognise a larger number of resemblances which I was afraid to admit for fear of 'dilution', though they were in fact to an important extent non-fortuitous.

On the other hand, his half-point entries are of a lower quality, as is to be expected, than his full-point; and it will be well to devote a few paragraphs to this question of 'quality' before we go on to discuss the arbitrated data.

2. Measures of Quality: At the beginning of an investigation of this kind we are, very properly, chiefly interested in deciding whether any real effect occurs at all: but as soon as our basic experiments have given a sufficiently significant result to satisfy us, even provisionally, that this is the case and that there is something worth studying in our material - and we certainly claim to have achieved this here - we shall want to answer questions of a much less simple character. For example, we shall want to know whether one group of percipients is relatively more successful than another, whether women do better than men, and so forth. In all

such cases it will often be necessary to use some measure of success which is independent of the total number of hits recorded, for in general this will merely be proportional to the number of percipients concerned. At the moment, it would be of interest to know whether some standards, or methods of judging, are more *efficient* than others, in the sense of enabling us to select from the raw material a group of drawings containing a higher ratio of genuinely cognised specimens to fortuitous resemblances, as opposed to being merely more *effective*, in the sense of yielding a higher level of significance.

For it is not to be imagined that all 'winner' resemblances, however close, in a sample showing a significant result, however high, are necessarily the result of genuine cognition; some are, we may be sure, for otherwise the significant result would not be obtained, but others will be purely fortuitous, and I see no means at present of telling which are which with any kind of assurance. And it might well be of importance, for some purposes, to employ material having the highest possible ratio of genuine to fortuitous hits.

Now it is easy to see that if there is a real effect, then of two samples of different sizes but identical constitution the larger will show the higher significance, simply because it is larger; indeed, a large sample of low quality may even prove more significant than a small sample of higher quality. To a close approximation, if there is a real effect and N is larger, the result of multiplying the size of the sample by n , will be to multiply the value of the normal deviate n ; thus, two samples of different sizes may reduce to level terms by dividing the normal deviate in each case by \sqrt{N} . But this gives a result approximately proportional to D/N . It is for this reason that I have here adopted the percentage of excess winners, $100D/N$, as the most appropriate criterion of quality to use here.

Referring now to Table V, let us see what light this throws on what has been going on. In the first place, we see that, although the significance falls as the standard is raised, the quality rises; but the rise over the first three columns is very small, and some way from significance, though in the expected sense. This would appear to imply that the full points given are not worth very much more than the half points, and to a certain extent I think this is true; for I have the impression that some of Mr. Hindson's full-point markings were a trifle optimistic, and we shall see later that a

considerably higher value of Q (as I may conveniently call all such measures of quality) is obtained from these full-point entries of which no one concerned had any doubts. On the other hand, we find a noticeable drop in quality, as we should expect, when the half points are isolated and treated by themselves, and this is still more marked when we consider only the halves rejected by the arbitrators.

From these facts we may fairly confidently draw the following conclusions, which are not without importance:

1. The cognitive process involved is not of an all or none character. That is to say, it may manifest itself in the production of drawings which are by no means unmistakable likenesses of the originals (witness: 'All Halves') and even in such as bear what many people would consider no more than faint resemblances to them (witness: 'Rejected Halves').

2. The lower limit of useful resemblance, if I may put it so, is probably being approached by Mr. Hindson's more optimistic half-point markings; for the Rejected Halves, of which we have the very considerable number of 546, only just creep over the conventional significance level. If the standard were to be made much lower, it seems probable that either there would be no effect at all, or at best that serious 'dilution' would begin to set in. I may easily be wrong here, and further work alone can show where, if anywhere the ultimate effective limit lies; but at present it looks as if the mesh of Mr. Hindson's net had been cunningly chosen so as to bring in pretty well all the worth-while fish.

3. The fact that the values of Q vary just about as we should expect them to goes some way towards assuring us that the effect is a well behaved citizen of the scientific world; it would have been very disconcerting if we had found, for example, that the fainter resemblances were of higher quality than the stronger!

3. Main Results from Arbitrated Figures: As I have emphasized above, results obtained from the M.T.H. unarbitrated figures are alone to be regarded as completely rigid from the strictest point of view; but, for reasons explained, the arbitrated figures have merits of their own. Since the process of arbitration has already been fully described, we may proceed at once to the results, which are given in Table VI below. To the

arbitrated results proper I have added, as a matter of interest, those obtained from my own marking and by treating all entries, whether by M.T.H. or myself, on the same lines as in the first column of table 1: There must be treated as suspect, and no conclusions are based on them. The calculations were, of course, as shown in example I.

Table VI

	W.W.C.	W.W.C. and M.T.H.	Arbitrated High	Low	Full Points
N	559	1329	558.5	486	314
E _w	128.88	260.53	122.24	113.46	71.41
O _w	166.5	323	160.5	140.5	90
D	37.62	62.47	38.26	27.04	18.59
D/σ	4.052	4.564	4.054	3.158	2.770
P	.000,1	10 ⁻⁵	.000,1	.01	.01
Q	6.730	4.701	6.850	5.564	5.920
Mean P				.001	(1/3,588)
Mean Q				6.252	

The first two columns are self-explanatory; the third gives the results obtained by taking the higher of the two arbitrations where there is a choice; the fourth, those of taking the lower: the fifth is obtained by discarding all entries about which anyone concerned had indicated a doubt by giving a half point or less, even if others had given a full point.

Note first the high values reached as regards both significance and quality by my own marking; it is fairly clear that either my knowledge was too much for my efforts to discount it, or that the quite legitimate factor of my understanding of what had been in the agents' minds proved valuable; unfortunately we cannot distinguish with certainty between the two hypotheses. But note also that these values are fractionally exceeded by the 'High' arbitrated figures; as there is no real reason for suspecting these at all seriously, it looks as if my efforts to discount my own bias had been reasonably successful and that the second alternative exerted (as I personally believe) an appreciable influence on the outcome.

Since there is nothing to choose, so far as I can see, between the High and the Low versions of the

arbitrated figures, except that the latter is the more 'conservative' - which is not necessarily a virtue - I have given, at the bottom of the table, values of P and Q obtained from the means of the two sets of figures. It will be noticed that they closely agree with the M.T.H. unmodified figures as regards significance, but achieve this result with the use of only 522 points (mean) instead of 848.5; that is to say, the quality of the sample obtained by the exercise of collective judgment is better than that from a single marker. This is only what one would expect, on the general ground that two heads are better than one; but it may be of some practical importance. It suggests the possibility that the best way to extract the most significant results from a given number of drawings may be to employ a plurality of judges working to a low standard rather than a single judge working to a high standard, and to test successively the data on which they all, all but one, all but two, etc., agree - thus finding empirically the optimum cut-off point, which is presumably that at which real resemblances begin to pass into fanciful.

4. Summary: The results just discussed, above all those of table 1, are of fundamental importance to the whole enquiry. They show clearly that it is in a very high degree improbable that the observed excesses of winning over losing hits are due to chance alone. If chance alone were operative, we should expect to find such a result as that of the M.T.H. 'All Entries' occurring on the average only once in some thirty-three thousand similar groups of experiments. It is accordingly very unlikely that we are being led astray by fortuitous resemblances. Sceptics must either find a flaw in the experimental procedure or conditions of marking, or show (which seems to me an impossible task) that the antecedent improbability of such a phenomenon is at least comparable with the figure just given, or take refuge in avowed intransigence. Meanwhile, we may regard the occurrence of a genuine cognitive process under the conditions described as provisionally demonstrated, to a point which amply justifies continued study and repetition by others of experiments of the kind described.

We also conclude that the cognitive process is not of an all or none character, so that comparatively remote resemblances may be genuine in the sense of being non-fortuitous; and that, consequently, the use of a very high standard is not necessarily the best way of separating the wheat from the chaff.

B. DISPLACEMENT

1. General: The apparent tendency for an original to be reproduced by the percipient on some occasion other than that on which it was exposed was noted in the first experiment and has already been mentioned. If this apparent tendency is real (as we shall see that it is) there can be no doubt that it is a matter of the first importance. If it occurs in one direction only, namely that of making the reproduction appear after the occasion of exposure of the original, there would be a suggestion that the impression received may require some period of gestation in the subconscious before it can emerge into consciousness and be recorded, or perhaps that it only emerges when some relevant extraneous stimulus gives it the necessary fillip; and either of these or similar conclusions would be of appreciable psychological interest. But if it occurs in the opposite sense, so that reproductions tend to appear before the exposure of the originals which they resemble we should be confronted (in so far as the effect was significant) with empirical evidence of a new type for the occurrence of a phenomenon of a precognitive character; and this would be of the very greatest interest from all manner of viewpoints beside the purely psychological.

The matter is accordingly one which must be investigated with the utmost care and rigour, and it will be well to begin by clearing the air of one or two potential misconceptions.

The one thing we must *not* do, of course, is to claim as precognitive every resemblance³³ that occurs before its original and as retrocognitive every one that occurs after it; for this would be merely begging the question. In other words, we must decide before we start on what the words precognitive and retrocognitive may reasonably and usefully be supposed to mean.

It is, as usual, a matter of observed as compared with expected occurrences. We can usefully and validly speak of retrocognitions only if we find a significantly greater number of resemblances than the hypothesis of chance would lead us to expect on some date or dates later than the relevant event; and *vice versa* for

33. The words 'resemblance', 'reproduction' and 'hit' may be regarded as interchangeable and as denoting a drawing sufficiently like an original to have been scored by the judge. The degree of resemblance does not concern us at this stage.

precognitions. Beyond this we cannot safely go before we have studied the facts, though our general experience might lead us to expect a continuous falling off in the frequency of the resemblances as the remoteness from the event increases. But we should have no right to complain, if some to be surprised, if it were found that the falling off were of a periodic or fluctuating character. But our general knowledge of the way in which all kinds of phenomena occur suggests that, if pre- and retro-cognition be both phenomena of nature, the relative frequency of hits on a given original is likely to increase as the occasion on which a drawing was made approaches that on which the original was exposed, to reach a maximum at or near the coincidence of the two, and to fall off again as the former recedes from the latter.

Our business therefore is to calculate from the observed data the number of hits expected to occur on occasions one, two, three, etc., places before, and one, two, three, etc., places after, the event concerned, namely the exposure of an original, and to compare these calculated values with the numbers actually found.

2. Illustration of Procedure: To illustrate the procedure, which becomes laborious when large numbers of occasions and originals are involved, we will take the relatively simple case, of no special importance intrinsically, of the originals and occasions of the first experiment.³⁴

The tabulated data and the results of the calculations are shown in Example IV, the first part of which should be read as follows: On the first occasion one hit was scored on the first original (Bracket), on the second and third occasions no hits, on the fourth occasion one hit, on the fifth two hits, etc., . . . a total of seven 'hits' or resemblances to the original bracket being found among the drawings of the first experiment percipients. Or alternatively: Among the drawings done on the first occasion there were found one resembling bracket, one resembling buffalo, none resembling fess or hand, two resembling cross-stitch, etc., etc., . . . a total of five hits on one or other of the originals of experiment I being found among the drawings done on the first occasion.

It will at once be realised that the figures in the 'leading diagonal', shown in heavy type, represent

34. I again use the 'M.T.H. All Entries' data, and shall do so henceforward unless otherwise expressly stated.

the number of hits scored on the several originals on the occasions on which they were exposed; that is to say, hits not displaced in either direction but coinciding with the event - or, more accurately, occurring within the period of exposure, which constitutes coincidence for our purpose. It is also easy to see that the next diagonal upwards and to the right, parallel to the leading diagonal, contains the numbers of hits scored on the occasion immediately *after* the event, of which there were one for buffalo, one for hand, two for cross-stitch, two for bottle, etc.; whereas the diagonal below and to the left of the leading diagonal shows the numbers of hits scored on the occasion immediately *before* the event, of which there were only one for buffalo, one for bat and two for net. Similarly, as we proceed outward from the leading diagonal, we find diagonals containing the numbers of hits scored on the second, third, fourth, etc., occasions before or after, as the case may be, that on which the original they resemble was exposed. Our task is to determine whether there is any significant tendency for the observed numbers in these diagonals to exceed their expected values as we approach the leading diagonal from either direction, or both.

The procedure is simple enough in principle, and is basically the same as that which we adopted for obtaining the expected number of hits on the originals of a particular experiment by the drawings of that experiment when we were calculating the main results *q.v.* The number of hits to be expected in any 'cell' or compartment of the 10 x 10 table of Example IV is obtained by multiplying together the two marginal totals concerned and dividing the product by the total number of hits. Taking the extreme bottom left hand corner as an example, we note that only one hit was scored by these percipients on the original anchor, being 1/81 of the total; altogether, five hits were scored on these ten originals on the first occasion; so, if no cause other than chance is operative, we should expect the same proportion of anchors among these first hits as in the total, *viz.*, 1/81, from which it follows that the actual expected number of hits will be 5/81, or .062. In practice, of course, one cannot have .062 of a hit, and the expectations in individual cells are far too small to be useful or interesting, but this does not affect the principle involved. For the next diagonal nearer the centre, the expectation will be (3 x 5)/81 for the beetle cell, and (1 x 7)/81 for the anchor cell, making a total of 22/81, or .272. In short, to determine the expected number of hits in any diagonal we sum the products of the successive pairs

of marginal totals concerned and divide by the total number of hits in the table. When we do this, we obtain the figures given under E in the lower part of the example, where 'Diagonal - 9' means the diagonal containing the hits made nine places before the original to which they refer, and so forth. O is the observed number of hits in these diagonals, as may be verified from the table, while the quantity $(O-E)/\sqrt{E}$ gives a measure of the deviation from expectation, reduced to a standard level, on the hypothesis that chance alone is responsible for the distribution in the various diagonals.

The Example, as I have said, is only illustrative of procedure and I do not propose to discuss the not very interesting results in detail here. The only points worth noting are, first, that the observed number of hits of zero displacement, i.e., made on the same occasion as that on which the original was exposed, is fractionally below expectation, instead of being handsomely above as we might perhaps have expected from the fact that the experiment as a whole scores 16.54 hits above the expected value; second, that, although the quantities $(O-E)/\sqrt{E}$ are distressingly ragged and not very informative as they stand, they do contrive to show an intrinsically (but not contextually) significant 'peak' at diagonal 2, while the observed numbers are greater than expectation in the positive (retrocognitive) part of the table and less in the negative (precognitive part); these last two facts accord, though somewhat feebly, with the qualitative impression first formed to the effect that hits tended to be deferred. The difference may be attributed to the fact that the impression was formed on the basis of the stronger resemblances only, whereas the data of Example IV include the feebler.

3. Compilation of the 50 x 50 table: As a matter of historical fact, I first satisfied myself that there was a displacement effect worth investigating by a study, on the lines indicated above, of the inter-experiment data given in Table I, but it will be more convenient to defer consideration of these for the present and to proceed at once to the crucial large-scale investigation of all occasions and all originals taken together.

This involves the preparation of a 50 x 50 table, a task which was complicated by the fact that in all cases except experiment II it was necessary first to 'decode' the random numbers which had been given to the 1,209 drawings involved for use in connection with the matching technique. In this table the cross-headings of the rows from top to bottom were the fifty originals in

the order in which they were exposed in the five successive experiments, and the headings of the columns from left to right were the fifty successive occasions on which drawings were made in these experiments. In each cell, formed by the intersection of a row with a column, was entered the number of drawings, made on a certain occasion, which the judge had assigned to a certain original. Nothing would be gained by presenting the full table here, particularly as we are not interested in individual cells, the expectations for which I have not calculated, but only in the total numbers expected and observed in the diagonals. Nor do I think it would be worth while to give full details for all the ninety-nine diagonals; but in Table II I give an illustrative sample consisting of three groups (the second, central and last but one, counting from the extreme bottom left hand corner of the 50 x 50 table) of nine diagonals each, and the sub-totals for the remaining eight groups of nine into which the ninety-nine diagonals may conveniently be divided. This will be sufficient to show the procedure adopted and the kind of results obtained. It should be noted that the values of $(O-E)/\sqrt{E}$ and of χ^2 given in the lines for Sub-totals are obtained from the sub-total values of O and E, and not by summing the results for individual diagonals; that is to say, in computing these quantities, the data are pooled for each group of diagonals.³⁵

It will be seen that the individual values of $(O-E)/\sqrt{E}$, in which we are chiefly interested, are distinctly irregular; but the sub-totals leave little doubt as to what is going on.

Note, however, before we pass on to discuss these, the very remarkable 'dip' at the centre (diagonal 0), with 'twin peaks' either side of it at -2 and 2. We will

35. The quantity E is, of course, only an approximation to the standard error of $(O-E)$. For the extreme case of a diagonal containing only a simple cell, it is clearly very nearly true since the probability of a hit falling in them is so small; for the leading diagonal of the whole 50 x 50 table it is 4.883, while the exact value of the standard error, calculated by Mr. Stevens' formula is 4.764. The discrepancy is negligible, but on the safe side in the sense of tending to make the values of $(O-E)/\sqrt{E}$, and consequently of χ^2 , somewhat too small. On the other hand, it should be noted that the values of χ^2 are not necessarily strictly additive, since the distributions in the various diagonals may not be independent. This, however, will not affect the implications of the highly systematic arrangement of the points in Fig. 1.

discuss the significance and interpretation of this at a later stage, but we may note here that the peak at diagonal 2, where there is an excess of no fewer than 19.384 hits over expectation, is very definitely significant even in the context of ninety-eight other diagonals, for the value 4.269 gives P less than .000,02, and even the symmetrically placed peak at -2 passes the 1 in 4,000 mark.

To bring out the salient features to better advantage, I retabulate the Group Sub-totals in table VII below, and show the values of $(O-E)/\sqrt{E}$ plotted in Figure I.

Table VII

Group	$(O-E)/\sqrt{E}$	
-5	-1.031	} Precognitive Group
-4	-1.973	
-3	-1.678	
-2	.524	
-1	1.186	
0	3.873	
1	.760	} Retrocognitive Group
2	-.922	
3	-1.262	
4	-2.083	
5	-.542	

The null hypothesis is, of course, to the effect that no diagonal or group of diagonals is more likely than another to show an excess or deficit of hits as compared with expectation. According to this, the values of $(O-E)/\sqrt{E}$ should be randomly distributed about a mean of zero with unit standard deviation. This, by inspection, they quite evidently are not, so that we hardly need the value of 32.09 for $S(\chi^2)$, which gives P just beyond the .001 level of significance, to tell us that the hypothesis is untenable. Moreover, deviations from expectation are very fairly symmetrical, as may be seen at a glance from Figure I, so we may conclude that the precognitive effect is just about as strongly indicated as the retrocognitive.

If we wish to satisfy ourselves as to the significance of the two sides separately, we may conveniently group the diagonals in seven groups of seven on each

side of the leading diagonal and find the value of $S(\chi^2)$ with seven degrees of freedom in each case. This gives:

	$S(\chi^2)$	P
Precognitive side	20.979	.01
Retrocognitive side	14.481	.05

Curiously enough, the precognitive effect, which most people would consider the less probable antecedently, appears to be the more strongly marked, or at least somewhat the better substantiated.

4. Introduction of the Quantity π : The quantity $(O-E)/\sqrt{E}$ is admirable for testing significance, but has certain disadvantages from other points of view. Since, from the nature of the calculations, the total observed and expected frequencies must be equal, it follows that any excess of the first over the second in one region must be balanced by a deficiency in another. Thus the fact that the line in Fig. 1 goes up to a peak in the middle necessitates a depression or depressions at one end or both ends; consequently the line is unlikely to represent to the best advantage the true relation between displacement and incidence of hits.

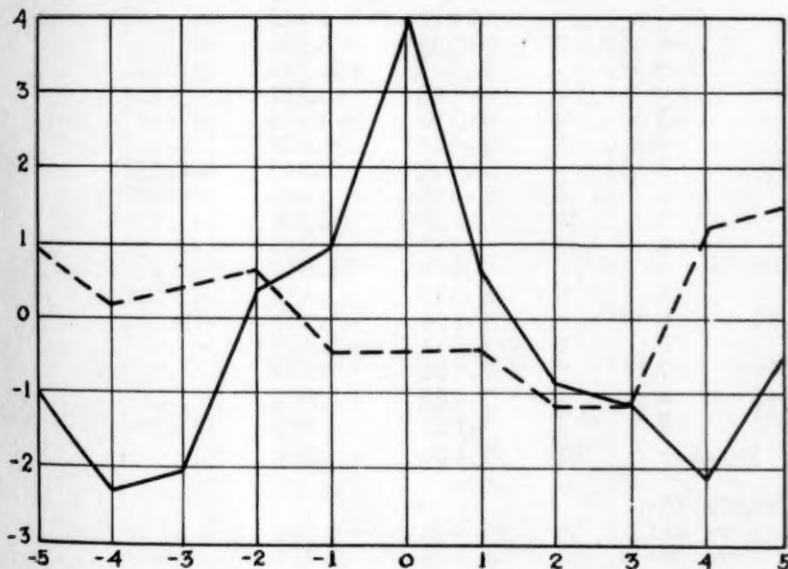


FIGURE 1

N.B. The dashed line shows the corresponding values for the Control Marking, q.v.

Table VIII

Example IV, Distribution of Hits on Originals
of Experiment I Made on Occasions of Experiment I

Original	Occasions										Total
	1	2	3	4	5	6	7	8	9	10	
1. Bracket	1	.	.	1	2	.	1	.	.	2	7
2. Buffalo	1	1	1	4	2	1	1	.	2	2	15
3. Embattled	.	.	1	.	2	1	4
4. Hand Fess	.	1	.	1	1	1	2	2	.	1	9
5. Cross-stitch	2	2	3	.	3	2	1	.	1	2	16
6. Bottle	.	2	.	2	.	.	2	3	.	.	9
7. Bat	1	.	.	2	.	1	.	1	3	1	9
8. Net	2	2	.	2	2	8
9. Beetle	1	.	1	1	3
10. Anchor	.	1	1
Totals	5	7	5	10	10	7	10	6	9	12	81

Then we have

Diagonal	0	E	(0-E)	$(0-E)/\sqrt{E}$
-9	.	.062	-.062	-.249
-8	1	.272	-.728	1.396
-7	.	.815	-.815	-.903
-6	1	1.556	-.556	-.446
-5	.	2.321	-2.321	-1.523
-4	4	3.763	.237	.122
-3	4	4.975	-.975	-.437
-2	9	5.371	3.629	1.905
-1	4	7.013	-3.013	-1.137
0	8	8.025	-.025	-.088
1	10	7.630	2.370	.858
2	16	8.383	7.617	2.630
3	6	7.692	-1.692	-.610
4	6	6.432	-.432	-.170
5	3	6.124	-3.124	-1.262
6	2	3.753	-1.753	-.905
7	3	2.778	.222	.133
8	2	3.000	-1.000	-.577
9	2	1.037	.963	.946
Totals:	81	81.002	-.002	
Totals for				
- ve half:	23	26.148	-3.148	-.616
+ ve half:	50	46.829	3.171	.463

Table IX

Samples and Group Sub-totals of the 99 Diagonals of the 50 x 50 Table (Groups of Nine Diagonals)

Diagonals	O	E	(O-E)	(O-E)/√E	χ ²
<u>-49 to -41</u>	43	50.312	- 7.312	- 1.031	1.06
<u>Sub-totals</u>					
-40	11	12.900	✓ 1.900	- .529	
-39	9	14.084	- 5.084	- 1.355	
-38	14	13.474	.526	.143	
-37	14	14.371	- .371	- .098	
-36	12	11.778	- .222	.065	
-35	6	8.556	- 2.566	- .877	
-34	4	9.143	- 5.143	- 1.701	
-33	3	8.933	- 5.933	- 1.985	
-32	10	9.776	.224	.072	
<u>Sub-totals</u>	83	103.025	-20.025	- 1.973	3.89
<u>-31 to -23</u>					
<u>Sub-totals</u>	63	77.801	-14.801	- 1.678	2.82
<u>-22 to -14</u>					
<u>Sub-totals</u>	121	115.370	5.630	.524	.27
<u>-13 to -5</u>					
<u>Sub-totals</u>	220	203.100	16.900	1.186	1.41
-4	30	24.298	5.702	1.156	
-3	19	23.995	-4.995	-1.020	
-2	40	22.544	17.456	3.676	
-1	33	24.288	8.712	1.767	
0	28	23.848	4.152	.850	
1	31	21.460	9.540	2.059	
2	40	20.616	19.384	4.269	
3	18	20.555	-2.555	- .563	
4	15	17.721	-2.721	- .646	
<u>Sub-totals</u>	254	199.325	54.675	3.873	15.00
<u>5 to 13</u>					
<u>Sub-totals</u>	139	130.325	8.675	.760	.58
<u>14 to 22</u>					
<u>Sub-totals</u>	113	123.239	-10.239	- .922	.85
<u>23 to 31</u>					
<u>Sub-totals</u>	103	116.631	-13.631	-1.262	1.59
32	8	11.278	- 3.278	- .976	
33	10	10.448	- .448	- .139	
34	10	10.088	- .088	- .028	
35	5	9.422	- 4.422	-1.441	
36	6	8.001	- 2.001	- .708	
37	3	6.641	- 3.641	-1.413	
38	3	5.596	- 2.596	-1.098	
39	6	4.840	1.160	.527	
40	2	4.173	- 2.173	-1.064	
<u>Sub-totals</u>	53	70.487	-17.487	-2.083	4.33
<u>41 to 49</u>					
<u>Sub-totals</u>	17	19.386	- 2.386	- .542	.29
TOTALS	1,209	1,209.001	- .001		32.09

Table X

M.T.H. 'All Entries' Data,
Showing Expectations and Differences

Hits by the Drawings of Experiment

On the Originals of	I	II	III	IVA	IVB	Total
0	81	10	9	53	33	186
I E	64.46	13.23	7.69	59.98	41.23	185.99
0-E	16.54	- 3.23	1.31	- 6.38	- 8.23	
0	125	26	11	76	44	282
II E	97.73	20.06	11.66	90.03	62.51	281.99
0-E	27.27	5.94	- .66	-14.03	-18.51	
0	61	15	9	66	52	203
III E	70.95	14.44	8.40	64.81	45.00	203.00
0-E	- 9.35	.56	.60	1.19	7.00	
0	43	11	7	58	33	152
IVA E	52.68	10.81	6.29	48.53	33.69	152.00
0-E	- 9.68	.19	.71	9.47	- .69	
0	109	24	14	133	106	396
IVB E	133.78	27.46	15.96	123.24	85.56	386.00
0-E	-24.78	- 3.46	- 1.96	9.76	20.44	
Total 0	419	86	50	386	268	1,209
Total E	419.00	86.00	50.00	385.99	267.99	2,208.98

NOTE: The expectation in any cell is given by the product of the appropriate marginal totals divided by the total number of hits (1,209).

The total expectation for the leading diagonal is 227.01; the total number of hits observed is 280; the difference is 52.99. Thus the value of χ^2 for the leading diagonal is very nearly

$$53^2/227 + 53^2/982 = 15.23, \text{ with } P \text{ less than } .000,1.$$

It would clearly be a matter of the greatest general interest if we could establish this relation definitely; if, that is to say, we could determine the 'law' connecting the magnitude of the real effect with displacement. This might well throw light on the nature of the process involved, or at least enable us to say that the facts are incompatible with certain hypotheses. But before we make even the most tentative steps in this direction we must be clear in our minds as to the senses in which we are using the terms 'magnitude of effect' and 'displacement'.

If we were dealing, let us say, with magnetic phenomena, there would be little room for ambiguity; the 'magnitude of effect' would be measured by the pull in dynes exerted by one magnetic pole on another or on a piece of iron, and 'displacement' would be the distance in centimeters between the two, and we should have little difficulty in ascertaining that the one varied inversely as the square of the other. In this case there is no 'magnitude' of this nature that we can measure; the phenomenon is a matter of the relative frequencies with which hits occur in different positions with respect to the originals on which they are made, and it is in terms of such frequencies, therefore, that we must conduct our study. The question of 'displacement' is not quite so simple, for we may either consider *order* alone, without reference to *time*, as we have done hitherto; or we might transform the horizontal scale of Fig. 1 into a true time scale (e.g., astronomical time) and work in terms of that. For the present, I shall continue to use the bare *order*, mainly because it is simplest and has yielded a reasonable looking arrangement of points; but it will be necessary, sooner or later, to substitute the mean values of the astronomical time intervals as abscissae, and to try to determine whether these, or *order* alone, are most relevant to the facts.³⁶

In these circumstances, the ideal would be an empirically determined measure of the *probability* of a hit occurring on an occasion at any given distance (measured in terms of *order*) from the original concerned; we could then test various hypotheses regarding the nature of the phenomenon, leading to correspondingly different theoretical relations between probability and

36. At a guess, and as a matter of rash prophecy, I have little doubt that it will not be astronomical time that will be found most relevant; but I do not wish to embark on an examination of this point at the present stage.

displacement, and see whether some were significantly more compatible with the facts than others.

Another way of approaching the point is to suppose that, in general, it is easier for the percipient to score hits on originals nearer to the occasion on which he is working than on those more remote, and to enquire how 'easiness' varies with 'remoteness'; from this point of view it should at once be clear that the only possible measure of 'easiness' is the relative frequency with which percipients do in fact, on the average, score hits at different remotenesses, other things being equal.

The sting of this is found in its tail, for our chief difficulty lies in the fact that other things are very far from being equal. In the first place, of course, the numbers of occasions of different type vary greatly; that is to say, with fifty originals, there are fifty occasions of zero displacement, forty-nine each with displacement plus one or minus one, forty-eight of displacement plus or minus two, and so on down to only one each with a displacement of plus or minus forty-nine positions. This would not particularly matter if equal numbers of hits were scored on all occasions and on all originals; all we should have to do would be to divide the number of hits occurring on each class of occasion by the number of occasions in that class (i.e., by the number of cells in the appropriate diagonal of the 50 x 50 table), and this would give us a series of numbers representing the relative frequencies of hits with respect to displacement, which is what we want.

Unfortunately, from this point of view, originals differ very widely in popularity, in the empirical sense that some are drawn very much more often than others. moreover, the number of percipients working on different occasions may also vary between wide limits while neither the 'markability' of their attempts nor the generosity of the judge can be assumed constant. Clearly, we shall have to make allowance for all these factors in attempting to estimate the quantity in which we are interested.

Note before we go on that any diagonal, e.g., of the 50 x 50 table, or of Example IV, is built up, so to say, of the hits scored on a number of different occasions and on an equal number of different originals; the common feature is that all the occasions are displaced to the same extent from the original on which the hits are scored. Consider in Example IV, for instance, the

'plus 2' diagonal, which is that starting in the top line and the third column, running downward and to the right, terminating opposite net, and reading . 4 2 1 1 3 3 2. Each of these entries represents the number of hits scored two occasions later than the original concerned; but both the occasion and the original necessarily vary from entry to entry. Thus, no hits are scored on bracket on the third occasion, which is two occasions later than that on which bracket was used as an original, while four were scored on buffalo on the fourth occasion, which is two occasions after that on which buffalo was used. But buffalo was much more popular than bracket (15 hits altogether as compared with 7), while the percipients of the experiment scored altogether 10 hits of sorts, on the originals of the experiment, on the fourth occasion as compared with five on the third. This should serve to make it clear why, in considering the relative merits of different positions considered *qua* displacement, we must correct the observed absolute number of hits, not only for the number of occasions of each type considered but also for the 'popularity' of the originals and the 'activity' of the percipients concerned. The first of these terms is self-explanatory, and is measured by the total number of hits made on the relevant original; by the second I refer to the combined effect of all those factors mentioned above which tend to increase or diminish the total number of hits scored by the percipients of that occasion on all originals, and is measured by that number.

Now it is evident that any increase in these numbers is likely to result in a proportionate increase of hits in the 'cell' or diagonal concerned, quite apart from its displacement; hence, as we are here only interested in the last named variable, the sensible thing to do is clearly to divide the observed number of hits in any cell by the appropriate 'activity' and 'popularity', or the number of hits in any diagonal or group of diagonals by the mean values of these, as well as by the number of cells. This will give us an estimate of what the number of hits per cell would have been if all the originals concerned had been of equal popularity and all the occasions productive of equal activity. For any diagonal, then, our estimate will be

Observed number of hits

(Number of cells) x (Mean Activity) x (Mean Popularity)

and since Mean Activity and Mean Popularity are found by dividing Total Activity and Total Popularity by the number of relevant occasions and originals respectively,

each of which is equal to the number of cells, this reduces to

$$\frac{(\text{Observed number of hits}) \times (\text{Number of cells})}{(\text{Total Activity}) \times (\text{Total Popularity})}$$

The terms in the numerator need no expansion; those in the denominator are obtained by summing the relevant marginal totals of the table concerned, for relevant occasions and originals respectively. Thus for the +2 diagonal of Example IV, we have eight cells and sixteen hits; the relevant occasion totals are 5, 10, 10, 7, 10, 6, 9, 12, giving a Total Activity of sixty-nine; and the relevant original totals are 7, 15, 4, 9, 16, 9, 9, 8, showing a Total Popularity of seventy-seven. Hence the value of our estimate, π , is $(8 \times 16)/(69 \times 77)$ or .024.

The values so obtained are not probabilities; but they may be regarded without much risk of being misled as measures of or indices to them, so that we may conveniently regard π_r as proportional to the frequency of hits that would have occurred on an occasion r places distant from that on which the original concerned was used, if all originals had been equally popular and all occasions equally active.

In Figure II, I show the values of π corresponding to those of $(O-E)/\sqrt{E}$ in Figure I, for the groups of nine diagonals each. It will be seen that, as we should expect, the general tendencies are very similar to those of Fig. I; but we can now say definitely that, ignoring for the moment the upward twists at the ends, the probability of scoring a hit on any original, other things being equal, increases as the occasion of display approaches, reaches a maximum in the near neighbourhood of that occasion, and diminishes as it recedes. The actual values of π multiplied by 10,000 for convenience are

Table XI

Group		
-5	7.027	} Precognitive Groups
-4	6.484	
-3	5.290	
-2	8.480	
-1	10.148	
0	10.276	
1	8.271	} Retrocognitive Groups
2	7.812	
3	7.359	
4	6.125	
5	7.286	

It may be thought that I have devoted an undue amount of space to the introduction of this quantity, which tells us but little more, now that we have evolved it, than we could infer by looking at Fig. 1. My reason is that, although the quantity $(O-E)/\sqrt{E}$ is perfectly satisfactory mathematically, it is too much of an abstraction to be a useful aid to thought; whereas the concepts of the differing proportions of hits occurring at differing remotenesses from the occasion of display, or of their relative frequencies, or of the probabilities of their occurrence (which are virtually interchangeable) are relatively intelligible and should enable us to form some idea of the kind of thing that is going on.

5. Summary and Comments: I fear I cannot claim that the foregoing has done more than scratch the surface of the enormously important subject of Displacement. We have, to be sure, established beyond any doubt (if the authenticity of the data be accepted) that displacement does occur in both the precognitive and retrocognitive senses, and this is a far from trivial conclusion. Not only is it likely to have interesting repercussions in the cosmological field and for psychology generally, but it affords a basis for a plausible explanation (other than fraud, mal-observation, etc.) of why there has been such difficulty in securing reliable repetition of card-guessing experiments. Displacement is easy enough to detect by inspection when we are dealing with drawings of buffalos, hands, guns or windmills, which cannot well be intended for anything else; but the successes of a percipient guessing Zener cards, for example, who consistently scored high but a place or two early or late, might easily escape notice unless specially searched for.³⁷

On the other hand, the number of questions we are not yet able to answer is legion. For example: Is the 'central dip' genuine or chance determined? (Probably genuine.) Are the upward twists at the ends of the $(O-E)/\sqrt{E}$ and π curves genuine, and do they

³⁷ It is very gratifying to find that a scrutiny of some of Mr. Soal's results from this point of view shows clearly that such an effect was very significantly present in some of his subjects. Such confirmation of the effect, coming from a completely independent worker, using entirely different material, a different method, and different percipients, constitutes as strong supporting evidence as could well be desired; and, of course, reciprocally. For these results, see Mr. Soal's paper in this number of *Proceedings*.

foreshadow lateral maxima? (With less assurance I think they probably are and do.) Is the asymmetry of these curves genuine? Does the retrocognitive side decline to a higher 'fixed value' than the precognitive? (Probably 'Yes' to both these) Is the peak for the whole curve, neglecting any local disturbance due to the general dip, significantly displaced in either direction? (Probably not.) Have the curves substantially the same shapes for strong and weak resemblances? And for near and distant percipients? Is bare order, or astronomical time, or something that we might describe as 'relevant psychological time' the proper basis for the horizontal scale?

Probably in all these cases, and undoubtedly in some, we could obtain a certain amount of information by an intensive analysis of the existing data; but I think it will be better to defer the investigation of these questions until more are available, or until further consideration of other points has suggested the most profitable lines of attack.

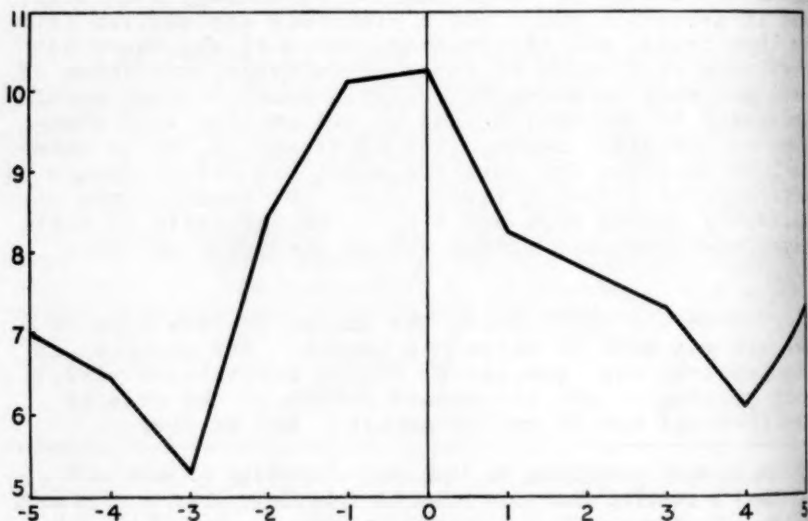


FIGURE II

Section V

RESULTS OF CONTROL MARKING

1. General: When the results described in the preceding Section began to emerge - particularly, of course, the Main Results of Table X and those for Displacement embodied in Fig. 1 - it was suggested that it might be of considerable interest, as a matter of comparison, if a 'control' marking of the same drawings against different originals could be arranged, and this suggestion was approved by those most closely associated with the work. This was not because any of those concerned had detected, or thought he had detected, any flaw in the procedure adopted, which is, indeed, logically impeccable. But it was felt that logical impeccability is not the only means, or even always the best means, of commanding assent and allaying doubts; and there was also, of course, the theoretical possibility that the combined intelligences of Professor Broad, Dr. Thouless, Dr. Irwin, Mr. Gatty and myself, not to mention others who had considered the problem, had failed to notice some serious methodological fallacy. This did not seem very probable; on the other hand many people would consider the results obtained to be also not very probable, and I think we were all agreed that we should feel easier in our minds - even though we might not be able to give logical justification for it - if a control marking were to be carried out and were to yield null results.

For this purpose we were so fortunate as to secure the co-operation of Mr. H. F. Saltmarsh, to whom I am very much indebted for his labours in this connection.

2. Procedure: For the purpose of this control a fresh series of fifty originals (or perhaps I should say 'pseudo-originals') was used. These were drawn by my friend Mrs. Aletta Lewis, to whom I am very much obliged for undertaking the work, and were primarily intended for use in the next experiment of the series, which lies outside the scope of this report. Inasmuch as some of them have been and others may be used for experimental purposes it would be undesirable to mention their nature here; it is sufficient to say that they were drawn in India ink on white paper in the same general style as the true originals, that they did not duplicate any of these used in experiments I to IVB, that their subjects were

selected by a strictly random method from a list of 216 possibilities submitted by me to the artist, and that they represented, on the whole, rather more everyday objects than did those already described. The following are the points relevant to the present issue.

As received from the artist, each original was enclosed, in accordance with my instructions, in a separate envelope and further protected by a sheet of paper against its nature being inadvertently ascertained in the course of ordinary handling. The envelopes were shuffled by Dr. Dingwall and myself, and ten of them were drawn at random for use in the further experiment mentioned above. These were later returned to the pack, so to say, and the whole fifty were re-shuffled by Dr. Dingwall, who then opened them one by one, calling out the title of each to me as he did so. It was agreed that the first ten thus arbitrarily selected should be deemed to be the originals of the first experiment for the purpose of the control marking; similarly, the second ten were allocated to the second experiment, the third ten to the third, and so forth. The allocations thus determined were listed by me at the time of selection under the heads of experiment I, experiment II, etc., in Dr. Dingwall's presence, and initialled by us both. All this was to guard against the possibility that, in the event of the control yielding a null result, as I confidently and rightly expected that it would, someone might suggest that I had arranged the pseudo-originals so as to secure it. As a matter of fact, it would be extraordinarily difficult to devise a better method of arrangement from this point of view than the shuffling procedure actually adopted, as a moment's reflection should show, but there seemed no harm in adopting a full precautionary ritual.

The pseudo-originals were then re-arranged in the alphabetical order of their titles, just as had been done with the true originals submitted to Mr. Hindson, and were sent to Mr. Saltmarsh in this order. I also sent him, by instalments, the same 2,194 drawings which Mr. Hindson had marked against the real originals; that is to say all the drawings of experiments I to IVB.

As regards instructions, these were substantially identical with those given to Mr. Hindson; and the reservation 'substantially' is made only because, as reference to Appendix IV will show, it was necessary to alter or delete a few words here and there, in the 'Guide to Scoring Hits', which would otherwise have indicated that it referred to originals other than those submitted.

Apart from these trifling alterations, which were made in consultation with Dr. Dingwall, the Instructions issued were identically similar in the two cases; but the 'Notes on Originals' were naturally not sent, nor was it thought necessary to fabricate a corresponding document.

Needless to say, Mr. Saltmarsh was not told that he was engaged on a control or dummy experiment; on the contrary, by a piece of innocently ambiguous verbiage such as is frequently necessary in this class of work, I was at pains to convey the impression that he was repeating rather than paralleling the work of an earlier marker.

The scoring sheets, showing half and whole points awarded to the various originals in respect of the drawings, were sent directly to Dr. Dingwall, and the data necessary for working the 'All Entries' figures, corresponding to the most important main result obtained from Mr. Hindson's marking, were extracted by us together. The remainder of the work was done by myself alone; but the original scoring sheets and the relevant calculations are, of course, available for inspection if desired.

3. Results: As was expected, the results were, if not quite 'icily regular', at any rate 'splendidly null'. In Table XII below I give the values of D/σ and of P for what I may term the four main modes of scoring, together with the corresponding figures for Mr. Hindson's marking against the true originals, taken from Table 4 for ease of comparison.

Table XII

	M.T.H.		H.F.S.	
	(Experiments)		(Control)	
	D/σ	P	D/σ	P
Full points only	2.772	<.01	-.613	~.54
Half points only	2.924	<.01	-1.699	~.09
All Entries	4.166	<.000,1	-1.318	~.19
Score ³⁸	3.572	<.001	-.984	~.33

There is evidently nothing at all significant about the control figures. Even the modest value of .09

³⁸ This merely means that in this case the full and half points have been allowed their due relative weights, instead of one or other or both being treated merely as 'entries', as in the first three lines of the Table.

must be to some extent discounted as being one of several, even though not fully independent, results; conversely, the interdependence robs the fact of all four results showing the same sign of whatever slight interest it might otherwise possess. In ordinary language, the percipients have scored on these pseudo-originals fractionally fewer hits in the right places, as compared to the wrong, than we should expect, but only to a quite non-significant extent entirely compatible with the hypothesis that no factor other than chance was involved.

In spite of this, I thought it worth while to make a job of it by repeating the calculations of the 50 x 50 Table, at least so far as was necessary to obtain the analogue to the eleven point Displacement graph of Figure 1. The values obtained are shown below:

Table XIII

Group	(O-E)/ \sqrt{E}	
	Experiment	Control
-5	-1.031	.834
-4	-1.973	.152
-3	-1.678	.530
-2	.524	.606
-1	1.186	-.502
0	3.873	-.476
1	.760	-.274
2	-.922	-1.089
3	-1.262	-1.288
4	-2.083	1.197
5	-.542	1.427

These values are shown by the broken line in Figure I, which, by contrast with the full line, well indicates the relatively random nature of the result. Inspection of the Table suggests a fairly strong central depression followed by an upward twist at the positive end; but the effect is not significant. Testing in the usual way, I find an initial variance of 8.1210; reduction of variance associated with the first two parameters of the regression line, 3.8610 with mean 1.9305; residual variance 6.1905 with mean 0.7738; variance ratio 2.495 ($n_1 = 2$, $n_2 = 8$); $P \sim .05$ or about .14.

That is to say, the Displacement results, like the Main results, show nothing at all that is not quite easily explicable on the hypothesis of chance.

4. Comments: Some readers may be inclined to think that this process of control marking was no more

than a gratuitous waste of Mr. Saltmarsh's time and my own; that it did no more than exemplify the obvious by showing that if you carry out a random process properly you will obtain a random result; and that it no more illumines the questions at issue than making ten thousand throws with a substantially true die would illumine the question of whether another was loaded. With such critics I have the liveliest sympathy, for these were precisely my own reactions when the project was first mooted. But I am now inclined to think that these reactions, though natural, were wrong, and that the procedure, though emphatically a psychological luxury rather than a methodological necessity, has a rational justification and a positive value greater than meet the superficial eye. Here are my reasons.

I implied just now that it would be waste of time to do a long series of throws with a substantially true die as a check on one which had shown significant bias. That is true, but it is only true because we know a great deal from practical experience about the behaviour of dice; it would emphatically not be true if no one had ever thrown a die before, and if all we had to go upon were an 'intuitive' conviction that because the thing was a cube it 'must', *ipso facto*, behave in a particular way; this would be just as rash, in principle, as to conclude that because a steel needle is symmetrically suspended in a mechanical sense it 'must' be indifferent to its orientation, which is false if it happens to be magnetised.

Now the application of statistical methods (i.e., of probability theory) to any experimental data is no more and no less than the transference to one type of empirical situation of systematized knowledge gained originally from experience of other empirical situations; for there are no such things, as was once erroneously supposed, as god-given *a priori* probabilities. And if the situation to which the methods are applied is not strictly analogous in its relevant features to those on which our theory was based, the transfer may be illegitimate, things may go wrong, and invalid conclusions be drawn. In most cases we are on reasonably safe ground, for common experience tells us that the behaviour of, for example, properly made cards or dice, approximates sufficiently closely for all practical purposes to that of the idealised versions usually postulated by theory. But it is not the less important to realise, as a matter of principle, that the assumptions we may then, quite legitimately, proceed to make are ultimately based on

practical experience and reasoning therefrom by analogy, not on any kind of divine right. Usually, of course, our 'intuitive' judgments, to the effect that in the absence of some systematic cause one thing is as likely to happen as another, will be reliable; at any rate in so far as they are expressions of properly assimilated if previously unformulated experience. But it would clearly be imprudent to rely on them too implicitly, or to pursue the chain of analogical reasoning too far; consequently, there is nothing to be lost and a certain assurance rightly to be gained by making fresh contact every now and then with the real world of empirically observable facts.

In this particular case, I have no doubt at all that we were completely within our logical rights in assuming that, in the absence of any systematic cause, the distribution of marks by an unbiassed judge would be in accordance with the terms of the postulates underlying the statistical treatment, or that it would be possible to justify this assumption with virtual rigidity. But I am also pretty sure that no one has hitherto taken the trouble to mark two thousand odd drawings against fifty thoroughly randomised and arbitrarily grouped pseudo-originals; so that the control marking has certainly established a much closer link between empirical experience and the assumptions of the experiment than could otherwise have been held to exist.

It seems to me, therefore, that although the work may well be deemed to have approximated to supererogation, it cannot possibly be considered to have been waste of time.

bou
kin
On
to
nob
val
pro
acc
of
sho
son
whi
to
hop
pre
eva
is
mer
oth
to
fi
adv
the
out
thi
But
the
pe
mai

di
fe
wi
ex
a
se
or
co
nu
fl
in

Section VI

ANTICIPATION OF CRITICISM

1. General: It is very meet, right and our bounden duty that we should subject experiments of this kind to the utmost rigour of all reasonable criticism. On the other hand, life is short and the amount of work to be done in the subject prodigious, so that it is to nobody's advantage to replicate controversy or to waste valuable time in arguing about matters which ought not, properly speaking, to come into dispute at all. I have accordingly thought it expedient to deal here with some of the types of criticism which private discussion has shown to be most likely to be raised, and to attempt some kind of logical analysis of the various ways in which an experiment of this general character might fail to satisfy scientific and evidential criteria. It is hoped that prophylactic treatment on these lines may prevent subsequent discussion from drifting into irrelevant backwaters. I need hardly say, I hope, that there is not, in my opinion, *any* valid criticism of a fundamental character which can be brought against the work - otherwise I should not be publishing it; but there seems to be quite a number of points which people find it difficult to grasp and these may better be dealt with in advance than in arrears. This is not intended to imply that the experiments could not have been better carried out; on the contrary, there are many details which I think I could now improve upon in the light of experience. But I do not believe there is any weakness either of theoretical method or practical procedure which is competent to invalidate, or even appreciably to weaken, the main conclusions reached.

Broadly speaking, we may expect to find critics divided into two main classes. First, there will be a few who, while admitting that there is nothing wrong with the experimental evidence, will contend that the existence of a cognitive relation between any person and a drawing the nature of which he cannot determine by any sensory process or by rational inference is impossible, or at least so antecedently improbable that no evidence could be considered convincing. Second, there will be a much larger class who will suggest that there is some flaw or weakness either in the general method adopted or in the manner in which it was carried out.

The first class are presumably of greater interest to the pathologist than they are ever likely to be to me; but for the benefit of the small and reasonable sub-class who are genuinely and semi-rationally troubled about antecedent probabilities two or three points are worth briefly noting which I do not think have hitherto been made.

First, it seems quite clear that all estimates of probability are either based on empirical experience or are worthless; hence there can be no such thing as a truly *a priori* probability, though there may be well-founded estimates of probability formed antecedent to the undertaking of any particular piece of investigation. Second, any empirical estimate of a probability must be deduced from the study of a sufficiently large aggregate of comparable instances, e.g., throws of a die or drawings from a pack, to enable a reasonably accurate estimate to be formed of the limiting ratio of the relative frequencies of the occurrences and non-occurrences of the event in question; it follows that the 'antecedent improbabilist' has the unenviable negative task of showing that the many indications of paranormal cognition which have from time to time been obtained are *not* in excess of what can reasonably be attributed to chance, fraud and the like; at the very best he must show (a) that so many cases are explicable in these terms that there is a strong probability that any new and similar case is thus explicable, and (b) that this case is similar. Third, if he declares that the occurrence of a faint but widely diffused power of paranormal cognition, such as I have found in the course of this work, is contrary to common experience, one can only reply that it is not; on the contrary, every second person one meets has a story suggestive of paranormal cognition in some form or other, either at first or second hand, while the whole literature, not merely of *Psychical Research*, but of *History*, *Anthropology* and *Biography*, is crammed with instances of varying degrees of authenticity. Fourth and finally, what would be contrary to common experience, to the point of incredibility, would be if no concatenation of circumstances had ever resulted in the kind and degree of ability I have found ever resulting in an incident sufficiently striking to be put on record.

In short, criticism on these lines, even in its most reasonable form, does not seem to me to have a leg to stand upon; for the whole of the antecedent and contextual evidence appears far more consonant with the

suppo
cogni
"twis
less
may l
desir
of th
anyt
conc

I ha
pany
to m
divi
with
the
cal
cept
the
obje
woul

'eri
sibl
of
ing

the
ate
dec
gar
ibl
whe
wor
H.S
if
som
of
tai
hav
ing
lea
sti
nur
on
hav
su

supposition that what we now call 'paranormal' modes of cognition constitute a fact in nature, often enough "twisted by knaves to make a trap for fools" but none the less real for that, while attempts to explain them away may be due to a wide range of motives - from a laudable desire to discourage superstition, through a vague dread of their ill-deduced implications, to a petty dislike of anything not smugly subsumable under the critic's own conception of how nature ought to behave.

Turning to the second class of critics, with whom I have every sympathy (for have I not been of their company where much other work has been concerned?) it seems to me that not-obviously-stupid criticism can again be divided into two main types, of which one is concerned with what I will call Real Possibilities of Error, while the other arises sole from Misapprehensions of the logical basis of the work. The possibility of sensory perception would be a perfect example of the first, while the delusion that the use of originals depicting common objects might possibly foster a spurious positive result would typically represent the second.

2. Real Possibilities of Error: Using the term 'error' in a somewhat wider sense than usual, these possibilities may conveniently be dealt with under the heads of fraud, misreporting, sensory leakage, inference, marking and statistical treatment.

As regards fraud, it is entirely proper that, in the present state of the art and in view of the unfortunate history of much of our subject, the possibility of deception should be seriously considered, without any regard at all to personal feelings, wherever it seems plausibly applicable; but I do not think that this is a case where it is. It certainly would have been if I had been working with professional music-hall telepathists, or if H.S.C. and I had set up as a pair of modern Zancigs, or if I had relied wholly or mainly on the performance of some one especially gifted percipient, or even on those of a very few; whereas none of these circumstances obtain. Apart from this, I very much doubt whether I could have put through so monumental a deception without arousing the suspicions of some at least of my various colleagues and consultants; and even if I could, it would still have been necessary either to conspire with a large number of my percipients, or else to fabricate drawings on a fairly extensive scale. The last would probably have been the best plan, and a skilled forger could presumably manage it. On the other hand, all drawings,

markings, tables and calculations are, of course, available for inspection by any interested and responsible person.

But considerations of this kind do not constitute the proper rebuttal of any such suggestions. As I indicated at the very outset of this paper, the development of a repeatable experiment - the most urgent desideratum in the subject - has been one of my most important objectives throughout. I believe I have succeeded in doing so; consequently, I am in the position of being able to make the only completely satisfactory reply (even better than pistols for two and coffee for one before breakfast), namely, "If you don't believe me, go and try for yourself".

Great stress has often been laid on the importance of predictability as the acid test of genuine phenomena. No sane person, of course, would agree with the pronouncement "Only when conditions can be so controlled that, e.g., a teacher can announce before-hand that, on such a day, hour, and place he will demonstrate these things" *viz*, telepathy or clairvoyance, "can they or will they be accepted by any sound scientific mind", but the general principle is sound enough.

I accordingly venture to predict to the following extent: Take a group of about thirty or more university students of mixed sexes and perform a ten day experiment as described in my account of experiment I: be scrupulously careful to exclude all sensory clues: Select the subjects of originals from some much larger list of reasonably common objects, avoiding ambiguity, by any substantially random method: repeat the experiment, preferably with different percipients, two or three months later, using different originals: have the drawings scored 'blind' against the twenty shuffled originals, by groups as described above, using a rather generous standard of scoring: if necessary repeat the experiment a third time. I am fairly confident that at least promising results will be obtained, though I would not care to guarantee significance; and when I say 'promising' I mean sufficiently good to justify any reasonable person in continuing the work. Naturally, a dozen small points come to mind, which can be gathered from the details given in the preceding pages; but there is one possibly important reservation which I think it fair to make. Inasmuch as the process may be and probably is of a 'telepathic' character it might very well be upset by a hostile attitude on the part of the experimenter, so I

think it legitimate to insist that such an experiment should be undertaken with a reasonable amount of good will. That is to say, I doubt whether an experimenter would be likely to obtain good results if he started with the attitude "This man is an adjectival charlatan, but I suppose it's my duty to show that there is nothing in it". Subject to this reservation, I fail to see why anyone should not repeat the experiments with success, substantially at any time and in any place and with any percipients he pleases - unless, of course, I happen to be an exceptionally good telepathic agent, which there is no reason for supposing and would be a singularly harsh piece of misfortune.

The possibility of misreporting, though one of the most important in some branches of our subject, is hardly one that is likely to be applicable here. There is no question of whether I did or did not hold hands in the dark, or of the squirmings of ectoplasmic pseudopodia in a dim light; nor even, to come a little nearer home, of making statements of the type "The percipient was seated so that she could not see the cards", which would need very careful expansion before acceptance. I think I am right in saying that all evidentially important statements made are of such a nature that they must be either lies or substantially accurate; but I shall be very pleased to elucidate any ambiguity which has escaped my notice.

The likelihood of percipients having obtained knowledge of the drawings by sensory means³⁹ may, I believe, be dismissed with even greater brevity. Not even Sam Weller's million magnifying power telescopes would enable people to see through brick walls, hundreds of intervening objects, or even curtains - let alone three thousand miles of bulging earth between here and North Carolina. I don't know whether anyone will be so rash as to try to attack on this front; but I'm afraid he will meet with but little success if he does.

Similarly, I find it hard to imagine that anyone will suggest that the percipients were able to infer by any rational process what the nature of the originals was likely to be, either in respect of any particular occasion or of any experiment taken as a whole. If a percipient had known (as none did) the precise method

39. By 'sensory means' I refer to any physiological process involving the stimulation of a peripheral receptor and the transmission of an impulse along an afferent nerve fibre to the central nervous system.

of selection we were using in the first, third and fourth experiments, and what dictionary was employed, he might have successfully inferred that originals as a whole were somewhat more likely to be drawn from one section of the dictionary than from another; but this would not have helped him to decide which page was most likely to be used on a given occasion or in a given experiment.

Personally, if I were coerced into leading a forlorn hope against an impregnable position of this kind I should concentrate on the marking, and should try to make out that the instructions and notes provided, wittingly or unwittingly, clues and guidances calculated to ensure an undue assignment of 'winning' points. It is to forestall attempts of this kind that I have published the relevant material in full in Appendix IV, and I do not think that any plausible case can be made out on these lines⁴⁰. In the very unlikely event of this occurring, we can always fall back on the results of the Arbitrated marking, or on the 314 drawings dealt with in the fifth column of Table VI on p. 75 above. In principle, I should be only too pleased to see these 314 drawings scored blind by any reasonable panel of unbiased (i.e., ignorant) judges; though in practice I should suggest that their time would be far better spent in repeating the experiment.⁴¹

We are left with the question of statistical treatment, and here, I think, I am on ground no less firm than elsewhere. The procedure whereby we calculate the expected number of winners, explained on p. 79 above, is no more than the simplest rule-of-three arithmetic; the method of obtaining the variance is due to Mr. W. L. Stevens and has not been challenged; or alternatively, where the numbers are as large as they are here, we may use a simple χ^2 method familiar nowadays to almost every scientific worker. The only point I can imagine being raised in this connection is the very silly and trivial one that my first method of assessment did not work, so that I was forced to shift the basis of the enquiry from

40. In case anyone wishes to make such an attempt, I hasten to point out that it must be completely specific; that is to say, expressions of opinion are not enough; it must be explicitly stated which Instruction or Note is thought likely to generate a spurious positive result, and how.
41. I have said "these 314 drawings", but strictly it should be the whole 2,194 unless the critic concerned is prepared to accept my assurances as to which the 314 were, and on other relevant points.

particular occasions to experiments as wholes; on this ground, in a narrowly literal sense, I might be accused of having 'selected my test of significance after I had seen the data'. The answer to this is "So then what? Do you suggest that the results are due to chance, or are you merely indulging a taste for dogmatic psittacism?" However, we need not worry about this unless someone is unwise enough to start it.

3. Misapprehensions: As will have been gathered from the foregoing I feel I have nothing to fear from attacks directed on legitimate objectives, that is to say on any feature of the work which is, in fact, relevant to its outcome. If there was no *mala fides* or gross misreporting, if the percipients could not obtain knowledge of the drawings by sensory means or by rational inference, if the marking was truly unbiassed, and if the statistical treatment was correct, then the experiments stand and the results must either be attributed to some mode of cognition not covered by the above, or to chance: and chance alone would yield such results only once, on the average, in some tens of thousands of such investigations.

On the other hand, I am very frightened indeed of the much more dangerous type of critic who does not (or sometimes will not) grasp fully what has been going on, but none the less avers, with much grave and pseudo-judgmental head-wagging, that he is "not altogether satisfied" about something or other - which usually he is incapable of stating clearly. To describe this kind of thing as 'irresponsible' is to carry courtesy to the point of fulsomeness; for it is not, in other contexts, considered honest to make disparaging allegations without being able at least to define, even if not to substantiate them.

A typical example is to be found among those whose resistance to accepting the conclusions reached (plain implications of the work) is expressed by a melancholy mewing to the effect that "it may be chance". Of course they may be chance; anything may be chance; the behaviors of such persons themselves may be chance - in fact this is often the most charitable interpretation of it. But some things are very much more likely to be due to chance than others, and events of the 'once in ten thousand trials' order are not strong candidates for the former class. This kind of thing is exasperating, but I believe that behind it there is often a genuine bewilderment which deserves a few lines of consideration. It

arises, I believe, from a failure, or refusal, to understand the kind of thing we are trying to do and the object of the tests of significance we apply. Let me try to explain.

Common experience tells us, before we start any experiments of this kind, that innumerable and unspecified factors of varying potency will be at work in the percipients' minds and may cause them to draw anything whatever, including some of the objects depicted in the originals we use. We may accordingly expect with some confidence that we shall obtain a certain number of apparent 'successes' (i.e., resemblances between drawings and originals) which will actually be due to these factors and not to any true cognitive relation between percipient and original. If we do not fully realise this we are very liable to be led by enthusiasm into overrating what happens and attributing to 'paranormal' cognition effects which are really due to these other factors. The object, and strictly the only object, of a test of significance is to tell us whether this is likely to have happened - and about how likely. When the level of significance is found to be high, as in these experiments, we conclude that the effects observed are not likely - indeed, are very unlikely - to be due to such causes; though, of course, it is always open to anyone to contend that, in his opinion, they are even less likely to be due to any alternative cause. This last, however, is of no importance; it would be of interest only if we were attempting to 'prove' the occurrence of the phenomena once and for all by means of a single experiment or group of experiments. But this is not, or should not be, the case. Rare and obscure phenomena are not 'proved' in this sort of way; they gradually become established and accepted through familiarity and through a gradual elucidation of the conditions of their occurrence and the way in which they work. If experiments of the kind here described are never successfully repeated, if we never succeed in discovering the 'laws' concerned, then the critical historian of the future will be perfectly correct in writing them off as some kind of a *lusus naturae*. But in the meantime we are perfectly entitled to say that the likelihood of their outcome being a chance effect is so remote that no one need fear wasting his time in pushing the investigation further.

Of the more specific misapprehensions that I have encountered there are two which seem to me to merit fairly detailed discussion here. Of these, the first is concerned with the desirability of employing a plurality of

judges, the second with the question of whether the result can be swayed in one direction or the other by the use of common or rare originals.

The first can be made to sound plausible, if it is put in some such form as "Surely you will not allow results of such importance to rest on the opinion of a single judge?"; but none the less it arises solely from confusion of thought as to the issues involved. It is true that I am relying, and am perfectly content to rely, on the markings of a single judge, but I am not relying at all on his opinion in any sense that would make this type of criticism valid. Yet, in another sense, I am relying on his opinions to such an extent that if they were usually ill-founded no significant effect would be likely to emerge. If we can clear up this apparent paradox, we shall not only settle the particular point at issue but a number of others also of a somewhat similar nature. We can put it very briefly by saying that the kind of opinion on which we do rely, and which is of great importance, is a judgment to the effect that "This drawing is sufficiently like that original to be worth giving it a mark (or half a mark)"; and that the kind of opinion on which we do not rely is a judgment to the effect that "This drawing is a Winner (i.e., a hit in the right place)". But this is unduly cryptic and calls for careful elucidation.

I think I can best clarify the issues, which are exceedingly important, by considering two apparently contradictory statements both of which are true. The first may be put in the form "It makes no difference what the judge does", and the second as "A good judge makes all the difference". The catch here is this, that the first statement is true and the second false if there is no real cognitive effect, but only chance, at work; while the second is true and the first false, if there is. The point will become clearer if we expand the first form "If there is no real effect, a judge who does not know the answer⁴² cannot generate a spurious one by any kind of wishful or misguided marking", and the second to the form "If there is a real effect, an intelligent and discriminating judge will bring it out better than one who is neither".

42. I use the phrase 'know the answer' as a convenient condensation of 'know which originals belong to which experiment', or, more generally, 'know what allocations or markings will favour a positive result.

Even these expanded forms need further consideration, and I think the best way to ensure it is to invite the reader to imagine himself in the position of a judge intensely anxious to secure (or prevent) the obtaining of a spurious positive result. How is he going to set about it? He is given one batch of drawings which he knows to be those of the percipients of experiment I, another of experiment II, and so on: he is also given fifty originals arranged in alphabetical order - apple, beaver, cat, cormorant, distaff, etc., down to xylophone, yoke and zeppelin, or whatever they may be. But (and this is vital) he does NOT know which of these originals were used in experiment I, which in experiment II, etc., nor has he any clue to guide him. He starts work on the first batch of drawings and soon comes across, we will suppose, a number of rather indeterminate birds. Is he to reckon these as hits on cormorant, or as half hits, or as not worth a mark at all? Which policy will best serve his ends? If cormorant is one of the originals of experiment I, then reckoning all these as full hits will tend to produce a positive result; but if it happens not to be, then the reverse. But this is exactly what he does not know, so that he will be quite unable to decide whether generosity or strictness will pay him best. And the same applies to every original of the fifty and to every batch of drawings. Thus it is literally impossible for any degree of fancifulness, of prejudice in either direction, or of eccentricity on the part of the judge to generate anything but a chance effect if there is no non-chance relationship between the drawings and the originals.

The same applies though not quite so obviously to capriciousness or lack of consistency. I have heard it asked "Suppose it just happens that⁴³ the judge is feeling generous when he is considering a drawing of experiment I and looking at an original of experiment I, but strict when he is considering a drawing of experiment I and looking at an original of experiment II?". The answer is "Suppose it 'just happens' the other way found, the effect will be reversed. Have you any reason for supposing that the one situation will arise appreciably more often than the other (if so, state it); and are not these fluctuations of the standard exactly what we refer to, among other things, when we use the word 'chance'?" It goes without saying that fluctuations of this kind will frequently occur and that they may

43. This is a common involuntary trick whereby the notion of chance is smuggled incognito into a sentence ostensibly not containing it.

often be the deciding factor in cases of doubt; but unless we postulate relevant knowledge on the part of the judge (excluded by hypothesis), or suppose him endowed with paranormal powers (which would be begging the question) they can have no more systematic influence on the outcome than would decisions obtained by tossing a half-penny or throwing a die.

Now turn to the other side of the question, and suppose that there is a real effect to which we are anxious to do justice. In these circumstances the intelligence and ability of the judge is likely to be of the very greatest importance. The material, on this hypothesis, will be made up of three types of drawing; there will be some which have no resemblance or connection with the originals at all; there will be some which show resemblances or associations determined by chance alone; and there will be a certain proportion of which the nature or form has been to some greater or less extent influenced by the cognitive process postulated. Now, unless we can hit on some infallible sign whereby we can distinguish lucky shots from genuine cognitions, we shall never be able to eliminate the second class of resemblances, which will always be present to dilute the real effect in some degree. The ideal judge would therefore be he who would reject all specimens of the first class, but successfully detect and correctly mark all members of the third; this would bring out to the fullest possible advantage whatever real effect the material might contain. At the other extreme, even if the material consisted wholly of genuine cognitions (class three) it could scarcely survive marking at the hands of a blind imbecile, who would presumably allot points in a completely haphazard manner having no reference to the resemblances actually observable.

It is in this sense, and in this only, that what the judge does is of importance; just as it is in the other sense, and in that only, that what he does can make no difference. If this is not now clear, I fear I must despair of ever making it so, and must be content to wait for the spontaneous growth of comprehension in the minds of those concerned.

It is worth noting in passing, however, that the evidence suggests that Mr. Hindson was a singularly 'good' judge in the sense just discussed; for it will be remembered that even those of his half-points which the arbitrator rejected as implausible resemblances showed a just significant result when treated alone. This is

very remarkable and strongly suggests that Mr. Hindson has a definite 'flair' for detecting the mind of remote resemblance which results from a very imperfect cognition. I shall always regret that I asked him, as recorded on p. 61, to raise his standard after scoring the drawings of the first experiment; if I had not done so, we might well have obtained much more information about what kind of distortion takes place and what the limit of 'genuine' resemblance is.

This brings us back to the question of the plurality of judges from a fresh angle. I greatly fear that, in spite of all that I have said above, I shall be urged to have the marking repeated by an independent judge or judges with a view to strengthening the validity of the results obtained; and if I decline to do so (as I shall) I am likely to be told that I am afraid of such a re-marking giving a null, or much less significant, result. This, of course, is rubbish, for a much less significant result by another judge would no more invalidate that already obtained than a much more significant result would strengthen it; it would only show the second judge to be inferior (in the sense discussed above) to Mr. Hindson, as *vice versa*. Chance, as an explanation of the effects observed is already out of court, except from the apriorist standpoint dealt with earlier, and no amount of replicate marking will bring it in again. Nothing is to be gained by wasting the time of judges in doing what we know to be unnecessary, when they might be doing something useful; or in bowing oneself in the house of illogicality when one's logical position is impregnable.

The position would be very different, and re-marking desirable to the point of necessity, if it could be shown that the instructions, etc., gave biasing guidance to the judge, or if it could be plausibly maintained, in face of his testimony and my own, that Mr. Hindson had, in fact, any notion of which originals were used in which experiment, or even any clues by which he might have known. Apart from this, re-marking in this case, or plurality of marking in general, will only throw light on the relative merits of the judges, not on the validity of results⁴⁴. Such questions as to extent to which

44. To cover a point which might be raised: if we were to submit to a judge, under the conditions described, material which we believed to embody a significant real effect, and were to obtain a null result, we should be perfectly entitled to re-submit to another or others in the hope that the first was incompetent -

judges differ, and why, or of what principles of judging are best calculated to bring out real effects in varying types of material, are of no small intrinsic interest, and will undoubtedly demand investigation in due course. But they have nothing whatever to do with whether the results of these or similar experiments are valid, and it is wholly illegitimate to contend that they have.

A great deal of what has just been said is applicable, *mutatis mutandis*, to the cognate suggestion that the outcome of the experiment may be influenced by the kind of originals used. If we substitute 'ignorance of percipients' for 'ignorance of the judge', we shall find a close parallel between the two situations, the second of which accordingly does not need detailed discussion here. Just as, in one sense, it makes no difference what the judge does, so, in the corresponding sense, it makes no difference what originals are used; and just as, in the other sense, some judges are better than others, so, in the corresponding sense, some originals are likely to be better than others. And just as it is impossible to show how a judge who does not know the answer can fabricate a spurious result by any process of mis-marking, so it is impossible to show how an experimenter can fabricate a spurious result, through the use of one sort of original rather than another, provided the percipients have no possibility of telling, by sensory means or rational inference, what originals are likely to be used on which occasions. Finally, again as in the preceding case, the best way of convincing oneself that this is true is to try to work out a plan for generating a spurious effect, by the use of any type of originals whatever, subject to the restrictions that the original to be used on any occasion is selected by a random method, and that the percipients are given no clue as to what it is.

4. Summing-up: I have no desire to appoint myself, dictator-like, as judge in my own cause, though it seems fair enough to insist, as I have been doing, that criticism should be directed towards points which might, in principle, be vulnerable rather than against those which could never be. It is true that my most anxious scrutiny has failed hitherto to find any source of systematic error in either the method or procedure adopted, and that I do not believe that there is one. But it is not by my opinion, or even by that of others however

(Footnote continued) always provided that we made due allowance, in the usual way, for the number of judges so employed making our final estimate of probability.

eminent, that the work must be judged as successful or otherwise; it is by the test of repeatability alone that it must ultimately stand or fall. If properly conducted attempts to repeat my results consistently fail to do so, no amount of argument on my part will off-set the fact; and if, in a fair proportion of such cases, similar results are obtained, argument will be superfluous.

So we may end this Section very much as we began it: I have tried to produce a repeatable experiment; I believe that I have succeeded; if, for whatever reason, you do distrust the conclusions reached - Try it yourselves.

thou
ciab
I wi
an e
in b
eral
foun

of d
thes
orig
were
250
draw
more
peri
bili
any
near
agai
clud
whic
whic
of t
is p
or
use
ati
any
woul
thi
res
sem
the

Section VII

SUMMARY AND CONCLUSION

This paper is already depressingly lengthy, though I doubt whether I could have shortened it appreciably without damaging omissions of important details. I will accordingly do no more by way of bringing it to an end than run over the essential features of the work in briefest summary and add a very few comments of general interest for which no appropriate place has been found elsewhere.

Five experiments⁴⁵ in the 'paranormal' cognition of distant drawings have been conducted. In each of these, ten different originals were used, and ten fresh originals were used in each experiment. All originals were selected by a substantially random method. About 250 percipients of both sexes, producing about 2,200 drawings, took part. Very few percipients, probably not more than ten or twelve, took part in more than one experiment. In no experiment was there the smallest possibility, humanly speaking, of any percipient obtaining any clue to the nature of any original by normal sensory means or by rational inference. The drawings were marked against the whole fifty originals by a judge who had no clue or information as to which originals were used in which experiment. A total of 1,209 drawings were found which were judged to be sufficiently like one or other of the originals to deserve mention. From the data it is possible to calculate how many of these resemblances or 'hits' would be 'winners', i.e., hits on originals used in their own experiment, if chance alone were operative, and how often this value would be exceeded by any given amount. It is found that the excess is such as would be equalled or surpassed only about once in some thirtythousand such investigations if chance alone were responsible. In other words, percipients' drawings resemble the originals (considered as a group) at which they are aimed more closely than they resemble originals

45. I say 'five' here because it will be more convenient henceforward to think of the experiments here described as experiments I to V, rather than I to IVB; otherwise we may find ourselves getting confused at some later stage by having, say, seven experiments with eighty originals. The next experiment will be numbered VI.

at which they were not aiming to an extent which cannot plausibly be attributed to chance.

Examination of the data from another point of view shows that these resemblances do not occur exclusively, or even most often, at the same time as the display of the original concerned.

But there is a fairly regular tendency for drawings which resemble a given original to occur relatively more frequently on occasions which are near that on which it was displayed than on those which are more remote. This effect is observable to a significant extent in both directions.

The main conclusions indicated by the facts are, first that there is a real cognitive relation of some kind (direct or indirect) between percipients and originals, second that this may be either of precognitive or retrocognitive form.

The above summary, if somewhat arid, appears to cover adequately the essential facts and inferences therefrom, but I may be permitted a few comments of a slightly more speculative nature.

I suppose most readers will want to ask "What kind of a process do you think is involved?". I do not think the time is ripe for saying, or even suggesting, what the process is; but I am prepared to record a few quite tentative impressions which suggest fairly strongly what it is *not*. First and foremost, I am as confident as I could be without special experimentation that it is in no sense a matter of the percipient copying or in any way 'seeing' the drawing. In contrast to M. Warcollier's recently published views, I have the strong impression that it is the 'idea' rather than the form that is cognised. For example, it is as if the percipient were told 'Draw a Hand' rather than 'Copy this Hand', for we get left hands and right hands, open hands and closed hands, apparently quite indiscriminately, which we surely would not do if it were a matter of copying something seen. Very seldom indeed have I received the impression that it is the uninterpreted form itself that has 'got across', and even then I fancy that it has been the 'idea of the form' rather than the lines themselves that have been concerned. In fact, on the strict understanding that this is entirely conjectural and 'for purposes of entertainment only', I increasingly incline to the view that the lines on the paper have nothing to do with it at all,

except perhaps as a focus for the thought or attention of the experimenter.

The astute reader will correctly deduce from this that my impression at present is one of a 'telepathic' rather than a 'clairvoyant' phenomenon. This is true as far as it goes, but only subject to the very important reservation that I am not at all sure that the current conceptions of either the one or the other will necessarily fit the facts. The truth of the matter is that we do not know what kind of a process is involved, and it would be a mistake to handicap ourselves by trying to tie it up prematurely with any preconceived notions.

It seems pretty clear that the percipients show a cognitive relation to the originals which cannot be attributed to any 'normal' cause; and I have no doubt at all that the experimenter's mind (H.S.C.'s or mine) plays an important part in establishing that relation. Indeed, certain work now in progress suggests on inspection (I will not say more) that the mind of a third part, neither experimenter nor percipient but still connected with the experiments in certain respects, may significantly influence the results; and the bare suspicion that such an effect may be exercised in a manner determinable by experiment should make us very cautious about formulating even the most tentative explanations.

I will end by drawing attention to the fact that the effect found, though highly significant, is intrinsically very faint. On the basis of the M.T.H. All Entries data, which I have so extensively used, a total of rather more than 2,000 drawings yields a crop of just over fifty hits above chance expectation; so we may reasonably speak of a $2\frac{1}{2}\%$ effect. Presumably this is an under-estimate, partly because I artificially restricted Mr. Hindson's natural judgment, partly because it may fairly be supposed that a perfect marker would have done a little better. But I should be inclined to doubt whether any marking however perfect would raise the figure much above 5%.

This, to my mind, is a very gratifying order of result. It is admittedly small, but so is the amount of radium in pitchblende which no one on that account denies is important and interesting. On the other hand, if I had found an effect of some ten times the size, I should have suspected the presence of many large flies in the ointment. As it is, it seems to me that what I have found (especially when we add the complications due to

displacement) is eminently compatible with both sides of common experience - with the knowledge that on the whole people very seldom show signs of paranormal cognition, and with the knowledge that none the less they occasionally do. Finally, the fact that, so far as I can judge by inspection, the ability concerned is pretty widely distributed, or at least not concentrated to any startling degree among a very few specially gifted persons, suggests that it is likely to prove an attribute common to all humanity, with nothing alarmingly magical about it; so that perhaps the adjective 'paranormal' is something of a misnomer after all.

ra
is
sh
sh
or
in
th
me

a:
a:

p
o
s
i

a
h
s
g
P
r

e
d
p
f
t
t

v

APPENDIX I

Copy of Instructions printed on the outer covers
of Percipients' Books used in Experiment I.

INSTRUCTIONS

1. Certain drawings (none of which will be elaborate) will be displayed in the room of which a photograph is provided, and in the position of the sheet of paper shown pinned to the bookcase. Only one drawing will be shown at a time and a different drawing will be shown on each of ten successive nights. Each drawing will be in position from 7:00 p.m. until 9:30 a.m., starting on the evening of Wednesday, February 1st, and ending on the morning of Saturday, February 11th.

You are asked to try to reproduce these as well as you can. Attempts to reproduce a picture may be made at any convenient time during the period of its exposure.

In order to be of use for the purpose of the experiment, it is essential that an attempt should be made on each of the ten nights. Please make an attempt of some kind, whether or not you feel that you are succeeding.

2. Drawings should not be altered or 'finished', and it does not matter if you cannot draw well. If you have several impressions, however crude or vague, make several drawings, *on the same sheet*, writing 'Best' against the one you think most likely to be successful. Please make *all* drawings on the sheet provided, not on rough paper first.

3. Write your full name and the time of making each attempt in the spaces provided on each sheet. Indicate under 'Visual Imagery' whether you had a clear picture 'in your mind's eye'. Write + or 0 after 'Confidence' according to whether you did or did not feel that you were getting a genuine impression of the picture.

4. Be careful to use the sheets in the order which they occupy in the book.

5. Do NOT write the date or any serial number on the sheet, or make any mark which would enable the order of the sheets to be ascertained if they were removed from the book. Do NOT make any Note which would give a clue to the date or ordinal position of the sheet.

6. Please insert the card provided between the sheet you are using and the next, so as to prevent the possibility of marks made on one sheet showing on the other.

7. The Space for Notes and Impressions is not intended for elaborate introspections, but for verbal recording of any 'ideas' which seem relevant but are too difficult to draw. There is no need to make any entry here unless you wish.

8. It is desirable that the reproductions by the various participants should be quite independent of each other; so please do the experiment by yourself and do not discuss your impressions until after the books have been handed in.

9. Return the sheets, in their cover, to Dr. Thouless when finished. On no account detach the sheets from the book or use them in the wrong order.

APPENDIX II

Results of the Method of Forced Matching

The Method of Forced Matching and its outcome have been fully discussed in general terms in the text: the subjoining table shows the actual scores obtained in the first three experiments by various judges. H.S.C. and I did the first and third experiments jointly, but with H.S.C. predominating; she necessarily worked alone in the case of experiment II, because I knew the order of both originals and drawings. For this experiment I give also the results obtained by the percipients themselves, which are fractionally worse than H.S.C.'s. As a matter of interest I add the figures for an independent matching of the thirty-seven sets of experiment I, which was very kindly undertaken by Mr. O. L. Zangwill.

Table XIV

Showing Numbers of Successes Obtained
by Forced Matching

Correct Matchings	Number of Sets Scoring								Total Percepts	Total Score	Mean Score
	0	1	2	3	4	5	6	7			
A. Matched by H.S.C. and W.W.C.											
Expt. I	14	12	6	4	.	.	1	.	37	42	1.135
" II	5	11	2	2	20	21	1.050
" III	5	2	2	2	11	12	1.091
Total	24	25	10	8	.	.	1	.	68	75	1.103
B. Matched by Mr. O. L. Zangwill											
Expt. I	12	16	8	1	37	37	.946
C. Matched by Percipients											
Expt. II	7	9	4	20	17	.850

The Table should be read as follows: When the thirty-seven sets of drawings constituting experiment I, were matched, set by set, against the originals by H.S.C. and myself, and the number of correct matchings ascertained as described in the text, it was found that in fourteen cases we had failed to match any drawing successfully against the original used on the corresponding occasion; in twelve cases we made one correct assignment, in six cases we made two, in four cases three, and in one case six.

Since the expected number of successes is in all cases equal to the number of percipients, there is evidently nothing of interest here except the solitary score of six in the first line, which is fully dealt with in the text.

APPENDIX III

Application of the Method of Decimal Scoring

1. Scoring of Experiments I, II, and III.

This was done by Mr. and Mrs. Oliver Gatty in two batches. The first consisted of eleven sets of experiment III mixed with eleven sets (nos. 1 - 10 and 12) taken from experiment I; the second included the remaining twenty-six sets of experiment I and the twenty sets of experiment II. Since I am relying for my conclusions

solely on the results of the third method of assessment, I need not detail here the innocent subterfuges I adopted to prevent the scorers unwittingly applying processes of rational inference to the work. I am, however, satisfied and the internal evidence suggested, that these were sufficient even if they may not have been necessary.

When the points were separated and added in the way illustrated in the text, the results were found to be:

a) Experiments I and III

	Judged to resemble		
	I	III	Total
Sets of Experiment I	10.0	1.0	11.0
" " " III	6.5	4.5	11.0
Total	16.5	5.5	22.0

Using the general matching formulae (p. 57 q.v.) this gives

E 11.0; O 14.5; D 3.5; σ^2 4.321; σ 2.079 which is not intrinsically significant, for D/σ 1.684 with $P \sim .09$.

b) Experiments I and II:

	Judged to resemble		
	I	II	Total
Sets of Experiment I	10.5	15.5	26.0
" " " II	5.0	15.0	20.0
Total	15.5	30.5	46.0

(N.B. A set is counted as half right and half wrong when the points awarded to the two series of originals are equal.)

This leads to

E 22.02; O 25.50; D 3.48; σ^2 10.327

For the two batches taken together we have

E 33.02; O 40.00; D 6.98; σ^2 14.648

whence we find D/σ 1.824 with $P \sim .07$, which is still not significant.

2. Scoring of eighty sets from Experiment IBA and IVB:

As explained in the text, the procedure of this experiment was specially devised to enable the somewhat

onerous task of scoring the three hundred odd sets which it was hoped to obtain to be undertaken by myself, and special precautions were adopted, as there described, to ensure that I could have no normal knowledge of whether any particular set was aimed at the A or B originals. As it turned out, however, many fewer sets were obtained than had been expected, while of those that were sent in a considerable number showed internal evidence which, as it happened, would have enabled me to judge that they were 'A' sets apart from any intrinsic resemblances of the drawings. However, when these had been weeded out by Prof. Broad, there remained eighty sets on which I could safely work. I again abstain from detailed discussion of the points involved on the ground that, although these results are perfectly reliable, I am not depending on them for my conclusions, and therefore need not establish every point meticulously here.

These eighty sets were given me in two groups, the first consisting of twenty 'A' and twenty 'B' sets, the second of ten 'A' and thirty 'B'; but, by a not unfortunate misunderstanding, I received the impression that the composition of the second group was the same as that of the first. Thus any bias I may have had in dealing with it was liable to mislead rather than to help: actually, it could hardly do either, for knowledge of the relative numbers of A and B sets in the group could not enable me to decide which were which.

The results obtained were:

	Judged to resemble		
	A	B	Total
Sets of Experiment IVA	9.0	21.0	30.0
" " " IVB	13.0	37.0	50.0
Total	22.0	58.0	80.0

Applying the same methods, we obtain

$$E \ 44.50; \ O \ 46.00; \ D \ 1.50; \ \sigma^2 \ 15.142$$

which is very nearly as bad as it could be, for D/σ is only 0.386 with P as large as .7. The result must be incorporated, but it is only of interest as showing the difficulty of separating the A and B portions of experiment IV, which we shall find persisting even with the much more successful third method of assessment.

3. Scoring of sixteen sets from the Individual Experiments:

In each of these eight experiments, which I do not propose to describe in detail at present, a selected

percipient was tested, together with either H.S.C. (first six experiments) or W.W.C. (last two). All H.S.C.'s drawings were scored by Dr. E. J. Dingwall, the six sets of the selected percipients in the first six experiments by Mr. B. E. Parr (to whom I should like to take this opportunity of expressing my gratitude), and the remainder by myself. All judges were kept in appropriate ignorance either as to which originals were which or as to which drawings were aimed at them, or both, as the circumstances required and permitted.

The scoring was done by pairs of experiments and need not be given in detail here. Combining the returns from the two judges

$$E \ 8.0; \ O \ 12.0; \ D \ 4.0; \ \sigma^2 \ 2.667$$

whence we find D/σ 2.449 with $P < .02$, which is significant.

Adding these quantities to those obtained in 1 and 2 above gives D 12.48; σ^2 32.456; D/σ 2.192; $P < .05$ which is just decently significant but no more. The substantially null results obtained from the eighty sets of experiment IV evidently vitiate an otherwise not unpromising method, but the final outcome is poor compared with that yielded by the more sensitive method used later.

APPENDIX IV

Copy of Instructions, etc., supplied to M. T. H.
for the application of the Method of Palpable Hits.

GUIDE TO SCORING HITS

1. The general idea is to compare every drawing with each of the (50) originals. If any drawing plainly represents the same object or activity as that depicted in one of the originals, a 'hit' is recorded; if not, not.

2. Scoring of hits should be done on the simplest commonsense principles.

The primary question to be asked in respect of any drawing is, "Does this drawing represent the same object or activity, or the same sort of object, etc., as one of the originals - or at least something extremely like it?". In other words, "Does the drawing indicate that the percipient had 'X' prominently in mind at some time while he was doing the drawing ('X' being the

subject of an original)?" If the answer to these is 'Yes', score a hit; if 'No', don't.

The second question, the answer to which may qualify the above, is "Does this drawing indicate that the percipient was thinking of the same kind of thing as was the agent when he produced the original?" Notes on what the agent actually had in mind when he drew the originals will be found in 'Notes on Originals'.

3. Pay no attention to the skill of drawing displayed. A badly drawn horse, say, is still a horse, provided it is clearly nothing else.

4. Pay as much attention to the written remarks as to the drawings; a mention of an X is just as good as a drawing of an X. One or two 'scorers' have missed very palpable hits here; please be careful.

5. Eschew far-fetched resemblances, 'puzzle-picture' methods, and all but the plainest and most universal associations, like the plague. Be very chary even of these last.

6. Half hits (but no other fractions) may be given in cases of doubt.

7. The greatest difficulty will probably be found in connection with composite drawings containing several elements. Sometimes it is easy enough; for example, if a drawing shows prominently both a tree and a horse, a hit should be recorded on tree and on horse (Jennet). But one is often in doubt as to whether the X depicted was 'prominently' in the mind of the percipient at the time of making the drawing (Cf. first question in 2 above).

It will probably be convenient to recognise four grades of 'prominence' as follows: 1. The X is the sole, or virtually sole, object depicted in the picture; 2. It is 'co-equal' with other objects; 3. It is 'secondary' but must none the less have required attention, or 'thinking of', for it to have been drawn; 4. It is purely 'incidental', i.e., jotted in, so to say, as a kind of 'trimming'.

Give hits for the first three categories, but not for the 4th.

8. In general, a part of an object may be taken as equivalent to the whole. E.g., It is unlikely that

anyone would draw a horse's head without having a whole horse more or less in mind; by contrast, someone might well draw a hand without having a whole man in mind, so a certain amount of common sense must be used in such cases.

9. Some objects, such as trees, boats, chairs, are very commonly drawn. In certain cases a hit must obviously be recorded, e.g., tree, but if the correspondence is doubtful caution should be exercised and points sparingly given; otherwise any real effect is liable to be masked or 'diluted' by points needlessly given to objects drawn only because they happen to be very familiar.

10. On the other hand, some originals are very 'difficult' and are seldom if ever drawn unmistakably, e.g., cross-stitch, embattled fess, stop-cock. Perhaps it would be legitimate to relax the usual standard to some extent in such cases; but this must be for the judgment of the marker.

11. General: Although a high standard should be maintained, so that, as a rule, there can be no reasonable doubt as to the correctness of the hits recorded (and only a half point given when there is doubt), it will be understood that we cannot rationally demand exact reproductions of the originals; drawings 'as good as can be expected' from a percipient who perhaps cannot draw well and may be supposed to have imperfectly cognised the original may be taken as acceptable.

In this connection it seems not unreasonable to suppose that whereas some subjects (e.g., anchor, buffalo, frog, hammer, prawn, violin, tree) will 'get across' in their entirety or not at all, others, which are built up of constituent parts - e.g., embattled fess (shield *plus* fess), flag (plain flag *plus* cross), Royal Standard (plain flag *plus* lion rampant), windmill (building *plus* sails), etc., might appear only partially; that is to say, one of the constituents might appear without the other. It may be worth while bearing this in mind in connection with awarding half points.

(Signed) W.W.C.

(Dated) 2.viii.39

PARANORMAL COGNITION OF PICTURES

Notes on first 50 Originals

The following notes are intended to indicate what was actually in the mind(s) of the agent(s) when the originals were drawn, and to draw the attention of scorers to certain doubtful points which have arisen. The latter are usually dealt with in the form of questions so as to minimise the risk of unduly influencing an otherwise independent scorer.

1. ANCESTOR: The idea in the agent's mind was that of an *old* man, hence the beard and the staff. Should a point or half point be given to any 'man' drawn regardless of whether he had a beard, is leaning on a staff, or shows other signs of age?
2. ANCHOR: Straightforward; no comments.
3. ARROW: Straightforward; as an example (which has not yet been observed) of 'the plainest and most universal associations' it is suggested that it would be legitimate to give at least a half point to a drawing of a bow, even if the arrow were not drawn.
4. BALANCE: Fairly straightforward; but what do you propose to do about (if any) steelyards, similar balanced mechanisms, see-saws, tight-rope walkers?
5. BAT: Drawn and thought of as BAT, though the word found in the dictionary was flittermouse. What, if anything, will you give for a mouse or mouse-like creature - birds, beetle, butterflies?
6. BEETLE: No comments.
7. BENCH: What about Chairs? Sofas?
8. BIRD: The dictionary word was 'Corn Bunting', but the original was thought of as BIRD, and the picture is not specifically of a Corn Bunting. The latter is a passerine bird and the picture agrees with this. How are you going to distinguish between this and DODO (q.v.) which was thought of as a 'duck-like' bird? Do you propose to give full or half hit for eagles, peacocks, ostriches, sea gulls, etc., or only for birds of passerine appearance?

9. BOAT: The dictionary word was SHIP and was illustrated by a full-rigged sailing ship, but the agent thought this too difficult to draw and drew the fore-and-aft rigged sailing boat shown. Always thought of as BOAT rather than SHIP. What do you propose to do about full-rigged ships, steamers, battleships, boats with masts but no sails, rowing boats, canoes, racing eights?

10. BOOT: Dictionary word was SHOE, but agent drew a BOOT. What are you going to give to shoes, horse-shoes, feet?

11. BOTTLE: This was based on dictionary word VACUUM BOTTLE, which was rejected as too difficult. As indicated, the agent thought of a glass wine bottle. What will you give to carafes, medicine bottles, rubber hot water bottles, jars, vases, etc.?

12. BRACKET: An unsatisfactory original, because something angular and structural can be found in very many drawings (e.g., chairs, houses, bridges, etc.) Cf. paragraph 9 of Guide and keep the standard high to avoid dilution.

13. BRIM: The word was BRIM, understood in the sense of the rim of a vessel, etc., not a hat-brim. While drawing, attention was focussed on a chalice or goblet shaped vessel as shown.

Do you propose to give full or half hits to all 'containers' - remembering that bottle and ewer (q.v.) must be distinguished, or to goblet shaped cubs only? What about vases, bowls, plates, tea-cups, saucepans, eggcups, teapots, etc.?

14. BUFFALO: It is generally agreed that all cows, bulls, etc., should count as hits on this original. What about deer, rhinoceri or animals with tusks resembling the buffalo's horns?

15. BULBS: Geissler potash bulbs were illustrated in the dictionary and copied. Agent unfamiliar with these. Probably chief idea in his mind was 'glass bulbs'. What about electric light bulbs, hourglasses, bulbs of the horticultural variety?

16. BUTTERFLY: No comment except that as an example (not yet observed) of 'extremely like' it would be considered correct to give a full hit on this for a moth.

17. CASTLE: Battlements (castellations) were chief feature in agent's mind.

18. CLEOPATRA'S NEEDLE: No special comments. common sense must be used in respect of drawings or mentions of obelisks, pyramids, memorials, monoliths, church steeples, etc. Ask yourself whether these are in fact visually or functionally similar to C's Needle.

19. COTTON APHID: No comments.

20. CROSS-STITCH: Cf. Guide 10. Consider whether it is likely that a percipient imperfectly cognising this original would produce a St. Andrew's Cross, a Latin cross, a mention of sewing or of embroidery. *In re* crosses, cf. FLAG

21. DODO: Thought of by the agent (wrongly as it happens) as a duck-like bird. Known to be flightless. Must be distinguished from BIRD (Corn Bunting) q.v. Consider what, if anything, you will give for ducks, geese, swans, etc., passerine birds, nondescript birds, eagles, gulls, peacocks, ostriches, etc.

22. EMBATTLED FESS: An unfortunate original. Agent thought primarily of the fess, not of the shield; but should anything be given for shields without fesses?

23. EWER: In the agent's mind the distinguishing features were the handle and the constricted neck. Should full or half hits be given for two-handled vases? For watering cans, teapots, etc., with handles but also spouts? For teacups, etc., with handles but no constriction? For saucepans, etc., with a different sort of handle? For vases, etc., with constriction but no handle? Cf. BOTTLE and BRIM.

24. EXFOLIATE: The agent was thinking of leaves, not trees. Should anything be given for these, or for leaves attached to flowers, fruit, etc.,

25. FAN: No comments.

26. FISH: No special sort of fish was intended.

27. FLAG: It will be necessary to distinguish this from royal standard (44). The flag drawn is a black flag with a strongly marked white Latin cross. Consider black flags (without crosses), plain flags, striped flags, Union Jacks; Latin crosses without flags, other

sorts of cross; also pennants, burgees, etc.

28. FLEUR-de-LYS: No comments.

29. FROG: No comments.

30. HAMMER: Presumably a mallet would deserve a point, but what about axes and picks?

31. HAND: It is agreed that a hand is a hand regardless of whether it is right or left, open or closed also that nothing should be given to human figures merely because they may be presumed to have hands. Consider (a) figures with hands prominently outstretched or displayed (b) hands without bodies holding things (c) gloves.

32. HORSE (Jennet): The dictionary word was JENNET, but the original was drawn and thought of as HORSE.

33. MOUSTACHE: An unlikely original for anyone to draw quite correctly. Should faces with prominent moustaches be given a whole point, half a point, or nothing?

34. NET: The dictionary word was NET BLOTCH. This was rejected and the agent decided to illustrate NET. The man, beach and waves form a setting for the illustration; the little fish are incidental.

35. PARNASSUS: The agent was more interested in the temple than in the mountain. Consider Greek-type temples (criteria, pillars, steps, pediment) with and without mountains, mountains in conjunction with other types of building, mountains without buildings.

36. PRAWN: Must be sharply distinguished from FISH. What about lobsters, crayfish, crabs, etc.?

37. SATURN: Combination of sphere and rings, the last being most in agent's mind on account of difficulty of drawing ellipses freehand.

38. SCISSORS: No comments.

39. SHELL: Definitely chonchological; artillery version not thought of. What about shells of somewhat different shapes - snails, oysters?

40. SHOOTING: See contemporary note on original.

41. SKULL: No comments.

42. SPECTACLES: No comments, except that as examples (not yet observed) it would be proper to give a hit for pince-nez, but not for two tumblers ('a pair of glasses').

43. SPINNING TOP: No comments.

44. STANDARD (Royal Standard): The dictionary word was standard, the agent thought of doing Royal Standard and was concerned to make it other than an ordinary sort of flag; hence the rampant beast intended for a lion.

45. STOP-COCK: No comments.

46. VIOLIN: No comments.

47. THRONE: The object at the top is supposed to represent a crown. Consider whether you will give full or half points to easy chairs, ordinary wooden chairs, wooden chairs with arms, chairs of definitely ceremonial appearance.

48. TREE: Trees are drawn with very great frequency and in every degree of prominence, which makes hits on this original difficult to assess. Single trees, whether 'sole', 'co-equal' or secondary (Cf. Guide, 7) must clearly be given hits; but it is not always easy to decide where 'secondary' passes into 'incidental trimming'.

It is also for consideration whether forests, woods, avenues, rows of trees, etc., indicate that the percipient had the idea of 'a tree' in his mind; also whether two, three, four . . . or how many trees should be given half or full hit.

49. TRIDENT: The dictionary word was SPEAR and an illustration of 'Head of Fishing Spear' was copied. TRIDENT was introduced by the agent.

Consider pitchforks with two prongs, garden forks with more than three prongs, table forks, rakes.

50. WINDMILL: No comments.

SOME OBSERVATIONS ON THE EXPERIMENTS WITH DRAWINGS

by

W. Whately Carington

Whereas I have done my best to exercise the greatest care and circumspection both in my experimental procedure and in my account thereof, the somewhat colloquial remarks which follow are deliberately of a more speculative character, and are to be taken on the express understanding that nothing I say now shall be 'used in evidence against me' later.

The method I have used suffers from the serious drawback of being very laborious and slow in yielding results. To induce ten or fifteen groups of percipients to co-operate involves an amount of correspondential organization which is large compared with the labour involved in merely making people guess cards in a pack; moreover, the fact that the effect sought does not seem to show itself within a 'ten-original' experiment of normal size aggravates this disadvantage. Another serious matter is that, if we are interested in the pre- and retro-cognitive effects which my results at least strongly suggest, we must be very careful not to duplicate originals and, so far as possible, not to let later percipients know what were used in earlier experiments. Again, so long as we feel obliged to keep 'test' conditions, it is necessary that all drawings of any group of experiments should be scored 'blind' against all the originals of that group by a suitably uninformed judge; but to score against more than about 50 originals at once is probably more than any ordinary judge can manage, and this limits the range over which pre- and retro-cognitive effects can conveniently be studied. The last difficulty can probably be met by some suitable method of itemizing and indexing the material, though this again will be laborious; as for the first, it is to be hoped that someone will soon devise a technique which will speed up the collection of material without sacrificing the various advantages which the use of drawings unquestionably offers.

One plan would be to use groups working together and doing ten drawings inside of an hour, as in my Expt. II; another is to work in the same way with selected

individual percipients; a third might be to use ready-made drawings (say 100 or 200) from which one would be randomly selected for each trial and merely looked at, or preferably copied, by the experimenter.

Against these drawbacks there seem to me to be several advantages worth mentioning:

First, it is an easy matter absolutely to exclude all question of sensory clues, namely by the simple expedient of never allowing the percipient to be in the same room with the original he is required to 'reproduce'. We know now (or I, at least, venture to think we do) that this condition is not a fatal impediment; so, for safety's sake we may as well adopt it till the whole subject is on a better established basis.

Next, the method eliminates all danger of errors in counting of the type discussed by Kennedy in *Jour. Parapsy.* III, 2. p.213. There is no question here of whether the judge was wrong or not in making a given judgment or scoring, in the sense that it would be wrong in marking Star as a hit or Circle; the question is whether, being ignorant of what will make for success, he significantly gives more marks to certain groups of drawings, in respect of certain originals, than he would if there were no systematic resemblances to guide him; his markings (unless he is a very stupid or wicked judge) may be taken as gospel, and thereafter it only remains to add up the points allotted, which is a matter of simple arithmetic checkable in five minutes.

Very much more important than either of these, however, is the fact that the use of drawings as test material offers, it seems to me, an enormously greater chance than does the use of cards, etc., of finding out something about the kind of think that is going on.

A Zener card guesser must say either Square, Star, Plus, Circle or Waves - and what is the poor man to do if he has an image of what one of my percipients magnificently described as "an equal-sided squared circle"? On the other hand, there can be little doubt, in my judgment, that the original drawing may be reproduced in a distorted - even in a highly distorted - form as indicated by the significant result given by Mr. Hindson's 'Rejected Halves' (cf. my paper, IV. 1.b, Table V). When we have found out how far this distortion or transformation may go, and where it takes place, we shall know very much more than we do at present about what the whole process consists of.

I have the very strong impression, though I am not prepared absolutely to guarantee it, that the percipient in no sense 'copies' the original. It looks to me much more as if the percipient were reacting to the stimulus "Draw a Hand", for example, than to the stimulus "Copy that drawing of a Hand"; for we find left hands and right hands and open hands and closed hands in various positions, whereas the original was a left hand open, fingers spread, back uppermost. This is the first example that comes to mind, but I have the same impression throughout. This kind of thing suggests that it is the word rather than the drawing that 'gets across', and the suggestion seems to be strengthened by certain cases in which (but this is terribly rash speculation) the word seems to have been taken in the wrong sense - as Rings of Saturn appearing as jewellers' Rings or Ruler being interpreted as either Governor or Scale. On the other hand, I have little doubt that the relatively numerous Cows which appeared in my first experiment represented the Buffalo original; but, if so, then what got across would appear to be the 'idea' of an unspecified sort of horned ruminant, while the only drawing which was specifically 'buffaloid' indicated a pair of unmistakable buffalo horns, mounted, and facing the observer, whereas the original drawing showed the whole animal sideways.

Cautious psychologists fight very shy of using the word 'idea', and rightly so; yet here we have a situation in which things seem to happen as if something - which we may as well call an 'idea' as anything else, provided we realise that the term means virtually no more than 'X' - were 'transferred' (but this must not be allowed to carry spatio-temporal connotations) from one mind to another, undergoing various kinds or degrees of modification 'on the way'.

Provided we stick rigidly to the methodologically righteous practice of constructing our picture of the referent of the term 'idea' (as used in this context) solely out of experimental observations, not allowing it to become invested by implication with the properties of waves, particles, colloidal conglomerates or astrological whatnoteries, we may, I think, - with luck - end by evolving a concept as valuable to psychology as a whole as that of the atom to classical chemistry, the wave to classical physics, or the cell to biology. But this again is highly speculative, and may be to a great extent the product of wishful thinking.

I spoke above of transference 'from one mind to another', as if what is commonly known as 'pure clairvoyance' were out of the question. Few things would be more rash than to affirm this; but, for what it is worth, I increasingly receive the impression that the whole process is 'mental' in the sense that the relation is between the percipient's mind and the experimenter's mind, and not between the percipient's mind and a piece of paper. Even in cases where the actual form seems too closely reproduced to be fortuitous, I still have the impression that it is the 'idea of the form' that has got across rather than that the form itself has been directly cognized. This is only one step better than guesswork; but the results of my latest experiment (unpublished and not yet fully digested) seem to show pretty clearly that knowledge of the original in the experimenter's mind is at least a significantly important factor. Personally, I suspect that knowledge in some mind not too remotely connected with the experiment is likely to prove all important, though it would be unwise to commit oneself at the present stage. To me, this seems natural enough; but it conflicts with the evidence in favour of pure clairvoyance, which many of my present readers will regard as very strong, and verification of it might force us to the rather distressing conclusion that cases of apparent clairvoyance, e.g., in experiments with cards, are really due to precognitive impressions of the scorer's mind at or after the time of scoring. However, this dilemma is likely to be a triviality compared to some of the conundrums we shall have to deal with in due course.

Incidentally, the fact (if it be confirmed) that relatively more hits are scored on known than on unknown originals suggests a neat way of deciding whether questionable transformations are genuine. If an original of an 'A' (e.g., Cat) has been used under both known and unknown conditions, and it has been observed (as my experiment suggests that it would be) that relatively more cats are drawn after it is known than before, and we want to know whether drawings of 'Bs' (e.g., puss-moths), appearing fairly copiously in the experiment concerned, are transformations of Cat, or merely fortuitous, then we may enquire whether they also are relatively more frequent after Cat was known than before; if they are, the hypothesis of genuine transformation would be to a corresponding extent supported, and *vice versa*. But here again I am extrapolating to an almost indecent extent.

To conclude: In these brief comments I have quite wittingly allowed myself to stray down the more

enticing of the speculative avenues that have suggested themselves; and it is only too likely that in due course I shall find myself eating my words - a wholesome if unpalatable diet - in respect of any or every suggestion made. But I believe that, provided it is not taken for more than it is meant to be, a reasonable enterprise in looking ahead is as legitimate in science as care in experiment and caution in inference are necessary. At any rate I hope that my remarks may serve to stimulate interest in a technique which should prove at least a useful supplement to those already in use.

ned
cor
The
tic
Mac
ing
art
ESI

Fa
on
at

no
re:
st
er
di
gr
ac
on
od

sh
ca
fe
sh
li
th
sc
in

st
th

AN EXAMINATION OF KENNEDY'S STUDY OF THE
MacFARLAND DATA

Charles E. Stuart

In a recent number of the *Journal*, John L. Kennedy elaborates in three articles his proposal of a recording error hypothesis to explain ESP research results. The second of these, "A Critical Review of 'Discrimination Shown Between Experimenters by Subjects,' by J. D. MacFarland," is not accompanied by remarks of the reviewing committee. It would be unfortunate to allow this article to exemplify permissible critical research in the ESP field without careful review.

There is no intention here of "defending" MacFarland's paper from critical attack as such. But it is only reasonable to require that the critic employ methods at least as reliable as those in the work under attack.

The review committee pointed out that Kennedy did not meet the latter requirement in his recording error research. It will be shown here that, similarly, in his study of MacFarland's errors he committed methodological errors of much greater effect than the observed errors discussed. That is, the observed MacFarland errors were great enough to cause concern but were not adequate to account for the research findings. Kennedy's conclusions, on the other hand, can be wholly attributed to his methodological errors.

Observed Discrepancies

The MacFarland data consisted of paired data sheets made out by J. M. and D. H. respectively. The call columns recorded the calls of a subject and, if perfectly recorded, should have been identical in the paired sheets. Kennedy found discrepancies, however, which he lists in his Table II. He does not examine the effect of these discrepancies as errors which would affect the scores "since it is impossible to determine from the data in which record the error was made."

The exact determination is impossible, but a straight-forward series of inferences from the data throws much light upon this important point. First, the

presence of discrepancies having no potential effect upon the score shows that some of the discrepancies are the result of simple observational errors with no relation to their effect upon the score. Second, it is natural to expect that errors will be motivationally directed to produce the effect upon the score sought by the experimenter. It is relatively easy to test the supposition that all the errors were wholly accidental by setting up an expected distribution upon a "null hypothesis." For example, suppose the call series records were perfectly identical and errorless. And then suppose that at n random points one or the other record were changed in a random fashion. There would then be n discrepancies independent of the scores. The probability of the way in which any given discrepancy would affect the true score can be computed, since the number of effects possible are clearly delimited.

Table A is an expansion of Kennedy's Table II and includes, plus a few corrections, the discrepancies in the DT work. In the "Exp." columns are the expected number of discrepancies for each classification upon a null hypothesis.

Table A

Possible effect upon score	GESP		DT	
	Obs.	Exp.	Obs.	Exp.
M+1 or N	28	9	5	4
H+1 or N	5	9	4	4
M+1 or H-1	8	3	0	1
H+1 or M-1	3	3	2	1
M+1 or H+1	13	6	1	3
N	10	28	13	12
N or M-1	4	9	5	4
N or H-1	6	9	3	4
	77	76	33	33

It is immediately apparent that the DT discrepancies fit the expectations of the null hypothesis so exactly that we may say with assurance that the discrepancies did not produce any positive or negative trend in the test scores. (That is, the errors are very close to what would be expected if 33 pairs of call records had been selected and altered in a purely random fashion independently of the adjacent columns)

The GESP discrepancies show, however, a definite positive effect, mainly in the classification which would raise M's score. It is plausible to suggest, therefore, that in these data there exist motivationally directed errors. The problem is to find their magnitude.

A third assumption is suggested directly by the figures in Table A. The DT discrepancies occurred in data numerically equal to the GESP work. They are simple observational errors. If it is assumed that these are typical of the general rate of observational error in the experiment, then the magnitude and location of the motivated errors may be approximated in the GESP data.

Table B

	J.M.			D.H.			
	Obs.	Exp.	Diff.	Obs.	Exp.	Diff.	
M+1 or N	28	4	+24	H+1 or N	5	4	+1
M+1 or H-1	8	1	+7	H+1 or M-1	3	1	+2
N or M-1	4	4	0	N or H-1	6	4	-2
			+31				+1

Table B thus compares the observed GESP discrepancies with the expected discrepancies of the DT work. It appears that J.M. is responsible for the positive trend and that D.H.'s discrepancies are independent of the scores. We may then safely attribute the +10 difference in the "M+1 or H+1" class, and also all transposition errors, to J.M., and estimate the magnitude of his total positive error as +45.

This inferred positive error would reduce the J.M. deviation in the GESP work from +394 to +349. The latter is still a highly significant deviant from chance.

Possibly it was recognition of the fact that the observed errors could not account adequately for the results that led Kennedy to devote the major argument of his analysis to hypothesized recording errors in the card columns.

Hypothesized Errors

Statement of the hypothesis: The form of error hypothesized is a transposition or inversion error. The experimenter, while copying the card order in a rapid manner, might accidentally invert the order of two

consecutive cards, and this accident would be more likely to occur when such an inversion would produce one or two hits where no true hit occurred. Kennedy proposes that such a form of error would increase the number of "isolated" hits, would increase the number of "different symbol pairs" of hits, and would have no effect upon "like symbol pairs" of hits. By categorizing the data into isolated, different symbol pairs and like symbol pairs of hits, he proposed to show that the hypothesis would be substantiated.

In spite of Kennedy's inversion of procedure--setting up the categories, adducing results, and then interpreting the results to favor his hypothesis--there is no logical basis for the proposal in the first place. Motivated inversion errors would increase the number of hits independently of the order of the symbols, their effect being simply to change the random card order to an order more closely similar to the call order. They would increase the like symbol pair category in the same proportions as the other categories. For example, suppose the call order were star-square-circle and the card order were waves-circle-plus. A transposition in the card order would produce a hit, but whether that hit would be an isolated hit, part of a same symbol pair, or part of an unlike symbol pair would depend wholly upon the status of the next card-call correspondence.

In other words, Kennedy categorized the hits as "isolated" and "paired" to get differential effects which would illustrate the hypothesis of transposition errors. The point here is that *whether or not there are transpositions, their effect could not be shown in this way.* This is the major error of his analysis.

Application of a Matching Analysis: With the prejudice in favor of differential effects as indicative of transpositions thus removed, the actual differential effects reported can be interpreted with greater objectivity. In Kennedy's Table III, a matching hypothesis is applied to the categories. For the questioned data, the unlike symbol pairs and the isolated hits are significantly above chance, and the like symbol pairs are chance. This is precisely Kennedy's prediction. *But it is also what might occur if the matching hypothesis were not applicable.* And since in his next section Kennedy finds that the matching hypothesis is not valid for this problem, he corrects the data of Table III in Table VII. That is, Table III is simply a false start, admittedly an erroneous analysis in respect to the main problem, and the figures in it are illustrative only of that improper

approach. No note accompanies the table; no hint occurs in the discussion nor in the mathematical appendix that the whole of Section III, with the mathematical appendix, is the elaboration of a methodological error.¹

Application of a Corrected Analysis: To compare correctly the unlike and like symbol pairs, the actual number of like symbol pairs called by the subject was used to predict an approximate chance expectation. In Table VII we find that, for GESP, unlike symbol pairs and isolated hits occur more frequently than chance, and that the same symbol pairs present a borderline case. Kennedy rests his case on this illustration, offering three interpretations: (1) "The analysis is invalid since it is based upon statistical artifacts caused by the discrepancy in the number of cases in the like and unlike symbol categories"; (2) the differences occur because of some ESP process; and (3) the differences are due to inversion errors. He favors the third interpretation.

But since we know that the data simply do not bear upon the third interpretation, let us look again at the first interpretation, which Kennedy pointedly ignores. That is, are the illustrated differences true and reliable differences? Surely this would seem to be a prerequisite to any interpretation whatsoever.

Kennedy's method has been to categorize these data and to find the individual extra-chance probability of each category by a Chi-square method. The probability set-up of the ESP deck does produce a necessary discrepancy in the number of cases in the various categories, as is mentioned. But the number of cases in the categories will itself produce differences in chi square values. In a large group of observations the total of which is extra-chance and in which some extra-chance factor is relatively constant throughout, it is always possible to select at random a group small enough (for example, a single observation) that usual statistical methods cannot show it to be extra-chance. In the same way, if a sample of the data is large enough, it is practically certain to be significant. Thus, by

-
1. Kennedy refers in a footnote to Greenwood and Stuart suggesting the matching hypothesis for evaluation of the mean expectation of ESP scores. The footnote is wholly accurate, but two implications must be emphatically denied: (1) that the writers mentioned would have approved Kennedy's application of the matching hypothesis; or (2) that their use of it embodies the same error as his.

Kennedy's method, it would be possible to illustrate a large number of spurious differences within perfectly homogeneous material.

To measure properly the difference in performance between two categories, it is necessary to find a statistic independent of the number of cases. A simple one for the present case would be the per cent of deviation from chance expectation. (Many others might be used, but this one is readily computable from Kennedy's table). If Kennedy's hypothesis were right, the additional observation could not but support it.

The ratio of deviation to expectation for the three exclusive categories in Table VII are, for GESP: .50, .49, and .13; for DT: .38, .09, and .09. The first and greatest ratio in each case is that for the same symbol pairs! Whatever the extra-chance factor in these data is attributable to, it was of slightly greater effectiveness in the same-symbol category, rather than absent from that category as Kennedy's analysis proposes. Proper comparison of these statistics would necessitate finding standard deviations of the differences (which I shall not labor to compute since the actual values are of no relevant concern here). The main point is that a first test of Kennedy's differences indicates their spurious nature.

Kennedy's final arithmetical argument neatly characterizes his whole approach. "If the assumption is made that the deviations on unlike symbol pairs and isolated hits are products of recording errors" and these are subtracted from the total reported, "such a subtraction would reduce (J.M.'s) results to a chance level!" The unlike symbol pairs and isolated hits categories include about 97 per cent of the data in this analysis. If it is valid to assume a conclusive explanation of 97 per cent of the data, then there is little point in suggesting another hypothesis, even a chance hypothesis, to explain the negligible 3 per cent remaining.

In summary it is only fair to say that there may be inversion errors in the MacFarland data as Kennedy concludes. But their existence would not show the effects he attempts to study and could not be shown by any analysis of the data known to the present writer. The work cannot be discredited on the grounds that no control of such error was utilized until clear experimental proof of the existence of the error, with a relevant magnitude of effect, is shown in other work. It

is unfortunate that the article which seeks to discredit the MacFarland study is open to so many objections of a much more serious character.

Summary:

a. An elaboration of Kennedy's study of MacFarland's call series discrepancies is offered to show that the errors indicated were of two kinds--simple observational errors, and motivationally directed errors.

b. It is shown that for the DT work, the errors were simple observational errors of a kind which do not invalidate the results.

c. It is shown that in the GESP work, D.H.'s errors are simple observational errors, while J.M.'s errors included some which might be motivationally directed. The magnitude of the effect of the latter was estimated to increase the score by 45 hits. Although serious from the standpoint of rate of error, this increase does not account for the observed deviations.

d. The logical basis for Kennedy's hypothecated card column errors is examined and found to be unwarranted.

e. Kennedy's application of a pure matching hypothesis to establish his differences is affirmed to be a methodological error.

f. A corrected mathematical method illustrates differences which are found to be spurious.

g. Kennedy's conclusions regarding his analytic establishment of inversion errors are found to be wholly unsubstantiated.

A FURTHER ATTEMPT TO TEST THE ROLE OF CHANCE
IN ESP EXPERIMENTS

R. Bugelski and S. Bugelski
University of Toledo

In a novel attempt to test the role of chance in ESP research, Leuba (1) used imaginary "subjects" (consisting of decks of ESP cards, one deck providing the target cards, the other call cards). The technique consisted of matching the two decks of cards, allowing an indefinite number of runs to represent one subject, another unspecified number of runs to represent another subject, etc. If a "subject" was doing well, that is, scoring above chance expectation, the runs were continued; if, however, the subject began to score poorly, he was dropped and a new "subject" was started. In this manner, Leuba discovered three subjects out of a total of 87 who scored well enough to merit the terms "good" or "very good" according to C. E. Stuart and J.G. Pratt's handbook (2). The inference suggested by Leuba was that actual real life subjects reported by previous investigators as earning scores higher than chance represented only chance "lucky" deviations around an expected chance average of 5.00. The notion was suggested that if one tests *enough* subjects, some will be found who will score above chance (some would score below chance in order to average 5.00 for *all* subjects) for some time, before they too would fall to chance results. His experiment was an attempt to demonstrate this reasoning.

The present report deals essentially with a repetition of Leuba's experiment with additional controls not exercised in his exploratory work. Certain advances in technique and method introduced by other students of ESP, such as mechanical shuffling, the double-check, etc. were not used by Leuba. These and other techniques used in the present study will be mentioned under procedure.

-
1. This experiment is essentially a repetition of the work reported by C. I. Leuba, An experiment to test the role of chance in ESP research; *Journal of Parapsychology*, 2, 1939, 217-221.

PROCEDURE

The aim of the method followed in the present study was to eliminate as many variables as possible and allow only chance to operate in determining the order of the cards. In order to accomplish this purpose, we did the following:

1. One of the variables not considered by Leuba was the possibility of a "psychic" shuffle. To eliminate this variable a mechanical shuffler supplied by the Duke laboratories was used in the suggested manner (3), i.e. three complete revolutions of the shuffler were matched. Further, all cuts were made with an ordinary table knife.
 2. Use was made of a double-check to avoid errors in tabulation. One experimenter recorded the target deck; the other the call deck; then each experimenter checked the tabulation of the other.
 3. In Leuba's experiment the decision to stop work with any one subject was left to such variables as amount of time, mood, interruptions, and, most important, when the subject seemed to be doing poorly. In our experiment the number of runs any subject obtained was determined strictly in *advance* by a mathematical criterion so that such indefinite factors as emotional mood, disappointment from continued poor scores or pleasure from consistently good scores were eliminated. The criterion adopted was arrived at in the following manner: from the table in Stuart and Pratt's Handbook (Appendix A), the number of hits necessary to score a critical ratio of 2.50 and of 3.00 were calculated for runs from 10-40 inclusive (see Table 1). The aim of the experiment then became to obtain as many subjects as possible reaching or exceeding a critical ratio of 2.5 for some number of runs above 10.² With Table 1 at hand it was a simple matter to determine after any run how far behind the desired score the subject stood. If at any time a subject met this criterion, he was automatically dropped and considered
-
2. Usually at least 8 runs are necessary before the scores should even be considered.

as "good". If he fell behind the necessary minimum number of hits after any of a series of 10-run sessions by more than 10 points he was dropped as probably unlikely to ever make a good subject.

4. After the first 25 subjects had received one session each (a session consisted of 10 runs) with the mechanical shuffler, sufficient data had been accumulated to draw purely chance subjects from the results of the first 25. This was done by matching systematically the target record for one run with the call records for each of the remaining 9 runs of the first session of the first 25 subjects. Then target record #2 would be matched similarly with all of the remaining call records and so on. In this fashion, data were accumulated by shuffling for 25 subjects, for hundreds of additional runs, all of which are theoretically purely chance combinations which were never intended to be matched with each other. In this fashion 75 new "subjects" were created by having consecutive series of 9 runs manufactured from matches of targets and calls of the first 25 subjects. A 10th run was given each subject by shuffling a deck mechanically, this 10th run could also have been obtained from the accumulated data, but this procedure would have introduced additional bookkeeping difficulties.

With the above technique it is believed that the scores obtained are due only to chance as nothing depends upon personal manipulations, errors, or moods of the experimenters.

RESULTS

Of the 100 subjects run, five were found whose scores were sufficiently high to merit the "good" rating as suggested by the handbook; i.e. their critical ratios were above 2.50. Of these, one subject, No. 75 scored a total of 75 hits in 10 runs, meriting a critical ratio of 3.87. The number of hits per run of these five lucky subjects are given in Table II. Note that although subject 75 was given only 10 runs, his success was so phenomenal that if he were given 15 more runs during which he should average only chance success or five hits per run, he would still earn a critical ratio of 2.5 after a total of 25 runs because of the very good performance in the first 10 runs.

To digress from the main point for a moment, it might be well to give a picture of the total results: the 100 subjects were given a total of 1618 runs, the number of runs per subject ranging from 10 to 50 (unless a subject scored better than 51 hits in the first 10 runs he was dropped). The average number of hits per subject was 5.055 with a range from 0 to 13. In accordance with previous results more subjects scored below average. Incidentally, of the five good subjects only one scored 11 hits during one run and another scored 12 hits during one run; since a total of 13 runs of over 10 hits were recorded, this would indicate that one unusually high score will not affect the eventual average unduly.

To return to the main issue at stake, it would seem that it is quite possible to secure, if enough subjects are tried out, some subjects whose scores will run sufficiently high to match the scores of real subjects whose results have been used to support ESP theory. The fact that if a subject is given enough runs he needs score only an average of six hits per run to earn a reliable critical ratio suggest that ESP researchers would do well to consider raising the criterion of good results somewhat above that figure, regardless of the fact that the critical ratio indicates unlikelihood due to chance to the tune of astronomical figures. The warnings cited by Leuba against selection of data of any one subject seem to be well taken.

CONCLUSIONS

The present experiment, like its predecessor, is simply a logical exercise to demonstrate the principle that in the ESP technique of matching decks of 25 cards the chance score of five hits is only an average chance score that should be read: in the long run, the average number of hits scored by chance will be five. In arriving at this chance average, some runs will result in scores of more than five, and some in less than five. If one discards all the runs that result in less than five hits and retains all the hits that result in more than five hits for any given subject, or if one starts out by getting some good runs for one subject, it is quite possible to discover subjects who will score consistently above chance for some time. To base claims for some unusual or "extrasensory" capacity possessed by those subjects who score well in these card matchings does not seem to be justified unless it can be shown that certain subjects always or nearly always, and for a long time score pretty well above average. Average

scores of 6 and 7 out of a possible 25 are not very convincing. This experiment does not pretend to prove that there is no such thing as extra-sensory perception. It simply suggests a closer scrutiny of the criteria so extensively used, up to now, to indicate the existence of ESP.

REFERENCES

- 1) Leuba, C. H. An experiment to test the role of chance in ESP research, *Jour. Parapsychol.*, 1938, 2, 218-221.
- 2) Stuart, C. E., and Pratt, J. G., A handbook for testing extrasensory perception, Farrar & Rinehart, New York.
- 3) Smith, B.M., Note on a shuffling machine, *Jour. Parapsychol.*, 1938, 2, 231-232.

Table I

The number of hits in n runs necessary to earn a score of "good" (critical ratio = 2.5) and "very good" (Critical ratio = 3.0).

Runs	S.D.	Good Score	Very Good Score
10	6.45	66.12	69.35
11	6.76	71.90	75.28
12	7.06	77.65	81.18
13	7.35	83.38	87.05
14	7.63	89.08	92.89
15	7.88	94.70	98.64
16	8.16	100.40	104.48
17	8.40	106.00	110.20
18	8.65	111.62	115.95
19	8.88	117.20	121.64
20	9.12	122.80	127.36
21	9.34	128.35	133.02
22	9.57	133.92	139.71
23	9.78	139.45	144.34
24	9.99	144.98	149.97
25	10.20	150.50	155.60
26	10.39	155.98	161.17
27	10.60	161.50	166.80
28	10.79	166.98	172.37
29	10.99	172.48	177.97
30	11.17	177.92	183.31
31	11.35	183.38	189.05
32	11.54	188.85	194.62
33	11.71	194.28	200.13
34	11.89	199.72	205.67
35	12.07	205.18	211.21
36	12.24	210.60	216.72
37	12.40	216.00	222.20
38	12.57	221.42	226.71
39	12.74	226.85	233.32
40	12.80	232.00	238.40

Table II

Hits made by five good subjects on each run

Runs	Subjects				
	30	32	68	71	75
1	11	2	8	5	7
2	8	66	7	4	8
3	6	3	4	4	4
4	3	6	6	8	5
5	6	8	5	7	7
6	3	8	8	5	10
7	5	3	6	8	10
8	6	5	7	7	8
9	5	9	7	2	7
10	7	4	5	4	9
11	8	6	2	5	
12	8	8	2	4	
13	3	6	7	4	
14	4	8	8	3	
15	8	2	6	8	
16	7	6	5	6	
17	8	2	4	9	
18		7	4	12	
19		9	8	7	
20		4	6	4	
21		6	9	8	
22		9	4	4	
23		6	10	6	
24		8	7	6	
25		6		6	
26		5		4	
27		10		9	
28				6	
29				6	
30				6	
31				7	
Total	106	162	145	184	75
Mean	6.24	6.00	6.04	5.94	7.50
C.R.	2.5	2.5	2.5	2.5	3.87

ESP AND THE INVERSE SQUARE LAW

B. Hoffmann
Dept. of Mathematics
Queens College

It has often been claimed that if the results obtained by Professor Rhine and his colleagues are not illusory, the agency of transmission involved in ESP must be of a type quite unlike the simple radiations familiar to physicists. In his interesting article¹ "A Theory of Extra-Sensory Perception" Professor Reiser states the argument clearly in the following words² (the italics are mine):

"In experiments where subjects were set to the task of calling cards at a distance of several hundred miles the results --- indicate that the ordinary laws of radiation do not hold, and this suggests that a non-radiant energy is at work in ESP. --- If Dr. Rhine's results are valid, they necessitate the acceptance of a kind of energetics not limited by the customary inverse-square law, i.e. there is no decrease of effectiveness of extra-sensory perception with increase of distance, as is the case for known energies."

In the present note I wish to show that this particular deduction is without foundation. I do not wish to imply that the agency of transmission involved in ESP must after all necessarily obey the familiar inverse square law, but simply that it is impossible, on the basis of the available evidence, to draw any definite conclusion as to whether the agency of transmission obeys this law or not.

The basis of the deduction made by Professor Reiser, and by many before him, is the experimental evidence which indicates that ESP is substantially unaffected by distance. For the purpose of argument we shall assume that the experimental results are valid. We are not concerned here with the question of their validity, but solely with the problem of their interpretation if they are valid.

To deduce from these experimental results that the agency of transmission of ESP cannot lose intensity

in proportion to the inverse square of the distance from the source would be fallacious. The fallacy rests to a large extent upon a confusion of the two quite different concepts *intensity* and *intelligibility*.

A simple example will serve to expose the weak places in the argument. It is readily admitted that the intensity of light falls off as the square of the distance from its source. But this does not at all imply that there is a corresponding falling off in the *intelligibility* of, say, an advertisement written in the sky with smoke from an airplane. On the contrary, the advertisement can be read with substantially equal facility over a wide range of distances. And, in fact, if one should approach quite close, as the sky-writing aviator must, one would find that the intelligibility rapidly *decreased*. The relationship between intensity and intelligibility is evidently no simple one.

The illustration of the sky-writing is made even more pertinent by the fact that as the light waves become attenuated with increasing distance the solid angle subtended at the eye by the sky-writing also decreases, and by precisely the same inverse square law, with the result that the apparent brightness to the eye remains constant. This is a well known law of optics, and illustrates one way in which the effects of the inverse square law can be very effectively masked.

Let us take another example, illustrating a somewhat different possibility. Consider two radio stations having the same power, one close by and the other far distant. The signals from the latter will be much weaker than those from the former.¹ But modern radio receivers are equipped with a device known as "automatic volume control" which brings in stations, within wide limits, at approximately the same volume. When the receiver is tuned to the distant station the automatic volume control acts so as to compensate for the weak intensity of the signals by automatically providing the extra amplification needed to bring the volume to the level determined by the hand operated volume control, and the distant station comes in with approximately the same loudness as

1. Owing to the existence of reflecting layers in the upper atmosphere, the law of attenuation of radio signals is, in practice, quite different from the normal inverse square law. However since the intensity of signals does decrease with increase of distance the analogy remains sound.

the nearby one. Thus in modern radio sets the attenuation of signal strength due to distance is effectively masked.

Moreover, still keeping to the radio analogy, even if we turn down the hand operated volume control to reduce the volume to a whisper, or turn it up to a roar, we can still make out the news bulletins with substantially uniform success. Within wide limits their intelligibility is little affected by their intensity.

In estimating the significance of Rhine's distance experiments there are thus two major factors to be considered, at the least. One is the possible existence within the ESP recipient, or outside him, of an automatic compensating device which may effectively mask variations of intensity over a wide range of intensities, and the other is the very great difference between intensity, whether objective or subjective, and intelligibility.

It is not unreasonable to suppose that the human recipient of ESP impulses should find these two factors operating to mask the effects of distance. As to "automatic volume control", does not the eye possess an iris to compensate for variations in brightness over a wide range? The ear, too, is equipped with a mechanism which compensates for variations in loudness. And coming closer to the possible seat of the ESP faculty, the realm of the mind and the emotions, we may point out that the intellect is well protected against the sudden impact of novel and startling ideas, as every ESP worker soon learns, while the human conscience, like the sense of smell, is to a notorious extent adaptable to the exigencies of the moment. It seems more than likely, therefore, that the ESP faculty is endowed with a similar automatic adaptability to extremes of intensity. Indeed it would be rather unreasonable to suppose the contrary, since automatic control mechanisms of this sort serve the important biological function of preventing injury to the organism.

The vital importance of the concept of intelligibility, as contrasted with intensity, in the interpretation of the Rhine distance experiments, is beyond question. One has but to recollect that it is possible to read a billboard with equal facility from near and far, and to understand a speech just as well, or should we perhaps say just as little, from the rear of a hall as from the front.

I submit, therefore, that the problem of the nature of the transmission agency of ESP is an open one. Until a method is discovered for measuring its objective intensity, as distinguished both from its subjective intensity and from the intelligibility of its message, the question seems unlikely to receive a definite answer. So far as present knowledge is concerned, no valid deduction seems permissible concerning the mathematical laws governing the propagation of the ESP impulses.

REFERENCES

- (1) *Jour. Parapsychol.*, 1939, 3, 167-193
- (2) p. 178.

LETTERS TO THE EDITORS

To the Editors:

Subject: Review of "Experiments on the Paranormal Cognition of Drawings," by W. W. Carington.

This paper has been criticized by the following members of the Committee: Drs. L. Dick, Ernest R. Hilgard, John L. Kennedy, Louis Long, Irving Lorge, R. R. Willoughby, and the writer.

The criticism covers materials, procedure, statistical analysis, and theoretical considerations. Comments not specifically attributed to another member of the Committee are the responsibility of the Chairman.

1. On the whole, the Committee believes that the full report, while interesting, is unnecessarily long.

2. Drs. Dick, Long, Lorge, Sells, and Willoughby have criticized the use of drawings in the experiments. "We believe that ESP cards make better experimental material than drawings; more and perhaps even better data may be obtained by the use of drawings, but the results have to be based on very subjective ratings. The application of statistical methods to data scored in an objective manner has greater justification than such application based upon subjective data." (Dick and Long).

Drs. Lorge and Sells pointed out similarly that the subjective methods used in scoring the data do violence to the analyses based upon the scores. While the author indicates that variations in the standards employed by various judges are, at best, arbitrary, it must be noted that lowering of the standards resulted in improved scores for Mr. Hindson.

3. There were several points raised concerning experimental procedure:

"A major feature of the design seems to have been omitted; namely, the inclusion of a control series, consisting of drawings made by the same percipients in the absence of the stimuli (original drawings) whose alleged effect was later studied. For lack of this the author is forced to fall back on two less satisfactory controls, viz, matching one series against another, and matching

drawings against pseudo-originals (the latter seems more satisfactory than the former)." (Willoughby).

"It is unfortunate that the units prepared for MTH were made after the author's extensive familiarity with the reproductions, since the questions asked could not fail to suggest associations to a judge of the pictures. The list is given in the appendix, p. 125. It is a coincidence that comments are given on 38 of the fifty drawings, and the excess of matches over chance reported for MTH comes out 38.2. That is, if each hint caused one picture to be included as a match which would not otherwise be included, the entire excess over chance would be accounted for." (Hilgard).

"I am more or less willing to agree with the author that 'it is by the test of repeatability alone that it (the experiment) must either stand or fall' (p. 112). One question on technique should be asked. Why was a different judge used for the control series (pp. 93 ff.)? It seems to me that in order completely to avoid the criticism that the results are due to the peculiarities of a single judge, Mr. Hindson should have done the control series matching as well as the experimental series, being at the time ignorant of which was which." (Kennedy).

"The principal result appears to be on page 82. The author's reasoning on the preceding pages seems persuasive, but I would like to consider another approach. If we have 1,000 drawings submitted by N people who are trying to reproduce 10 originals, is not the appropriate null hypothesis that they are as likely to hit them (and no more) as they are any other originals? That is, the number of hits found among the 1,000 drawings of Series I varies only in a chance manner from the number of hits on Series II or III or IVA or IVB? If so, then what does it mean that in only one case out of five was the number of hits on the target series not exceeded by the number of hits on some series that was not the target--in three cases on some series that had yet to occur? Perhaps it means *precognition*, but I don't think so." (Willoughby).

4. Almost all of the reviewers commented on the significance of *precognition* in relation to the experimental results. Dr. Willoughby's remarks are representative.

"It is not clear why the author thus neglects the possibility of pre- and retrocognitions beyond the

bounds of a single experiment while stressing it so much within an experiment. While retrocognition is reasonable within the unusual premises of the whole approach, precognition, or the notion that a response can precede its stimulus, seems to me on the verge of the unbelievable. If *any* association with a small chance probability is to be taken as a causal relationship, without reference to time, place, or what we naive thinkers have thought of as possibility, where are we going to stop?" (Willoughby).

The Committee members generally agree that the experiment ought to be repeated. However, the complications of procedure encountered by the author may make this virtually impossible in a strict sense. The author's implication that only a "believer" is qualified to attempt repetition ought not to go unchallenged. Anyone willing to approach the problem as objectively as the author, regardless of belief, should be qualified.

Saul B. Sells

Chairman, Board of Review.

July 18, 1940

To the Editors of the Journal of Parapsychology:

May I ask the courtesy of your Journal to say a further word about a question which was brought up in the Symposium on ESP Methods at the meeting of the Southern Society for Philosophy and Psychology in the spring of 1939 and which Dr. Lemmon has again referred to in his critical comments contained in the recent volume: "Extra-Sensory Perception after Sixty Years."

I do not feel there is any essential disagreement between Dr. Lemmon and myself, and I sincerely regret having misunderstood the implication of certain statements in the paper he contributed to the symposium. On the other hand, I do not feel that he has quite correctly stated my position in the matter, which I wish to clarify through this public statement. Prior to the symposium I had done considerable mathematical work on the question of optional stopping, and the space devoted to the subject in my own symposium paper should make it sufficiently clear that I did not regard this question as one that could be disposed of merely by reasoning about it, and that I certainly did not regard my remarks in the discussion as an "answer" to the optional stopping hypothesis.

My position is very concisely summed up in the following statement from page 228 of the monograph:

"Series pooled lose their consecutive identity; only the last item included in the pool is subject to the optional stopping error."

Since Dr. Lemmon, in his footnote, does not take exception to this statement, I assume he agrees with it. However, certain statements in his symposium paper lend themselves to the interpretation that the apparent significance of a set of data may be influenced by the choice of stopping points other than the final one. On page 105 he says:

"For example, the experimenter might decide that since the subject was doing well today he would set a limit of twice as many runs as usual."

On a chance assumption, it is difficult to see how the setting of such a limit could affect the final result unless this particular day's work was the last of the entire series, or unless he had in mind evaluating this day's work separately and offering it as evidence of ESP -- something which has not been done in any of the recent ESP work. It is not my intention to criticize this statement which I apparently misunderstood, but merely to point out that it is open to the interpretation I placed upon it.

I am not sure I understand Dr. Lemmon's comment on page 226 of the new monograph to the effect that the optional stopping hypothesis "concerns experimental procedure rather than mathematical theory." It is true that one can adopt an experimental procedure which definitely rules out the hypothesis. However, if this is not done, it remains a mathematical problem to determine what correction should be introduced.

In closing, let me reiterate my conviction that there is no essential disagreement between Dr. Lemmon and me, and at the same time express my hearty endorsement of the very sane and impartial point of view he has consistently taken with regard to ESP, both in its mathematical and other aspects.

Cordially yours,

T. N. E. Greville.