TOWARD THE ENHANCEMENT OF INTER-LABORATORY AND INTER-EXPERIMENTER REPLICABILITY IN PSI RESEARCH

REX G. STANFORD

Even while laboratory procedures such as hypnosis or Ganzfeld seem to favor ESP and have produced significant rates of replicability across long series of studies (Honorton, 1977; Stanford, in press), it is clear that the successful studies contributing to those success rates come from a limited number of investigators. Certain investigators have not, for example, reported success in their Ganzfeld work (see Stanford, 1982), even with repeated studies. To make matters worse, some investigators in parapsychology seem unable to obtain psi evidence whatever methods they use.

The cornerstone of science is the public character of its observations. In it there are no private revelations. When the ability to confirm a scientific claim depends upon who attempts to confirm it, there is reason for doubting that the claim to knowledge has attained genuinely scientific stature.

On the other hand, the situation with regard to parapsychological claims simply is not analogous to a hypothetical circumstance in which some astronomers claim to have observed that a certain star is exploding, but other astronomers cannot observe it even though they have suitable telescopes and observatory sites and know the proper celestial coordinates. That circumstance might justifiably lead either to charges of incompetence or fraud. The crisis would have been precipitated because of a claim which could not be confirmed by anyone wishing to confirm it under circumstances in which everyone would agree that it should be confirmed.

In psi research we are recognizably in the realm of interactions between the organism (usually *homo sapiens*) and the environment and thus are, whether we like it or not, in the realm of behavioral science. Behavioral science is widely acknowledged to lack the precision and, usually, the degree of replicability of the physical sciences. In parapsychology, even when we can specify a general set of conditions,

such as an hypnotic induction or the Ganzfeld procedure, which seems to favor ESP, experimenters, subject populations, experimental settings, experimental apparatus and procedures, and ways of measuring the dependent variable (ESP performance) often vary considerably, even from study to study within a given laboratory, not to mention across laboratories. Furthermore, it is widely acknowledged in behavioral science that experimental outcomes can be influenced by such factors. We are not, in short, in a situation within parapsychology where some failures to replicate across investigators or laboratories need necessarily be envisioned as a life-or-death crisis either for a specific claim or for parapsychology as a scientific endeavor.

Such a defense of our status quo would, however, be unacceptable if parapsychologists were unwilling or unready to embark upon serious, systematic efforts to reduce the mystery surrounding the circumstances necessary for the elicitation of psi events (assuming such circumstances exist). It is my objective here to discuss specific, systematic ways in which researchers can approach such problems in the hope of reducing the present uncertainty about why some investigators and/or laboratories seem unable to produce significant ESP results, even with putatively psi-favorable circumstances, while others can do so. The hope is that through such efforts we may either specify and implement circumstances under which cross-laboratory (or cross-experimenter) replicability can be enhanced or at least will be able to predict *a priori* the successes and failures. If parapsychology should be unable to meet these objectives it might usefully be laid to rest as a scientifically tractable endeavor.

In order to provide a uniformity and coherence in the treatment of these problems, research in the promising Ganzfeld-ESP area will provide an example here for the application of some general strategies for investigating and reducing inter-laboratory and inter-experimenter variations in the outcomes of psi research.

Experimenters as Social Stimuli

The social roles of a psi experimenter are often said to include: (a) providing clear, simple explanations for participants; (b) insuring that the explanations are understood; (c) furthering the participants' interest in and positive motivation concerning the experiment without creating apprehension concerning failure (often considered to involve creating a challenge while providing psychological support); (d) helping to reduce any fears or apprehensions the subject might have

concerning participation in the study; (e) providing a setting in which psi events, even personal ones, are expected and enjoyable, not threatening; (f) providing an atmosphere of openness and acceptance such that communication is free and easy and defensiveness is not aroused and (g) implementing the procedures (e.g., applying hemispheres to the eyes or setting the noise level) with adequate explanation and with care for the participant's comfort and feelings. Some of these roles imply empathic abilities combined with real social skills.

Though the above list should be viewed as neither definitive nor exhaustive, it should be sufficient to suggest that experimenters might vary widely in their ability to play the requisite roles. There is little if anything in scientific training which specifically prepares persons for such roles, so individual differences in social skills are likely to be large and fully manifest, even in a single experimental paradigm.

To illustrate such differences, it is useful to describe a pre-experimental briefing experienced by a friend of mine who participated in a psi study at a major "center." My friend described how, while being instructed by the experimenter, that experimenter stared fixedly at the ceiling for the duration of the lengthy explanation. As a pointed joke to aid the experimenter in understanding the absurdity of the situation, my friend slipped from his seat and hid in front of the experimenter's desk. Near the conclusion of the explanation, the experimenter finally looked back in the direction of the participant's chair, was startled to see him gone and began to call out his name. It would appear that this experimenter understood little about the need for two-way communication in preparing participants for the study.

At the other extreme, in being tested, myself, in a particular laboratory—again, a major "center"—I have always felt the sense of an interest in my personal involvement in the study and a real desire to share with me the excitement of the procedure or study. There has always been the touch of personal concern, empathy and understanding.

Sometimes aspects of equipment and procedures (to be discussed at more length later) overlap with social concerns, as when an investigator uses cellophane tape (stiff, sticky, hot and generally unpleasant) for applying hemispheres to the eyes. (Please do not laugh, for it has happened in a professional setting.) Procedural offenses are among the major concerns in psi research as to possible sources of variability in outcomes. Nevertheless, real procedural offenses are, in many cases, doubtlessly just an overt manifestation

of a general social insensitivity on the part of the experimenter or the person who planned the study.

Has anything already been done to illuminate the possibility that social factors are of importance in such research or to reduce any variability from such sources? Examination of John Palmer's scholarly reviews of the ESP literature (Palmer, 1978, 1982) shows clearly that studies designed to illuminate specific, personal styles which might influence ESP outcomes are very few in number, though, ironically, the results of the few existing such studies suggest clearly that personal-style factors on the part of experimenters may be important to psi performance. Are we a bit reluctant to examine the possibility that our own social deficiencies can contribute to our failures to obtain significant psi results?

As to what can be done, there are several possibilities.

(1) There is a need for systematic efforts to study actual experimenter-subject interactions in Ganzfeld-ESP studies and to see whether indices related to such interaction differentially predict success of the various experimenters in obtaining significant psi hitting with subjects sampled from the same population for all experimenters. Such work would, ideally, be done in a single laboratory to control for setting, equipment and subject characteristics. Each investigator would test an equal number of subjects in a standard procedure. Investigators could be deliberately preselected to emphasize certain social characteristics. It might be well to include some skeptical experimenters to learn whether their style in relating to subjects might differ and thus influence success. (The latter might help to suggest ways in which even skeptics could be trained to get "psi" results.) This approach is an ambitious one, and it would require that those involved in preparing the study become familiar with the literature on the study of social interactions.

Studies of social interaction in the psi experiment can also examine interactional differences within the same experimenter, including the role of those differences in success and failure. Such an approach might be very useful, but it will not be discussed at length here because of the present focus upon inter-experimenter differences. It does, however, have potential relevance to the present discussion in that the success of some experimenters might depend upon their shifting the style of interaction to meet the needs of the individual subject. This suggests that the variance of certain social-interaction measures (across subjects tested by a given experimenter) may be of value as one index of inter-experimenter differences.

While the approach being discussed here may be time-consuming, costly and demanding, it may hold maximal promise for helping us to understand the role of social factors in determining the outcomes of psi experiments. It has such potential value precisely because it is focused upon what actually happens in experimenter-subject interactions in psi studies. It does not rely, basically, upon paper-and-pencil tests of supposed experimenter social differences as reported by subjects, though such questionnaires conceivably could provide useful supplementation (if the presence of strong reporting biases are acknowledged and the questionnaires are answered prior to feedback concerning success). Nor does this suggested approach study experimenter social differences based upon situations divorced from actual experimentation.

It is fairly common to see unsuccessful psi experimenters ready to believe that experimenters as "psi sources," rather than as social stimuli, account for the variance in our experimental outcomes. Either that, or they begin to suspect that loose conditions or even fraud account for some of the differences. Despite such opinions, it is to be hoped that some of the unsuccessful experimenters might volunteer to participate, as historically unsuccessful experimenters, in these kinds of studies.

The reliability and validity of judgments of social interaction in studies such as these can be greatly enhanced by having permanent audio or audiovisual records of the interactions for blind judges to examine. ("Blindness" here refers to ignorance of the psi outcomes of the session and of the judgments of other raters.) The availability of such records allows repeated examination of interactional sequences and interjudge reliability is greatly enhanced (Cairns and Green, 1979).

(2) Another approach involves experimental manipulation of the characteristics of experimenter-subject interaction along the lines traditionally believed to influence ESP performance in order to learn whether such factors do influence psi performance. An example of this approach is a study by Honorton, Ramsey and Cabibbo (1975). While this approach is certainly useful and has the virtue of relative simplicity, it has some potentially important limitations. First, it necessarily starts from a set of *a priori,* and possibly incorrect or incomplete, assumptions about which factors in the interaction are important and in what ways they are important. If these asumptions are incomplete or wrong, such studies may produce misleading conclusions. This approach is not as "open-ended" as the one discussed previously (though the latter is not entirely free of assump-

tions because of the need to select categories of behavior for scoring or rating); the necessity of manipulations seriously limits the scope of the findings. Second, such manipulations must either be done by individuals preselected for ability to take such roles in a convincing, natural-appearing fashion (in which case we may still wonder about the generalizability of the results to other experimenters to whom the role is certainly a natural one) or must be superimposed upon existing experimenters whose feelings about the roles may contaminate how they are played or their consequences for subjects' psi performance. Third, this approach may well miss something which is crucial to understanding the problem at hand, for it may not sufficiently consider the truly interactional character of many experimental sessions. That is, the necessity for "preprogramming" actors may simply miss the truly "emergent" qualities of interactions which could influence psi outcomes.

(3) Some of the inter-laboratory and inter-experimenter differences in psi outcomes might be reduced if individuals wishing to experiment with a particular paradigm, such as Ganzfeld, could obtain a period of training in the administration of that procedure at a laboratory which has had success with it. Such training could result in increased uniformity of procedural and technical matters ranging from how properly and gently to apply and remove hemispheres to lighting arrangements (for visual Ganzfeld). Hopefully, such training would affect how subjects are actually treated during the course of the session, though the mere learning of a "proper procedure" for "handling" subjects in such a paradigm might or might not influence the social-interaction variables potentially crucial to psi outcomes. Of course, one hopes for a salutary effect, one in which the experimenter trainees somehow assimilate some of the "spirit" of the work with subjects at the laboratory, as well as the nitty-gritty details of its procedure. Perhaps these social-interactive features of testing can best be taught by analogy with clinical procedures in psychology. Videotapes of sessions by the trainees could be made and then viewed by an experienced, successful experimenter, notes taken, and the videotape then viewed by both trainee and trainer with feedback given and with full opportunity for discussion. Also, a second investigator could interview subjects afterward and this information could be discussed with the trainee. The objective of this general approach is not to produce investigators who are social-psychological "clones" of some successful investigator, but to teach general principles of effective interaction in such settings. This approach is an optimistic one in that it assumes that some fairly fundamental aspects of

individual styles of social interaction can be successfully modified in a relatively short training period. This may be easier than it seems, however, for parapsychologically nonproductive experimenters may simply be unaware of the kinds of things needed to aid their subjects' psi performance. Some, at least, may not be generally inept in social interactions with experimental subjects, but may lack some special approaches which are helpful to subjects.

A training approach such as this would, in all probability, be most useful after there has been systematic study, as discussed earlier, of precisely how social interaction influences psi performance. It would then be more evident which features of interaction are most important for the training. The potential importance of this approach is emphasized by the recent development at the Psychophysical Research Laboratories, Princeton, New Jersey, of an automated, computerized Ganzfeld procedure which is aimed at allowing more interaction with the subject by an experimenter who does not have to be preoccupied with details such as security precautions. If the investigator thus has more or freer opportunity for interaction with the subject, the quality of his or her social skills might be all the more important.

It is quite conceivable that laboratories wishing to employ optimally psi-conducive experimenters would profit by preselection of socially skilled individuals as experimenter trainees. These individuals could then be trained in the manner discussed above. This would very likely prove the most efficient and effective approach. Selection of experimenter trainees could be facilitated by the videotaping of sessions conducted by potential experimenters and reviewing them later for the purposes of selection. Discussions with their subjects might also prove enlightening.

Experimenters as Psi Sources

Some writers (e.g., Millar, 1979) have argued that, usually, it is not our subjects who have psi abilities, but that the successful investigators are themselves powerful psi sources who, presumably unconsciously, generate their favorable outcomes through their own psi abilities. It might appear, at first, that if inter-laboratory differences in psi-research outcomes are due to inter-experimenter differences in the ability to use psi to confirm their hypotheses, then there would be little to be done about it. It is not possible either to eliminate experimenters entirely or to make any such effects disappear by wishing them to disappear, since they involve nonintentional effects presumably controlled by the experimenter's motivation.

However, it seems likely that some laboratories—and they include many of the successful ones—introduce various unnecessary loci within their studies for possible experimenter-psi manipulation of outcomes. Among the areas of special concern in this regard is the selection of ESP targets, *de novo* for each trial, by running the random event generator. Similarly suspect is the use of the random event generator for finding an entry point into a fixed random-number table for the individual experiment or subject, since unnecessary degrees of freedom are introduced here and, in any event, Schmidt (1981) has empirically demonstrated possible PK influence in a rather analogous setting. If laboratories accustomed to such use of random event generators would deliberately make an effort to reduce or eliminate such loci for psi operation, we might find that some of the inter-laboratory differences would disappear.

At the very least, perhaps such laboratories will derive a part of their data by thus eliminating unnecessary loci for possible psi manipulation, so that some of their data will be free of such possible influence. This allows the possibility of examining whether such loci do play a role in psi-mediated experimenter effects. It is very important, however, that if a laboratory generates ESP targets both by running a random event generator and by sequentially sampling a fixed, long sequence of pre-generated random numbers, the decision to develop targets from a given one of these two sources on a given occasion not itself be made on the basis of running the random event generator. To do so would mean that the fixed sequence could, theoretically, be sampled only when it would provide optimal targets (if that accords with the needs of the experimenter).

This general approach would seem to be the only way directly to gain a handle on the experimenter psi problem, though it is certain that the problem has potential ramifications which are broader than those addressed by this approach. Therefore, even though this approach cannot address all the ramifications of this sticky problem area, it would seem to hold promise of effectively addressing at least one of them.

Subject Differences

There can be little doubt that the populations of subjects sampled in studies at different laboratories and by different investigators vary considerably. This is due to the differing locations of the laboratories, the kinds of individuals most readily available within a given locale, varied means of recruitment of subjects and, possibly, recruiter differences.

Parapsychology centers are likely to attract as subjects individuals with definite personal interests in psi events, persons who have taken the initiative, in many instances, to contact a center themselves to inquire about ways of becoming involved. University-centered laboratories, on the other hand, often test primarily undergraduate students who, for some reason, feel obligated or simply wish to participate, may be required to participate or may pick up some extra course credit for participation. Of course, they do test some individuals with considerable interest in psi phenomena, but that is not the rule. Among their most motivated subjects are psychology majors who simply want to get involved with laboratory research as a part of their learning process. Such subjects often lack a special interest in psi research, even though they are often very open-minded concerning psi events.

In many other ways, too, subjects differ who are tested at different laboratories. Parapsychology centers are likely to attract a substantial number of persons who believe ahead of time that they personally have psi ability, but would like for personal, social or even financial reasons to have some external validation of that belief. This is not usually true of university settings not identified as special "centers." Factors such as age, education and socio-economic status of volunteers vary widely, in both mean and variance, from laboratory to laboratory. Some investigators seem to use largely personal friends and acquaintances as subjects, and those subjects often participate in several studies. The extent to which laboratories use primarily experienced or "recycled" subjects varies greatly, and many university-based investigators use primarily naive subjects previously unknown to the experimenter. Some of the above factors influence the probability that the subjects believe in the reality of psi phenomena and expect psi to happen in the study and even in their own session. Investigators who tend to recycle subjects may also tend to select them for psi performance, and subjects who are asked to participate again may select themselves on the basis of prior performance.

Which, if any, of these many subject-population differences will influence psi performance in a given study is uncertain. Nevertheless, many researchers would surely feel that such subject-population differences might be important. The extensive literature on the effects upon ESP performance of factors such as belief and various personality attributes is difficult to ignore.

What, then, can realistically be done to address this problem? One simple, straightforward approach is for investigators in the several laboratories who are working with a given paradigm (e.g., Ganzfeld)

to develop and retain a file of information on the characteristics of the individuals who participate in their studies. It would be useful for investigators to reach some degree of consensus about which measures of subject differences to use. This would aid in helping to determine whether inter-laboratory differences in outcomes might be attributable to such differences. The careful selection of such measures is of the greatest importance, and such measures should surely include those deemed relevant to psi performance on the basis of prior research. Such measures should certainly have demonstrated reliability and validity, aside from whatever parapsychological usefulness they might have. Also, the measures should be so small in number and compact in form that subjects are not burdened with answering immense numbers of questions. Such a file on subject differences might profitably include, as judged by the reviews of Palmer (1978, 1982), various sheep-goat measures, an extraversion measure and a neuroticism measure, as well as records of previous psi testing and basic demographic information.

If it appears that differences in psi outcomes across laboratories might be related to subject differences, then an "unsuccessful" laboratory might find it useful to preselect subjects so that their characteristics are more similar to those at "successful" laboratories.

Experimental Settings

Laboratory settings used by major investigators in this country range in appearance from visually delightful to visually tolerable. However, the psi-research literature contains no systematic studies of whether such factors influence ESP or PK performance. If such factors do influence psi performance, it is conceivable that inter-laboratory difference in these regards might, at least in part, be compensated for by the presence as interactive experimenters of persons who exhibit a high level of professionalism, have good social skills and enjoy working with people. External appearances may have less of an influence upon subjects when they are working with an experimenter who can facilitate their interest in the task at hand, though the impression created by a dirty or disordered laboratory might not be overridden by experimenter skills. Dirty or disordered laboratories reek of nonprofessionalism and I have never known of such a laboratory which reported consistent success in psi research. As for attractiveness per se, ideas for the improvement of laboratory decor can be had by visiting some of the "successful" laboratories. Improvements in attractiveness need not be expensive.

Laboratories also differ in the amount and kinds of equipment present to the view of subjects. The presence of impressive laboratory equipment might influence the morale and even the motivation of subjects. A well-equipped laboratory is more prestigious, and subjects might feel that "surely these investigators know what they are doing or they would not have such good funding." The amount and sophistication of equipment visible might influence actual expectations that psi will occur in the study. Once again, however, the literature contains no studies of the effects of such variables upon psychological or psi measures. If the equipment variable does influence psi performance, it is possible that inter-laboratory differences in this respect can be at least partially compensated for by the maintenance of a high level of professionalism in the conduct of experimental sessions. If a subject sees evidence of competence, well-grounded confidence and carefulness on the part of the investigator in all phases of the investigation, the lack of impressive equipment might be far less salient for that subject. One at least hopes that professionalism is more fundamental and important for subjects than is equipment per se. Impressive equipment might enhance the image of professionalism when professionalism is already present in terms of the conduct of experimental sessions, but the presence of such equipment could, conceivably, "backfire" upon subject morale when true professionalism is lacking. Does the presence of such equipment cause subjects to expect a high level of professionalism which can easily be frustrated if not met? There are many unanswered questions here and a great need for some systematic study of how professionalism and equipment may jointly influence subject morale, motivation and psi performance. For the present, investigators in less-well-equipped laboratories surely could not go wrong in making sure that the level of professionalism in their studies is high. Probably, the presence of an interactive experimenter with some social skills should be considered part of the professionalism which is of potential importance here.

Another avenue by which the appearance of a laboratory and how well it is equipped might influence both the psychological and psi outcomes of a study is that such factors could influence the morale of investigators and, thereby, the performance of subjects.

Experimental Apparatus and Procedures

Even within the Ganzfeld paradigm, apparatus and procedures have differed radically. One investigator used cellophane tape to attach the hemispheres; some use halved electrode "collars"; others

use paper-fiber tape of the easily-removed surgical type. Most experimenters have used the traditional, halved, off-white ping-pong balls as Ganzfeld hemispheres, but one study was done wherein the subject was allowed to select either red or green translucent hemispheres. (In the latter case, since a red stimulating light source was used, the green hemisphere must have produced a very dark visual field surrounded, wherever leaks were present, by seemingly very saturated red light.) Also, some Ganzfeld investigators have used lights with red filters; others have not used a filter. Some have used fluourescent tubes; others, incandescent bulbs. Luminous flux, physical arrangement (e.g., reflector?) and distance of light source, like intensity and kind of auditory stimulation, have certainly not been standardized. (Indeed, actual level of auditory Ganzfeld stimulation has come to be specified only in some of the most recent research.) The subject's physical situation during Ganzfeld has also varied (e.g., whether a recliner chair or bed is used). Which, if any, of these differences is important to ESP performance is presently unknown. In the face of the serious claims for replicability with Ganzfeld and the disputes concerning such claims because of inter-investigator (or -laboratory) differences in Ganzfeld ESP success, investigators (who are not already doing so) should begin to specify clearly features of apparatus and procedures and, where realistic and desirable, to standardize them in accord with those used at successful laboratories or in accord with results of systematic studies of the effects of such variations in apparatus and procedure.

The target pools from which targets and control pictures have been sampled have also varied widely from laboratory to laboratory and even from study to study within the same laboratory. This is a potentially very important fact since there is already experimental evidence suggesting that some targets are more ESP-favorable than others (e.g., Sondow, 1979), and this evidence does not seem explainable on the basis of response biases. The most immediate need, then, is for considerable systematic research on the effectiveness of various kinds of target pictures. Later, perhaps, some standardization would be useful in this area.

Differences between studies in the method of random selection of targets may also be important to psi outcomes. Earlier discussion was directed to the problem of possible experimenter psi when random number generators are used in target selection. It is also important to recognize that some ESP studies (see Stanford, 1982) involving Ganzfeld have used potentially deficient methods of randomization which might have contaminated their results. Also, a number of

reports have given almost no information about the crucial matter of random selection of targets. What is minimally needed at present is clear specification of target-selection procedures and an upgrading of the deficient methods which have been used by a few investigators.

Although some standardization in equipment, apparatus and procedures would surely be useful, such standardization must follow in the wake of careful research and planning. A rush into standardization could, in the face of a technological or judgmental error, be ruinous for large numbers of research projects. Nor should any efforts at standardization discourage new methodological exploration. Such a turn of events could cause researchers to settle for less than optimal methodology and might result in a period of methodological stagnation.

Measurement of the Dependent Variable

In free-response ESP work an investigator must decide: (a) who (subjects or outside judges) will do the judging of correspondence between session experiences (or recorded utterances) and the pictures (target and controls); (b) how that will be done (usually, ranking or rating) and (c) how the information thus gained is to be treated statistically to determine the inference of possible ESP (or to contrast conditions or compute correlations among variables). Ganzfeld studies certainly exemplify the diversity of choices which have been made in confronting these issues. There are central issues in making such choices, and they will be aired here in the hope of fostering some convergence upon either useful, efficient methodology or research which would point in that direction.

Who Should Judge?

Subjects' verbal reports of Ganzfeld mentation will always be less than complete. Nevertheless, when their mentation report is played back to them (or summarized) at the end of the session prior to judging, memory traces related to unreported details may be redintegrated, giving them access to most of what transpired during the session. Outside judges, who would merely hear or read the subjects' mentation reports, would thus lack some of the information about Ganzfeld mentation which subjects themselves would have in judging. Outside judges have no interior traces related to the session which can be redintegrated by exposure to the subjects' mentation reports,

so they should have a disadvantage in judging as contrasted with subjects, all else being equal. There are some indications in the Ganzfeld literature that subjects do have such an advantage when their judging is contrasted with that of only one or two outside judges (Stanford, 1982).

Despite this information advantage for subjects, as opposed to outsiders, as judges, it is clear that judging by any single individual is likely to be a fairly imprecise affair. Even if ESP occurred, a single judge (subject or outsider) might not have the skill or the cognitive make-up to detect the type of information encoding which occurred. Thus, if several outside judges were employed, the opportunity for detecting any extrasensory information transfer might be optimized. There is, indeed, one study in the literature which suggests that the use of multiple outside judges (specifically, eight) is essentially as effective as using subjects' own judgments, despite the information disadvantage of such judges which was discussed earlier (Child and Levi, 1980). With the use of more or better-trained judges, the use of multiple outside judges might actually have an advantage.

The interest in outside judges is, however, based upon more than simply the realization that multiple judges might do a good job when their judgments are taken collectively. Some investigators have suggested that when subjects judge their own sessions (and are, thus, exposed to both target and control pictures), they may tend to displace their ESP to control pictures (for reasons which cannot be discussed here). There are, in fact, two empirical studies (Child and Levi, 1980; Palmer, Bogart, Jones and Tart, 1977) providing evidence of such displacement when subjects have been exposed to both targets and control pictures in a judging procedure. There is now a need for studies additionally including the critical control group in which subjects get feedback by seeing the target, but not the control pictures. Thus, in one group, subjects (as a manipulation) would judge their own sessions; in the other, they would get feedback simply by seeing the target picture. Then, outside judges would use the mentation reports of both groups to provide ESP scores; multiple judges would be desirable. The hope here would be that the elimination of subject judging, combined with multiple outside judging, would increase the possibility of getting evidence of ESP—both because of the use of multiple judges and the avoidance of displacement. More work is needed, in general, on the problem of who should judge. The question of how best to prepare or train individuals for judging also could profit by some research.

Rating or Ranking?

It is clearly wasteful to have judges do ranking when they can do ratings. The latter are probably less stressful for them to do, anyhow. Ratings contain all the information contained in ranking, but they contain, additionally, considerable information about quality of correspondences. Ratings can either be used directly in statistical analyses which contrast mean ratings for target and control pictures (Child and Levi, 1980) or can be used to derive *Z* scores (Stanford and Mayer, 1974) which can then be used in a variety of statistical analyses. Z-score-based dependent measures are very useful in contrasting experimental conditions and in performing correlational analyses; they can often be used in parametric statistical procedures, but can also profitably be used with nonparametric analyses (Stanford and Sargent, 1983).

Statistical Treatment of Data from Judging

Historically speaking, Ganzfeld researchers seem to have had great difficulty in deciding whether to do statistical analyses of ESP occurrence on the basis of binary hits, direct hits, sum-of-ranks or other measures such as mean *Z* score. Recently, however, investigators seem to be increasingly aware of the fallacies of failing to select the statistical analysis on *a priori*, rational grounds (Stanford, 1982). There is a growing realization that there is usually no justification for using methods (such as evaluating binary or direct hits) which lose considerable information and that we are logically required to use methods which consider most or all of the information available in our data, such as sum-of-ranks analysis (Morris, 1972; Solfvin, Kelly and Burdick, 1978), ratings-based analyses built around *Z* scores (Stanford and Mayer, 1974; Stanford and Sargent, 1983), or parametric analysis based upon rating data (Child and Levi, 1980). The choice among such methods is beyond the scope of this paper (but see Stanford, 1982 and Stanford and Sargent, 1983).

Consistent use of these evaluative guidelines should aid in the assessment of the degree of replicability of findings with techniques such as Ganzfeld, and it might even aid in the reduction of interstudy variability of outcomes.

Concluding Comments

If parapsychology is to have any hope of moving from the status of "science as the use of scientific method" to "science as the

discovery of publicly verifiable knowledge through acceptable rules of inference," researchers shall have to embark upon at least some of the kinds of ventures suggested here. The future of parapsychology as a science will depend heavily upon our willingness and readiness to come to grips with the problem of replicability. What that future will hold if parapsychologists are thus willing and ready is uncertain, but psi research does not need another one hundred years in scientific limbo.

BIBLIOGRAPHY

Cairns, R. B. and Green, J. A. "How to assess personality and social patterns: Observations or ratings?" In R. B. Cairns (Ed.), *The analysis of Social Interactions: Methods, Issues, and Illustrations.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1979, pp. 209–226.

Child, I. L. and Levi, A. "The use of judges' ratings to test hypotheses about psi processes." *Journal of the American Society for Psychical Research,* 1980, *74,* 171–181.

Honorton, C. "Psi and internal attention states." In B. B. Wolman (Ed.), *Handbook of Parapsychology.* New York: Van Nostrand Reinhold, 1977, pp. 435–472.

Honorton, C., Ramsey, M. and Cabibbo, C. "Experimenter effects in extrasensory perception." *Journal of the American Society for Psychical Research,* 1975, *69,* 135–149.

Millar, B. "The distribution of psi." *European Journal of Parapsychology,* 1979, *3,* 78–110.

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

Palmer, J. "Extrasensory perception: Research findings." In S. Krippner (Ed.), *Advances in Parapsychological Research, Vol. 2: Extrasensory Perception.* New York: Plenum Press, 1978, pp. 59–243.

Palmer, J. "ESP research findings: 1976-1978." In S. Krippner (Ed.), *Advances in Parapsychological Research, Vol. 3.* New York: Plenum Press, 1982, pp. 41–82.

Palmer, J., Bogart, D. N., Jones, S. M. and Tart, C. T. "Scoring patterns in an ESP ganzfeld experiment. *Journal of the American Society for Psychical Research,* 1977, *71,* 121–145.

Schmidt, H. "PK tests with pre-recorded and pre-inspected seed numbers. *Journal of Parapsychology,* 1981, *45,* 87–98.

Solfvin, G. F., Kelly, E. F. and Burdick, D. S. "Some new methods of analysis for preferential-ranking data." *Journal of the American Society for Psychical Research,* 1978, *72,* 93–110.

Sondow, N. "Effects of associations and feedback on psi in the ganzfeld: Is there more than meets the judge's eye?" *Journal of the American Society for Psychical Research,* 1979, *73,* 123–150.

Stanford, R. G. "Ganzfeld-ESP research: A survey and critical analysis." In S. Krippner (Ed.), *Advances in Parapsychological Research, Vol. 3.* Jefferson, NC: McFarland & Company, Inc., 1982.

Stanford, R. G. and Mayer, B. "Relaxation as a psi-conducive state: A replication and exploration of parameters." *Journal of the American Society for Psychical Research,* 1974, *68,* 182–191.

Stanford, R. G. and Sargent, C. L. "Z scores in free-response methodology: Comments on their utility and correction of an error." *Journal of the American Society for Psychical Research,* 1983, *77,* 319–326.

DISCUSSION

HONORTON: In terms of social interaction variables, it just occurred to me listening to your talk that when we at Maimonides began Ganzfeld work we were coming out of the dream studies. When you are doing dream research you become very sensitive to the vulnerability of the person you are working with. After all, they are coming into the laboratory, they are sleeping there, they are exposing themselves to you in a way that is very unnatural. And so you learn to become very sensitive and supportive in the way you waken them, in the way you talk to them during the night. Moving from the dream studies into the Ganzfeld work it was natural to carry over at least some of that—I think on a pretty automatic basis without giving much conscious thought to it as something that might be important to describe or report. I have long felt a great deal of sympathy for the psi-inhibitory experimenters who are frustrated by the endless stream of ad hoc explanations for failure that seem to come from those of us who have reported psi results in our data and social interaction variables are very difficult to quantify. But clearly it is going to be necessary in order to extend the range of replicability to larger numbers of experimenters that we develop some of the kinds of procedures that you have described here, training in particular. There may very well be certain things that can not be adequately communicated in written reports that can only be picked up through direct exposure.

On the Ganzfeld parameters—light intensity is something that in the non-parapsychological Ganzfeld research is very important in terms of determining the type of experience a person will have, how quickly color disappears, color saturation takes place and so on. It is very important that we start, at least for several years, beginning to measure and report light intensity because this is varied, I think as you may have mentioned, everywhere from a 25 watt bulb to a 500 or 600 watt flood light. The degree to which a true Ganzfeld is produced could vary considerably from one situation to another. Also at the Cambridge PA meeting, in the little informal session that we had with Ganzfeld researchers there, it became quite evident that a number of people had to work in situations where they did not really have adequate sound isolation, where the subjects were in rooms where there were street noises, in some cases where the subject was aroused at the end of the Ganzfeld session by a harsh telephone ringing to alert him to begin the judging period. These are the things that we need to start reporting.

In trying to do a meta-analysis on the Ganzfeld research, I was very frustrated in attempting to deal in any way with targets. I think this is symptomatic of the problem we have in quantifying free-response material. No one describes targets in a useful way. Targets are pictures, targets are art prints. Except for the Maimonides target set, which has a predefined composition, there is really no way to know how target sets may differ from one laboratory to another, the degree to which target pictures within a judging pool are orthogonal enough so that the subject does not have to select between two fairly similar scenes and so on. This is a complete unknown. We must develop better means of reporting these things if we are going to get any further.

STANFORD: I certainly concur with those remarks about light and there is also evidence in the psychological literature about the effects of different levels of white noise upon arousal and performance. We haven't even reported these points in the literature, much less held them constant. Another factor that I would mention in connection with light that might be important to bear in mind is that most light sources produce heat. You could have a more distant light source that might not produce as much heat directly on the subjects, or you could have one not so intense that would be close and would produce a lot of heat. We should take a look at that. If a person feels as though he is getting the third degree in the Ganzfeld, I don't know what he might start to perceive. Sound isolation is another factor. We have little information about the sound attenuating capacities of whatever room people have and I would just generally agree with what Chuck Honorton says. We must have more specificity in the literature. When we start to specify these things in our reports, then we become conscious about them, and we may start to take a closer look at them and try to standardize them or try to study them in some way or another.

BERGER: I want to mention a few things about our automated Ganzfeld system, which I'd like to first describe. It is controlled by an Apple computer which, using a hardware random number generator, randomizes the target selection process. The computer then instructs the randomizer (who is not the experimenter) which one of four videocassettes to insert into a videorecorder. The computer then accesses the target and presents it six times during a 30 minute period to the sender in a distant room. During the judging process, the computer shows the receiver four targets from the target pool, one of which was the actual target which the sender viewed. The receiver rates each target for its correspondence with their Ganzfeld

experience by putting their responses into the computer with a game paddle. The experimenter is maintained blind throughout the session.

I want to say a few words about experimenter sensitivity toward research participants. In the automated Ganzfeld, the role of the experimenter is quite different from that of a traditional Ganzfeld. Since most of the procedure is under computer control, the experimenter mostly responds to computer commands to flip switches, turn over audio tapes, etc. The automation frees the experimenter to focus more effort on the needs of the participants. We spend a great deal of time with them both prior to and following sessions. We try to be aware of things like leaving the receiver unattended for long periods of time after being set up. Having been through this procedure many times myself, I know that one can feel extremely vulnerable when left alone in a sound-isolation room with one's vision obscured by a homogeneous visual field.

In the auto-Ganzfeld we are doing a few things that are new. One is the use of ambient music during periods, such as the initial setup, to make the environment a little more pleasant and less stressful for the participant. Participants are told to loosen their clothes, take off their shoes if they wish and relax and get comfortable. We try to make the transition from Ganzfeld stimulation back to a normal state a gradual process, as it can be a very abrupt experience to have the noise and light suddenly go off after a long period of stimulation. Both are very gradually faded off.

With regard to giving feedback, we try very hard to reinforce whatever positive effects we saw, even if the outcome is not a first choice. We try to point out what the participant did right, rather than what he did wrong.

One last comment about the heat problem. We are the laboratory with the 600 watt light. It is on a dimmer and is adjusted to the participant's comfort. I have measured the temperature increase in that room and have found that it will heat up six degrees during the course of the experiment with the light at full intensity. Bearing this in mind, we precool the room prior to a session, especially in the summer.

STANFORD: This attention to the concerns, the comfort of the subject, is precisely the kind of thing that I am referring to. It is not verbal; it is really part of the interaction. If you make the subject uncomfortable in an experiment, you communicate to that subject something about yourself; that you are either blind or you don't care. That is an important communication. If you are blind and don't care, then they may not trust you and they may not open up.

Of course, there are other factors, too, that may interfere with psi performance.

About the matter of always trying to play the role of the participant. You wouldn't run an experiment before you participated in it. I quite agree. I have had this rule for many years myself. I don't know if all parapsychologists abide by this or not, but it is really very instructive. I think the chief investigator ought to do that and, especially with a modification procedure, maybe a number of people ought to go through it because there are individual differences and responses to the situation. Running three of four people through a procedure is pretty cheap and easy if it is going to prevent your "blowing" your entire experiment, so to speak, with the change in procedure.

SCHLITZ: We are stressing the reporting of detail. I think that many of us have cited the Collins work in our papers. He references this enculturational model with which there are a lot of tacit bits of information within our experimental protocol that do not get integrated within our experimental reports. I would definitely agree that this material needs to be included and this is particularly directed towards people who have a more mechanical type of orientation towards psi, in that they are a little more regimented and algorithmical rather than enculturational in their approach.

With regard to Rex's idea of the subject/experimenter interaction, I think that in a case like this a phenomenological approach, if I may use that term, is really useful. You have emphasized the need for administering various psychological tests and, while I agree with that as a starting point, I don't agree that that would be a good end point. I think that we have done a lot of psychological correlations with ESP and I don't believe that has been a particularly fruitful area. I think that it might be some more unknown aspects of the interaction as well as of the subject's method of response that might be gleaned from doing a phenomenological profile. Also, by doing some type of phenomenology we might come up with a common way of reporting this kind of information, which is difficult to report, but needs to be reported.

STANFORD: I think you have a misapprehension, probably conveyed from my verbal report here, about what I proposed doing with regard to social interaction. I do not believe that we should rely in a heavy way on numerous psychological tests of the subjects for this. I think this would be totally off base and totally contaminated by the subject's desire to please the experimenter. You can query them afterwards, you can even have them fill out questionnaires, if you

want. But the major thrust of the emphasis is upon actual recording of the interaction, so that we can look at its features. Social psychologists have developed a lot of ways of looking at interaction. Some of these are difficult and they are in the process of being developed. We are going to have to learn what will work best within this area. But I think that we can get at some of the emergent qualities of the interactions that come out of the two-way flow of communication. That is important. I am not claiming that this is going to be easy. It may take years to develop, but I think if we do not develop some of these approaches we are going to remain ignorant of something that is going on here, that, otherwise, we can simply speculate about. We could continue the myth of the bad and the good experimenter for years, but have no documentation, if we don't make this kind of approach.

SCHLITZ: I agree with that wholeheartedly. My second point has come up a couple of times in the course of this discussion and has not been elaborated on. It is the concept of displacement within the target pool in free response experiments. Now, what do we mean by that? I understand a little about this concept of the observational theory and the role of future observers. When we are talking about displacement within the target pool, are we talking about psi displacement going to the targets at the time of the ESP experience, or are we talking about the displacement occurring at the time of the judging? In the latter case, the subject is receiving information in the judging period which then causes the displacement. I think these are really important questions to address, because we haven't dealt with the rationale for talking about this displacement idea.

STANFORD: It seems to me you are raising a theoretical question here that requires investigation at a theoretical level. But from the pragmatic standpoint, if subjects displace, we do not have what we primarily look for in these studies, namely direct hitting on the target. And this is what I am concerned with. When subjects are doing their own judging, you can get significant displacement onto control pictures. There may be some tendency for that to happen. When people can look at those several pictures at the end of the session, for a few moments they do not know which is the actual target. If there is one that might be easy for them to encode in some way or another because, maybe, relevant memories are in their heads, they can temporarily at least get a lot of reinforcement by this. Ah ha! There it is! That is it! It doesn't make any difference which one it is, necessarily; at that time it is undifferentiated for them. Maybe there is implicit theory behind what I just said—there

probably is, but that is beside the point. We do have evidence that displacement occurs and the question is how do we control for it? What I am advocating is that we do some studies that will help us pin down whether or not this is really the factor we think it might be.

SCHLITZ: I think that it is really important that we address where we think the displacement is coming from, because it does have important implications for how we design our experiments in the future. If you feel that it is in the judging process, that is different than if we feel that it is in the ESP process.

STANFORD: I don't understand that differentiation you are making. I mean a subject's mentation occurs during the ESP process. The feedback comes at the end.

SCHLITZ: But for some people the ESP comes with the feedback.

STANFORD: I don't understand that.

SCHLITZ: Can I make one more comment? This is a response to Rick Berger's point about you as the experimenter having familiarity with the target pool. I am just wondering if, in your interaction with the subjects, in helping them to do the judging, you are also interacting with them during the time when they are doing their mentation? If so, then that is a place for possible displacement also, but not on a non-psi level.

BERGER: If I understand your question correctly, do we interact during the mentation process?

SCHLITZ: Yes.

BERGER: No. At the beginning of the session we play a relaxation induction tape. At the end of the tape are instructions to the receiver to mentate. At the end of 30 minutes of white noise, we read back their transcript to them. When this is done, the computer shows the receiver four targets, one of which was the one viewed by the sender. The receiver is asked to give his associations to each as they are viewed. The receiver can see each target as many times as he likes. The experimenter can review any of the mentation that the receiver wants to hear, but the experimenter does not help the receiver with his judging. (Bear in mind that the experimenter is blind to the actual target). After the receiver is satisfied with his rating, the data are automatically entered into the computer and the sender, experimenter and receiver all converge in the isolation room where the computer presents feedback to the actual target.

BALLARD: First of all, I totally agree that the social aspects of the laboratory and what goes on in the laboratory are extremely important. One thing that I have found very useful in addition to the

typical debriefing and getting direct feedback from the participants, is providing participants with anonymous questionnaires so that they can provide feedback after they leave the laboratory. I get a return rate of about 50 percent and I get feedback that is usually very positive. I often get some very good suggestions that I never thought of. For example, in one experiment using relaxation, one person told me I had asked him to imagine that he was next to the ocean, something to that effect. The subject said, "Gosh, I was fine up to then, but I hate the ocean, and waves make me seasick, and you just lost me!" Now, I never would have thought of that. I always thought that most people would have found that positive. I think that is one technique that I would offer for some people in getting feedback.

The second is, I think it is important that we recognize that if we had a lot of people doing research, many of these factors would randomize out, if we were correctly identifying the necessary dependent and independent variables. I think that we will find, as an outgrowth of this kind of standardization, that there will be variables related to specific theories that will help us a lot in our future efforts. I would propose, based on some reading and research that I have done, that one such variable might be the concept of affective assessment—that is, as people come into the laboratory they react positively and negatively to all aspects. The more that we can make the task predication positive, the more we can make the experience positive, I think the better off we will be.

STANFORD: What we are talking about here is in some ways a far cry from the traditional view of experimental psychology, where the whole experiment is a machine that you have to put the subject into, and then he is like a cog in a machine. I really believe that we can have the best of several possible worlds. I don't think we have to abandon good experimentation to have good humanistic concerns in our studies. Incidentally, parapsychology may have a contribution to make to psychology in this regard. I am convinced, talking with colleagues and students, that many psychologists including those who work with humans, not just animals, are quite unaware of the importance and the possible relevance even to the statistical power of one's own test, of bringing all subjects to a similar point of understanding and emotional response, to the experimental setting, because that reduces error variance in a study. When you mention this to most psychologists they look at you as if you came from the planet Mars. They are likely to ask, "Don't you know that standardization is the name of the game?" My question is, "What are we

standardizing, the stimuli or where the subject starts in doing the study?" Even rat learning people know that you have to put rats into the maze beforehand to let them run around and get familiarized with it. Some will take ten minutes to do that, some will take twenty minutes, some will take two minutes to do that—but you let them have their own time in doing it. Yet, we seem not to recognize this in dealing with humans.

SCHMIDT: I would like to clarify Marilyn's question that asked "Does psi occur at the time of the session or at the time of the feedback?" To understand this you don't need quantum theory. I think the general question is this: if you have an ESP experience, how does it get into your brain? And I think one quite acceptable tentative explanation is that you don't go out there, it happens just in your brain; you see in your brain, what later you will perceive. Supposing you have four pictures to guess which one is right, then perhaps you don't mystically reach out to where the pictures are. What you aim at is what you are shown later, when the experimenter says "This is the right picture." If this is true, then it would be very important not to show the person all the pictures, but just the target picture. Could one by this procedure avoid the mix up?

STANFORD: Yes, I do understand *that* point. And, indeed, that was precisely the perspective which I was suggesting in my paper. Maybe I didn't make it clear in my presentation, but the thought was indeed that what is central here is the feedback to all the pictures. I don't see how else you get displacement, if you are talking about something happening contemporaneously while you are sitting there in the Ganzfeld. It doesn't make any sense that people are psychically running all over the lab looking at all the pictures. But the central idea is that the pictures are all seen together, at some point in the experiment. From that perspective then, you set up the kind of study that I proposed, for one group at least. All they see at session's end is a target, just to give a feedback about the target, but someone else does the outside judging.

HONORTON: This is something we hope to get at with the experimenter judging procedure as a comparison, because there the receiver will only see the actual target. Now, informally it sometimes appears to us that the subject latches on to whatever the first target is that he sees during the judging procedure. This would make a certain amount of sense simply from the standpoint of the effects of sensory deprivation and the stimulus hunger that builds up during that deprivation. It leads a person to be much more influenced by

whatever the first thing is that he sees, when he comes out of the Ganzfeld. We hope to get at that question.

SCHLITZ: Important for designing experiments is the question: What if you don't have feedback to the subject? Helmut articulated the one side of the coin which was perhaps that psi comes at the feedback level. Perhaps it is the subject responding to what he is going to see in the future. Okay, what is he actually going to see? The other side of the coin is that the ESP impressions come during the time when the participants are sitting in the Ganzfeld situation, for example. And that would have implications in terms of our development of a target pool prior to the experimental situation. In other words, if the subject is displacing during the experimental period, then he could be displacing to those potential future targets. And a way of getting around that is to choose your target pool after the experiment. You have your one target and then you have your pool that is developed afterwards, so that there is a clear channel between the subject and the target, having nothing to do with the feedback. The feedback question is then a separate question.

STANFORD: It is certainly a study that is worth doing. I don't know whether any of us can guess what the results would be ahead of time, but it is one of those variables you can look at.

BERGER: Marilyn, are you suggesting that the target for the session be generated after the receiver has mentated?

SCHLITZ: Well, that is one possibility. That would be precognitive. No, what I am suggesting here is that the target be generated, be created or in existence prior to the experiment and during the experimental episode. In a classic example, you have an agent staring at a target picture. Oftentimes in experimental set ups there is a target pool that you put together before the experiment. That target pool is already there and if the psi is occurring during the experimental episode then there is the potential that the subject is reaching out to that target pool and that there is a great source of displacement during the psi episode. One way of getting around that is to have only one target and then build your target pool afterwards. Of course, it would have to be someone who hadn't heard the mentation, but that is one way of getting around that if it is the problem. And that is why I think it is important to make that distinction.

RAO: How would you rule out precognition?

SCHLITZ: In terms of this potential protocol? The precognition comes in. It is the question of the observation matter again. I am thinking in terms of just a classic approach to the subject describing

during real time what the target is. If there are four targets existing out there, it might be easier for the subject to pick up all four of them than to associate just one. That doesn't deal with precognition.

HONORTON: It is not really very useful to engage in detailed speculation of what kind of psi is taking place before we are able to produce psi in general, unless there is some specific reason that would increase the strength of the effect. I think a lot of time was perhaps not used as well as it could have been in the 1940's in the parapsychological equivalent of how many angels can dance on the head of a pin, whether it is precognition, telepathy or clairvoyance.

STANFORD: I feel the same way. I think that the history of psi research has shown us that it isn't too profitable to look at these differentiations, and we haven't been able to succeed in separating precognition and this and that. The things that have seemed to make a difference so far are things that really have some kind of psychological impact on the subject. Now, I may well be wrong. I am not saying that the experiment isn't worth doing, but for my own priority I would put the more directly psychological variables first, especially in terms of trying to enhance our replicability, which is the focus of this specific discussion.