Journal of Parapsychology, Vol. 50, December 1986

COMMENTS ON THE GANZFELD CONTROVERSY

BY IRVIN L. CHILD

ABSTRACT: Parapsychologist Charles Honorton and psychologist Ray Hyman have presented meta-analyses of ESP ganzfeld experiments, directed at the question of whether these experiments provide replicable evidence for ESP. Three comments on their reviews are presented here: (1) Hyman misrepresents the standard design of ESP experiments when he says they lack control conditions. In fact, they appropriately use within-subject control rather than between-subject control but are reported in terminology that may make the presence of experimental control not obvious. (2) The pooling of results from several groups or conditions, which was done in several of the experimental reports, seems more likely to conceal a real effect than to erroneously identify one. (3) Because of the insensitivity of the measure used, the experiments are likely to understate evidence for a real effect that may be present. I give an example of a subject's two ganzfeld sessions for which this insensitive measure yields a *p* of only .07, whereas a more sensitive measure based on ratings by outside judges yields a *p* of .005.

This issue of the *Journal of Parapsychology* is to contain, I am told, a statement by Charles Honorton and Ray Hyman about points on which they now agree. This should be the most authoritative single outcome of the recent ganzfeld debate, since its authors are presumably the two people with the most thorough command of the experimental literature. While awaiting their statement, I would like to make some comments on issues of a general nature, to which their statement is not likely to be addressed.

THE QUESTION OF CONTROL

In presenting the general guidelines he intended to follow in his critical appraisal of ganzfeld research, Hyman (1985, p. 6) said he would focus on two questions: (a) whether the ganzfeld experiments "supply evidence for the existence of psi" and (b) whether they "yield evidence for psi that is replicable." He then continued:

> The basic index for both these questions is some measure of hitting or target matching compared with a chance baseline. This creates special problems when compared with more conventional measures of effect that depend on empirical comparisons between two or more groups.

He then proceeded (Hyman, 1985, p. 7) to characterize in a similar way the design of the first published ganzfeld psi study (Honorton & Harper, 1974), saying:

> It is relatively uncomplicated in that it consists of a single uniform condition for all subjects. No comparison or control conditions exist, and the only meaningful way to evaluate the observed number of hits is against some theoretical expectation.

It is simply not true that in this experiment (or other typical psi experiments) "no comparison or control conditions exist."

In the Honorton and Harper study, and in typical psi experiments generally, there is for each trial a pool of potential targets. One member of that pool is chosen at random and segregated as the actual target; the other members of the pool remain as decoys. Segregation of an item as the target is the experimental condition; being left as decoy is the control condition. A test is made of whether, through all the trials in the experiment, the subject's calls show more resemblance to items in the experimental condition than they do to items in the control condition. There is nothing unconventional or special about this test and no lack of control condition. The experimental-versus-control manipulation is within subjects rather than between subjects. The statistical test is thus comparable to a *t* test for paired observations rather than to a *t* test for means of independent samples—a distinction familiar to all psychologists—and involves a comparison between observation and theoretical expectation in no sense other than that which characterizes all uses of statistical inference.

In considering other fields of psychological experimentation, Hyman would not, I suppose, be likely to think that only a between-subjects design embodies a control condition and that a within-subjects design lacks a control condition. We see here a misunderstanding that grows out of the specialized vocabulary of parapsychology—a misunderstanding that we may expect to disappear from Hyman's writings if, as is to be hoped, he continues to give detailed attention to parapsychological research, and from the critical writings of other psychologists if they follow his admirable example.

Interdependencies Among Sampling Units

In the course of a generally reasonable discussion of how to define the units that are to be summarized in a research review, Hyman makes some confusing statements that seem inconsistent with his initial guidelines. On page 9 he speaks of "interdependencies among the sampling units," and later on he returns to similar considerations in simpler terminology:

> A very common practice was to include within a single experimental condition trials in which agents were friends of the percipient along with trials in which agents were members of the laboratory staff. Just how serious such violations of independence are is difficult to decide. One can imagine possible models in which they make no difference. But all such models assume that randomization has been optimal and that, on the null hypothesis, no psi exists. (p. 26)

Hyman is thus presenting the null hypothesis of no psi as though it were somehow restrictive and undesirable (perhaps especially to parapsychologists). It is, in fact, the null hypothesis to which his guidelines restrict his discussion; and it is also the null hypothesis used in any parapsychological research that is directed at the question of evidence for psi.

These sections of Hyman's article are likely to give the impression that combining data from different persons or different conditions poses a risk of creating an impression of significant findings not justified by the facts. Exactly the opposite seems more likely to be true. Combining data is appropriate for a test of whether, overall, the rate of hitting is significant. But if the individuals or groups vary systematically one from another in rate of hitting, the genuinely high performance of some may well be buried by the chance performance of many others, with the result that the null hypothesis is mistakenly accepted. For the question to which Hyman's critique is addressed, then—the question of whether the ganzfeld experiments provide evidence of psi—the pooling of individuals or of groups is more likely to prevent a positive conclusion than to sustain one erroneously.

Insensitivity of the Measure

The measure of performance most used in the studies reviewed is the occurrence of a direct hit—that is, success in selecting the true target as the pool item most similar to imagery during the ganzfeld session. Hyman (1985) has pointed out that in choosing sometimes this measure and sometimes others researchers have exaggerated the statistical significance of their findings. Honorton (1985) has shown that when this exaggeration is prevented, by restricting the review to this one most commonly used measure, the overall results still provide impressive evidence for psi. In this context of evaluating the overall evidence for psi, reliance on a widely available uniform measure is highly desirable, even if the measure is an insensitive one, for it permits elimination of an important potential source of error.

But this measure is indeed insensitive. If the pool consists of four items, chance alone would predict a hit on 25% of trials. Yet, to make the measure more sensitive, the subject cannot judge a much larger pool of items without the threat of boredom, confusion, and conflict and the certainty of delaying the feedback that may be important to his continued interest or even to the occurrence of psi on this occasion. When researchers' interests turn to other questions, such as personality correlates of psi performance or the effects of experimental variables, more sensitive measures may be more important or even, at times, essential. There is some danger that the great advances made with the insensitive measure may lead to continued reliance on it alone even where other measures are now more appropriate. Some of the other measures may, however, be expensive and time-consuming and thus not likely to be used except where especially needed.

In many ganzfeld sessions, the subject's imagery report shows striking similarities to the target that would seem likely to arise by chance much more rarely than 25% of the time. Is there any useful way of quantifying this improbability so that the facts about the direct hit may provide a more sensitive measure of possible psi functioning?

One approach is by explicit analysis of aspects of the imagery and the target (as well as decoys, where appropriate). Honorton (1975) began this approach by constructing a set of slides in which the presence or absence of ten features was varied systematically and orthogonally. The number of features on which presence versus absence was correctly judged could then provide the basis for a measure of psi performance. (It was used in some of the studies reviewed by Hyman; he identifies it by the symbol BC, for binary coding.) Jahn, Dunne, and Jahn (1980) have developed a similar kind of measure appropriate for the targets and imagery in "remote-viewing" experiments, and so have May, Humphrey, and Mathews (1985). Their approach could be applied to suitable sets of items used in ganzfeld experiments as well. One problem with all these approaches so far is that they may be based on analyses too superficial for maximum sensitivity and for penetration into the processes involved in the psi performance. A. J. Maren (1986) has recently begun to draw on the discipline of artificial intelligence for modes of analysis that might be superior in these respects, though perhaps so costly that they can be used only under special circumstances.

Another approach is to enlarge the pool for purposes of analysis without altering the task of the subject. This can be done by having judges rate the subject's imagery report from a ganzfeld session for its similarity to each of a larger number of items drawn at random from the population of potential targets from which the target and decoys were drawn.

I used this procedure as a class exercise in methodology in a college course on parapsychology (Child, 1978). Two ganzfeld sessions with direct hits that were conducted at the Division of Parapsychology and Psychophysics of the Maimonides Medical Center were used for this purpose. For one session (the subject's first ganzfeld session), the target was imbedded in a random position among 11 other slides randomly selected from the 512 colored slides of the Maimonides series mentioned above. The students in my class judged each of the 12 slides for their similarity to the imagery report, and the target emerged with the highest rating. The following year, another group of students performed the same task but with a different set of comparison slides; this time the target slide received the second highest average rating. Recently I followed the same procedure with a group of summer-program students at the Institute for Parapsychology, and the target slide had a higher average rating than any of the 11 new comparison slides. The target slide can at this point be compared with 36 other slides: the 3 decoys shown to the subject and the 33 shown to various groups of students. Only one of the 33 received a higher rating than the target slide. At this point, then, the one trial gives positive results expectable by chance only 5.4% of the time, instead of the 25% when the subject's judgment alone was available.

A second ganzfeld session of the same subject, which also had yielded a direct hit, was used in a similar exercise some years ago, and this too was repeated in connection with the recent summer program at the Institute for Parapsychology. On the first occasion it received the highest rating; on the second occasion, the next to highest. Thus, it can now be compared with 25 other slides; so high a standing would be expected by chance only 7.7% of the time.

These were this subject's first two trials with the standard ganzfeld procedure. How surprising would this result be on a basis of chance? There are 37 possible ranks for the target from the first session, and 26 for the target from the second session. There are thus 37 × 26, or 962, possible pairs of rankings. Of these, only 4 (1 and 1, 1 and 2, 2 and 1, 2 and 2) are as extreme as the pair of rankings obtained. The outcome is thus significant beyond the .005 level (p = .0042).[1] When the original finding of two direct hits is evaluated in the same way, it yields a p of only 0.0625.

For each trial, the enlargement of the sample of items available for comparison has very substantially increased the sensitivity of the measure. Wherever there is in fact occurring the kind of consistency that is called psi, and it is expressed in hits rather than misses, an increase in the size of the comparison sample would be likely to have this effect.

There may be some danger that emphasis on avoiding multiple testing might lead researchers to report only the one measure they have decided in advance to use in testing the significance of evidence for psi. This could have very destructive consequences. Honorton's (1985) meta-analysis illustrates this fact. The size of his sample of studies would have been greatly reduced if all the researchers had reported results only by their one most favored measure; by reporting several measures, they made possible the analysis Honorton performed at a later time.

Ganzfeld research is expensive in both time and money. The records of subjects' imagery and the materials from which target and decoy are drawn are an investment that may be used in later analyses, even for purposes not originally contemplated. I hope that one effect of the current controversy will be to clear the way for expansion and preservation of these research resources, and for their fuller use in some future era of more adequate financial support for basic parapsychological research.

References

Child, I. L. (1978). Some educational uses of the Maimonides slides. *Parapsychology Review,* **9**(5), 24–25.

Honorton, C. (1975). Objective determination of stimulus incorporation in ESP tasks with pictorial targets. In J. D. Morris, W. G. Roll, & R. L. Morris (Eds.), *Research in parapsychology 1974* (pp. 112–115). Metuchen, NJ: Scarecrow Press.

[1]I am indebted to George P. Hansen for suggesting this procedure, which had been used by Whitson, Bogart, Palmer, and Tart (1976). The reasoning is in essence identical with Morris's (1972) presentation of a sum-of-ranks analysis of preferential matching data.

HONORTON, C. (1985). Meta-analysis of psi ganzfeld research: A response to Hyman. *Journal of Parapsychology,* **49,** 51–91.

HONORTON, C., & HARPER, S. (1974). Psi-mediated imagery and ideation in an experimental procedure for regulating perceptual input. *Journal of the American Society for Psychical Research,* **68,** 156–168.

HYMAN, R. (1985). The ganzfeld psi experiment: A critical appraisal. *Journal of Parapsychology,* **49,** 3–49.

JAHN, R. G., DUNNE, B. J., & JAHN, E. G. (1980). Analytical judging procedure for remote perception experiments. *Journal of Parapsychology,* **44,** 207–231.

MAREN, A. J. (1986, August). *Representation and performance evaluation approaches in psi free-response tasks.* Paper presented at the 29th Annual Convention of the Parapsychological Association. Rohnert Park, CA.

MAY, E. C., HUMPHREY, B. S., & MATHEWS, C. (1985, August). *A figure of merit analysis for free-response material.* Paper presented at the 28th Annual Convention of the Parapsychological Association. Medford, MA.

MORRIS, R. L. (1972). An exact method for evaluating preferentially matched free-response material. *Journal of the American Society for Psychical Research,* **66,** 401–407.

WHITSON, T. W., BOGART, D. N., PALMER, J., & TART, C. T. (1976). Preliminary experiments in group "remote viewing." *Proceedings of the Institute of Electrical and Electronic Engineers,* **64,** 1550–1551.

Department of Psychology
Yale University
P. O. Box 11A, Yale Station
New Haven, CT 06520-7447