METHODOLOGICAL APPROACHES IN PSI RESEARCH: AN OVERVIEW

A. R. George Owen

Trinity College
University of Cambridge
Cambridge, England

Owen: Methodology can be thought of in a restricted sense as a theory of experimental design, or more broadly, in terms of imaginative choice of avenues of approach such as the wide vistas opened by Dr. Tart and Dr. Ehrenwald. Both aspects are to some extent represented in my handout which reviews the ideas used by investigators during recent decades. I apologize for this review being neither concise nor comprehensive, and also to anyone present whose name might inadvertently have been omitted.*

This morning I shall speak in terms of the lower methodology. I take the low road—the side of experimental design. As is very proper in a speaker, I begin with a paradox: the theory of experimental design cannot of itself tell one how to do a perfect experiment. This is because

---

* [In addition to the topics more extensively discussed in this presentation (logico-statistical considerations and target choice) Dr. Owen's schematic review included the psychology of subjects (correlation between ESP ability and personality traits, motivation), the dynamics of the experimental situation (role of the experimenter, various agent-percipient relationships, dramatization of the set-up), subconscious psi tests (plethysmography, GSR, reduction of stimulus below sensory threshold, sensory-psi perception interaction), and systematic study of spontaneous material (recorded concomitant associations of agent and percipient, dream diaries).

All evidence referred to by Dr. Owen is based on published material. A comprehensive parapsychological bibliography can be obtained from the Parapsychology Foundation, Inc. — Editor.]

the precision and therefore the structure of the optimal experiment usually depends on the magnitude of the very effect to be demonstrated. But the nature and magnitude of an effect is just the thing that the experiment has to determine. Thus, as often as not, if one is blessed by encountering a perfect experiment, it was probably a result of designing by hindsight, after the actual real discoveries had been made. Perfect experiments are usually textbook experiments—they are propaganda, and are not part of the actual process of discovery.

Professor Flew said very rightly and properly that we must have repeatability. However, this seems to me a question pertinent to the particular stage in research at which we are. I am thinking of the common enterprise in which we are engaged, which is one of discovery, i.e., informing ourselves in the first instance, and not just seeking to convert a skeptical scientific public. That can come later (and it will) if we succeed in our enterprise. We seek to manufacture a high-scoring, repeatable ESP sensitive. If we are successful, we shall have repeatability, which is a proper scientific desideratum. If we fail, we shall not. The matter seems to me as simple as that. But, at the outset, our purpose is still practical and not academic or expository, and this means that we have to be somewhat opportunistic.

For example, we have to be guided by the actual performance of our subjects as we go along, and be prepared to change the form of experiment. Nonetheless, opportunism cannot be unbridled—major principles of design must be kept in view. Because the possible combinations of treatments (for example, hypnosis and/or drug A or drug B) are numerous, each phase of experimentation should be kept as short as possible. This is important on account of repeatability, because in long experiments we encounter the well known problems of boredom and decline effect, which presumably reflect the fact that each experiment is in itself a psychological treatment operating on the subject.

But, to be short and to have any informative value at all, an experiment must be sensitive. The more unique the target (that is to say, the more unpredictable) the more sensitive the experiment will be, and the more competent to detect ESP ability which, though present, may be low in magnitude. With Zener cards, the predictability of each target is fairly high, namely 0.2, and conventional experiments are statistically speaking relatively insensitive. Supposing, for the sake of argument, that one had centimal targets (say a hundred different symbols), unpredictability would be high and sensitivity very much increased, as shown in Table 1. I think this rather elementary though somewhat neglected point deserves consideration.

TABLE I

				A	1	4	9	99	∞
				p	0.5	0.2	0.1	0.01	0
Targets	T	C	P		*Number of Trials**				
Unibit	2	1	0.5		57	390	1590	160,000	∞
Zener	5	4	0.2		20	120	450	40,000	∞
Decimal	10	9	0.1		14	70	240	18,500	∞
Centimal	100	99	0.01		11	27	80	2,000	∞
Unique	∞	∞	0		6	17	35	374	∞

T = number of targets in target pool

$P = \frac{1}{T}$ = predictability (assuming all targets known to subjects, and presented randomly with equi-probability)

$C = \frac{(1\text{-}P)}{P}$ = odds against being right by chance

p = chance of being right by psi

$A = \frac{(1\text{-}p)}{p}$ = odds against being right by psi

$P + p - pP$ = chance of guess being correct

* Number of guesses which will be needed for each type of target to detect psi effects with 97.5% probability (at 2.5% significance level).

The situation where you have a target of low predictability approximates very much that of spontaneous cases, which I submit are sometimes very convincing. Although I could say in passing, as a worker in the spontaneous field, that I sympathize with Freud who lamented that he was embarrassed because his case histories read like short stories rather than scientific papers. This feeling that spontaneous studies are somehow less scientific than laboratory ones is partly justified, but by no means entirely.

To get back to the question of targets of low predictability, which I call quasi-unique ones, indeed we run into various problems which perhaps belong to the higher methodology such as the moot point whether the target range should be unknown or should be available to the subject's consciousness. However, I would like to make a rather strong plea for tending to approximate the laboratory situation to the spontaneous field cases situation insofar as targetry is concerned.

To reduce subjectivity by submitting the judges' assessments to statistical treatment, the target range could be chosen according to the principles of elementary information theory. For instance, a set of portraits could be made in accordance with the dichotomies: adult/child, male/female, nude/clothed, ancient/modern, indoors/outdoors, alone/in a crowd, colored/monotone, blonde/brunette. This would give 256 possibilities, each specified by 8 bits of information, and would therefore allow analysis of partial hits and of Gestalt effects (diametrical perception).

Another possibility would be to use targets tending to uniqueness, as often are those endowed with an affective or emotional quality, like art reproductions, surprise or erotic pictures, familiar names, color combinations, favorite music, toys (in children tests) etc. All these have been used by various investigators, but under so different conditions as to invalidate any attempt at comparing their intrinsic advantages.

Incidentally, with this kind of affective material the percipients' preferences as shown in their calls might themselves assist in characterizing their personalities, thereby providing an internal feedback to assist the experimenter in designing targets increasingly better tailored to his subjects' personalities. By the way, one of the more down to earth by-products of this conference might perhaps be a recommendation as to the right battery of personality tests to use in psi research, and how to weave them into the experimental program, bearing in mind that a psychological test might well be in itself a psychological treatment.

As a geneticist, I must confess to a certain pessimism about our success in attempting to determine psi conditioning parameters. Psi ability might depend on very specific biological variables, genetically regulated. If psi were monogenically determined, the answer might be essentially biochemical. At the state of the art, to achieve pharmacologically the required central nervous system state would call for a methodology of random search, that is to say a certain regulated wildness in experiment. On the other hand the situation may be better if psi were polygenically determined and therefore achievable through a delicate balance of many factors. In this hope, experimenters tend to try treatments in combination. In this respect I would like to stress the importance of balanced or factorial designs. These are vital in avoiding false correlations like Professor Flew's two clocks. Also, they allow maximum economy of scale of experimentation. This is because, in the absence of interactions, main effects are accurately determinable. If however there is a significant interaction present, this is in itself a valuable discovery. Factorial designs are therefore like two-headed pennies. There is, of course, no guarantee that in particular cases both sides of the penny are not labeled "Fail"; nevertheless they are very good for this kind of purpose.

Reverting to the question of what I call "Spencer-Brown" effects, all these awkward statistical problems can be very much reduced by using unique or unpredictable targets. One unfortunate by-product of a change in that direction might however be that we could not very easily measure psi missing. This might be a blessing in disguise, but I leave it to others to give a more definite opinion thereon.

Finally, I suggest that when working with a promising subject, we should, if possible, follow his psi ability in the round. That is to say, the laboratory experiments should be supplemented by continuous study of his spontaneous ESP.

This opportunistic kind of procedure probably involves the crime of optional stopping, but be of good cheer, there are other spheres where optional stopping, under a different name, namely sequential testing, is regarded not as a crime, but as one of the marvels of recent age. Consequently, in our initial enterprise situation, one should not attach too great an importance to statistical conservatism.

MARGENAU: Thank you very much, Mr. Owen. Ladies and gentlemen, if you desire, you may raise questions.

KRIPPNER: Professor Owen's discussion contained many relevant ideas and it reminded me of some problems we have with target pictures. In our experimental sessions items pertinent to the experimenter or to the person who prepared or handled the targets, but irrelevant to the target itself, sometimes produce results more interesting than those elicited by the actual target picture. For example, on the night of July 7, 1964, an agent was attempting to telepathically transmit Van Gogh's "The Starry Night." Another staff member, while handling the routine experimental duties and monitoring the receiver's EEG, began to read the July 10 issue of *Life* magazine at 5:50 A.M. The experimenter read two stories: "Topless Swimsuits" and "MacArthur's Reminiscences." At 6:50 A.M., the following dream report was given by the subject:

> In a park, and we were talking about two busts—women's busts—busts of Cleopatra . . . And we were arguing about that. . . .
>
> And a friend of mine was explaining to me that he wanted to go—that he had to go—on several travels, but it was . . . to the island of Gulliver, and people told him . . . , "Oh, you will be terribly lonely. . . . That island is so lonely far away in the Pacific. . . ."

The article on "Topless Swimsuits" contained a picture of a bare-breasted Minoan goddess and a reference to "early Egyptian culture"; the MacArthur article contained a detailed description of the general's campaigns in the Pacific. Very little of this subject's dream content resembled the target picture by Van Gogh. Nevertheless, the correspondence of his dreams with the *Life* articles being read by the experimenter was rather direct.

TART: A very important point has been challenged in the last three papers, and that is our conventional paradigmatic view of an ESP experiment, where a neutral observer (the experimenter) is not involved, while one or two subjects experience communication of relatively neutral information. There is evidence suggesting that this may not be the right view at all. The experimenter is definitely involved. We know that in parapsychology some experimenters seem to get consistently positive results with subjects, while others seem to get negative results. I think this is an indication of a much more complex interaction between personality factors of all people involved in the experimental situation than we usually conceive.

Dr. Owen wondered if this meeting might come up with suggestions for personality tests that would be generally useful in psi research. I would like to propose one, although the test has not been developed yet. The dimension to be measured is what I call "openness," i.e., the degree to which people are not afraid to have the private aspects of their life revealed. In the psychoanalytic examples given earlier quite often the patient comes up with something about the analyst's life. Often this is hostile to a certain extent. It is a demonstration of "I have some power over you; I can see what is going on here."

If we had ways of assessing this dimension of a person's openness, his ability to share his private life, this might help to differentiate among experimenters in parapsychology and the kinds of results they get. For instance, a very open experimenter might tend to get more positive results, while an experimenter who has a very strong need to guard his private life might have more psi missing come up.

SILVERMAN: Fitzgerald at the University of California, Berkeley, devised an "openness to experience inventory."[1] It combines aspects of Hilgard's hypnotic suggestibility scale, and some other scales that he put together himself. This may be a good instrument for differentiating between psi-sensitive and non-sensitive subjects.

OWEN: I think it is important to select your most promising subjects and stick with them. In fact, it would be very useful to use the spontaneous approach to do your selection before getting them into the laboratory.

MARGENAU: Mr. Owen, you spoke about experimental approaches without reliance upon any theory at all. You thereby set yourself an extremely difficult task, because most experiments are inspired by theoretical expectations. An experiment without theory is usually blind; therefore, if you could supplement theoretical ideas, they would make your job a little easier.

I would like to ask you another question. You advocate the use of extremely unlikely choices in statistical experiments. I have a feeling that if you make the a priori likelihood of a target-event very small (thus approaching the spontaneous situation) you end up with a binomial formula by which you judge the veridicality of your results, which becomes indiscriminate with respect to the number of failures or successes. But the reliability decreases with increasing likelihood of failure. Is that not true?

OWEN: Well, that would be true for infinitely unlikely targets. For the moment I do not want to go as far as that. I was suggesting that we meet infinity half way.

MARGENAU: In other words, make the likelihood not .2, but .01 or something like that.

OWEN: Yes. As regards your comment on theoretical support to experimental approaches, I was envisaging a situation in which the experimenter would concentrate on checking a theory of his choice, not mine. I am speaking only as a methodologist. It would be quite improper for me to suggest what theory a man should start with. I was trying to suggest that he should be prepared to modify his theory and try new experiments in the light of his findings. He should be flexible enough even to adapt the type and length of each experiment to the results he will be obtaining by the combination of his theoretical expectations with his subject's performance. In the enterprise of discovery this seems to me the only kind of real situation. The question of ultimate proof is difficult.

MARGENAU: Then I misunderstood you in thinking that your use of the term methodology was confined to experimental procedures.

Going back to the question of targets with low predictability, just as in spontaneous cases, you cannot assign a figure to their level of significance.

OWEN: Perhaps there is a way of getting around that. I agree with you that even in a very highly convincing spontaneous case, you often are in a situation where you know the occurrence is statistically significant but you cannot put a quantity to it. You have the confident knowledge that the significance level was really very low, and it seems to me that this is real knowledge even though it was not quantified. All the same, I do sympathize with the person who wants to put limits on the significance. That is why I proposed a way with which perhaps both spontaneity and some degree of quantification could be achieved, while keeping affectivity in the target.

TART: The problem of quantification is not that difficult: you can always count the frequency of occurrence of such characteristics in a comparable reference population. For example in studying token object reading with the psychic P. Hurkos[2] we obtained a "rarity rating," i.e., an actual count of the occurrence in a reference population of characteristics mentioned in the "psychic" readings.

OWEN: Often in research on humans, questions cannot be answered from genetic theory. One has to get field statistics and give an estimate.

SILVERMAN: In reference to the rules of evidence, it seems to me that the issue would be whether to accept as a fact the statement that someone is reading information from a source or to consider it as some kind of a probability statement.

OWEN: Students always ask this type of question. They say, "Why do you reject or accept data at a 5 percent confidence level?" And I always say, "You accept it or reject it depending on what you mean to do next or what kind of degree of certainty you want."

MARGENAU: That, incidentally, is no defect of parapsychology, because every scientist knows he has to set some limit of tolerance to the validity of his results. When we choose the probable error in science, we are doing something that is exactly as arbitrary as what you just described.

SILVERMAN: But it may be that you are putting some constraints on the definition of stimulus by the way you are playing your probability game and it may be that part of the methodology will involve a redefinition of stimulus.

EHRENWALD: Another difficulty in experimental design has transpired in our discussion. Sometimes it is not the target the experimenter has in mind that is being hit, and then we seem to get a failure. Although there is psychological evidence that we got a meaningful response, it is very difficult to make allowance for it in our experimental design. I described it many years ago as "paraexperimental telepathy." I think the experimental design has to be flexible enough to accommodate different qualities of the psi function. We must not try to force it into the straightjacket of our preconceived experimental design.

OWEN: I agree very heartily with that. We do not want to be like the man who went to South Africa to look for gold and when he came back he said it was a total failure, "I found very little gold; only a few diamonds."

REFERENCES

1. FITZGERALD, E. T.: "Measurement of Openness to Experience: A Study of Regression in the Service of the Ego," *J. Pers. Soc. Psychol., 4* (1966), 655.
2. TART, C. T., and SMITH, J.: "Two Token Object Studies with the 'Psychic' Peter Hurkos," *J. Amer. Soc. Psych. Res., 62* (1968), 143.