CONFRONTING THE EXPERIMENTER EFFECT

JOHN PALMER

The Experimenter Effect and Replicability

The experimenter effect (EE) in parapsychology is closely linked to the broader problem of replicability. If the experimenter's identity or behavior is a crucial factor in determining whether a psi experiment succeeds or fails, then replication by a wide range of different investigators will be difficult to achieve unless whatever causal factors the EE represents can be identified and controlled.

Although formal survey data are lacking, it is widely acknowledged by both parapsychologists and their critics that some investigators and laboratories have been more successful than others in eliciting psi effects. Such consistency is often evident across a variety of test procedures and with selected as well as unselected subjects. This is not to deny that some cross-laboratory replicability does exist in parapsychology, but successful results are not distributed evenly enough within the community of psi researchers to inspire a great deal of confidence that comparable rates of success would occur outside this community.

This more broadly based replicability is essential if scientists outside of parapsychology are to be persuaded that psi anomalies are worthy of their attention, let alone that they are paranormal. Replicability within the fairly circumscribed parapsychological community is unlikely to have much impact. Such recalcitrance by our scientific brethren may be unjustified and even unfair, but that is beside the point. If parapsychology is to compete in the larger scientific arena, we must have procedures that we can export to outside scientists with reasonable assurance that they will succeed at the same rate they do within parapsychology. This does not mean that we must accede to the demand that hostile conventionalists be able to replicate our results, but it does mean that a healthy cross-section of benevolent conventionalists and neutrals be able to do so.

Two more specific factors can be mentioned that suggest to me the prospects of exporting psi success may not be as great as is sometimes assumed. First, some of the statistical methods currently being used in our meta-analyses and other literature reviews may be giving us inflated estimates of the degree of replicability that exists *within* parapsychology, at least in the sense that is relevant to the export issue. A case in point is the Stouffer Z statistic (Rosenthal, 1978). This test gives a valid and powerful estimate of the degree of statistical significance within a data base. Along with the complementary file-drawer statistic (Rosenthal, 1979), it can obviate objections that unreported nonsignificant studies in the parent population can wash out the results of the significant ones. However, it does not give a very good indication of how significance is dispersed within a population. Like the traditional CR method, it provides no empirical estimate of the variability of outcomes among experimental units. It is quite similar to the sum-of-squared-CR method introduced by Rhine and Pratt (1957), except of course that with Stouffer's Z the component Zs are not squared and the sign of the deviation is thus taken into account. However, in both cases one very large component can result in a significant summed value even when balanced by several components near zero. Indeed, it is this very characteristic that makes file drawer analyses so powerful.

To illustrate that statistics like Stouffer's Z can produce inflated estimates of replicability, consider the 28 direct-hit ganzfeld studies recently reviewed by Honorton (1985). Using Stouffer's Z, the aggregate result of these studies is highly significant ($Z = 6.60, p < 10^{-9}$), 1-tailed, indicating (quite properly) strong evidence for psi. The result is essentially the same when component Zs are computed for all the studies conducted by a given investigator and summed over investigators ($Z = 6.16, p < 10^{-9}$).

Compare the above significance levels to those obtained using the single-mean t test, a method that does provide an empirical variance estimate. Using the *experiment* as the unit, the outcome is still very significant, although not quite as significant as with Stouffer's Z ($t(27) = 4.22, p < .0005$). However, when one takes the *experimenter* as the unit, the significance is much less than what is obtained with the comparable Stouffer Z and drops into the marginal range ($t(9) = 2.58, p < .02$). Moreover, the mean component Zs of three of the 10 investigators are negative.

As noted by Stanford and Palmer (1972), it is this empirical variance reflected in the error term of the test statistic that allows the statistic to serve an inferential function, i.e., to support generalization beyond the sample. Of course, it is this generalization that is at the heart of the export issue. While even the t-test analysis offers encouragement in this regard, it is much more equivocal than the encouragement one might draw from looking at the Stouffer Z.

Again, let me stress that the above discussion is not meant to deny the obvious statistical significance of the ganzfeld database, nor does it deny that within the sample the results cannot be wholly attributed to one or two investigators. Nonetheless, one should be cautious in promoting the ganzfeld as a widely replicable technique for eliciting psi effects until success is demonstrated on a broader scale. Although Honorton (1985) and other researchers have not made this claim, terms like "replicable" could be interpreted as implying it unless they are qualified.

A second, and more fundamental, reason for caution about the exportability of psi success is that the population of parapsychologists is probably not representative of the broader population of scientists (even excluding hostile conventionalists) on variables that many parapsychologists believe are relevant to psi outcomes. Parapsychologists, for example, are more likely to have positive attitudes toward psi, to be absorbed in the subject matter, and to be more highly motivated to achieve positive results than even benevolent outsiders. Thus even if one views the replication rate within parapsychology as providing a sound basis for generalizing to other parapsychologists, it is not really a sound basis for generalizing beyond that.

I do not wish to imply that parapsychologists are unaware of or insensitive to the replicability problem in parapsychology. Indeed, a recent Parapsychology Foundation conference was devoted entirely to this issue (Shapin & Coly, 1985). Likewise, attention has been paid to the EE and some progress has been made in coming to grips with it. Three major parapsychological explanatory frameworks have been proposed: differences in subject populations, differences in the way experimenters interact with their subjects ("experimenter psychology"), and experimenter psi.[1]

Subject Populations

One possibility, suggested by Honorton (personal communication), is that the EE, at least in part, might be an artifact of the tendency of different experimenters to work with different subject populations. There has been some suspicion for years that ordinary college students

[1] There, of course, is also the conventionalist interpretation of the EE in terms of error and fraud by the experimenter. I have chosen not to address this possibility in the present paper, although I have done so elsewhere (Edge, Morris, Palmer, & Rush, 1986).

are, generally speaking, not a very good population for psi tests, and that better results are more likely to be obtained by concentrating one's efforts on volunteers who are especially interested in psi, practice mind-development techniques, are "successful" in life or their profession, or are engaged in occupations like art or clinical psychology that capitalize upon intuitive skills.

An important asset of this hypothesis is its easy testability. It can be confirmed by showing (a) that subject populations at successful and unsuccessful laboratories differ on certain operationally defined characteristics and (b) that psi scores at each of these laboratories are correlated with these characteristics.

A test of this hypothesis is currently underway involving Psychophysical Research Laboratories (PRL) (a successful laboratory with the ganzfeld) and Foundation for Research on the Nature of Man (FRNM) (a not-so-successful laboratory with this technique). Honorton and Schechter (1986) have shown that first-time subjects in the ganzfeld at PRL who have participated in other psi tests, practice a mental discipline, had many different types of psi experiences, and have a certain profile on the Myers-Briggs Type Indicator (MBTI) are more likely to score positively than other subjects. (Schmidt & Schlitz, 1988, who work at a laboratory noted for success in REG PK experiments, have recently replicated the MBTI pattern for this latter type of experiment). Kanthamani is currently in the process of analyzing comparable data from ganzfeld subjects tested at FRNM. It is already evident from preliminary inspection of the data that the PRL and FRNM samples differ on relevant demographic characteristics. Should the significant mean differences and correlational effects found at PRL be confirmed by formal analyses of the FRNM sample, a major piece of supporting evidence for this interpretation of the EF will be uncovered, at least insofar as the ganzfeld is concerned.

Experimenter Psychology

There is a good deal of experimental evidence suggesting that the ability of experimenters to make their subjects motivated to succeed, confident, and at ease in the test situation facilitates psi scoring (White, 1977). However, exactly what is crucial about the subject-experimenter interaction is not so well understood. Research in psychology on experimenter bias effects, such as the work of Rosenthal and his colleagues (Rosenthal, 1966), suggests that such factors might be quite subtle and include nonverbal behavior.

These contributing factors are likely to be mediated to some degree by the experimenter's own belief in psi and its likelihood of appearing in the experiment. This could be a stumbling block in getting other scientists to successfully replicate psi experiments, as noted previously. One solution, suggested to me by Truzzi, would be to have such persons (including, in particular, unsympathetic conventionalists) farm out the actual testing of subjects to successful psi experimenters while maintaining in their own hands crucial control features such as the target order. A prototype for such an approach has already been developed (Schmidt, Morris, & Rudolph, 1986). Care would need to be taken, however, to be sure that the unsympathetic investigator does not create a negative climate that could adversely affect the mood or confidence of the experimenter and/or subjects. One case where this may have happened in the past was the unsuccessful attempt by Foulkes and colleagues to replicate the Maimonides dream experiments (Van de Castle, 1977). Such "climatological" factors might also help explain why different experimenters at the same laboratory often seem to get the same kinds of results; they absorb the attitudes and confidence level of the principal investigator or lab director.

A frequently overlooked aspect of this problem is the effect of experimenters' past track records on the attitudes with which they approach subsequent experiments. Experimenters who have gotten significant results in the past, especially if they have done so frequently, will likely approach their next experiment with an air of confidence that is bound to be communicated to their subjects. By the same token, unsuccessful experimenters are likely to find it difficult to avoid communicating a corresponding lack of confidence, however hard they may try. In other words, success breeds success and failure breeds failure. This factor might help explain why the divergence of results at different laboratories seems to be so stable over time.

One sanguine observation that can be made about all this is that experimenters' beliefs and attitudes are not communicated to subjects directly but indirectly by means of verbal and nonverbal behavior. If we knew more precisely what those behaviors are, we might be able to train experimenters to act accordingly, whatever their underlying attitudes happen to be. Alternatively, we might be able to find experimenters with the desired attitudes who act this way naturally. This obviously is easier said than done, but the idea is worth exploring. Specifically, we might consider creative use of videotaping. A good first step was taken several years ago in experiments where students rated videotaped excerpts of psi-conducive and psi-inhibitory experimenters speaking at a Parapsychological Association convention (Schmeidler & Maher, 1981; Edge & Farkash, 1982). It would be more useful, however, to have videotapes of such experimenters actually conducting a test session, including the all-important orientation or get-acquainted period. Analysis of such videotapes, informed by what social psychologists have already learned about factors influencing attribution of traits, attitudes, etc. might provide hypotheses that could be tested by systematically manipulating the experimenter behavior variable (either between or within experimenters) and observing its effects not only on psi scores, but also on various behavioral, cognitive, affective, and even physiological variables. Such a research program would be difficult and time-consuming, but it might also be very important in helping us to conquer the EE. Other useful suggestions for tackling this problem have been offered by Stanford (1985).

Experimenter Psi

The other major alternative interpretation of the EE is psi input by the experimenter. Traditionally, it has been assumed (implicitly, at least) that conscious intention and effort are necessary to produce psi in the laboratory. In an REG experiment, for example, where a subject and experimenter are seated side-by-side and the experimenter instructs the subject to attempt to bias the output of the machine paranormally, it was assumed that the subject is the only potential psi source, the reason being that only the subject is making a conscious effort to influence the machine. This linkage of psi to conscious effort was never very plausible, for a number of reasons. First, it was never supported by a theoretical rationale and, in fact, much emphasis has traditionally been placed on the unconscious nature of psi (Rhine & Pratt, 1957). Second, psi in the real world (e.g., poltergeists) does not seem to involve conscious effort or even conscious intent on the part of the agent or percipient.

However, what should have destroyed once and for all the linkage between psi and conscious effort or intent is the solid body of evidence in support of covert psi effects (Schechter, 1977; Stanford, 1977). In what is perhaps the cleanest demonstration of covert psi, Stanford, Zenhausern, Taylor, and Dwyer (1975) showed that an REG can be biased when *no one* is trying to influence it consciously, but the REG output has positive or negative consequences for the subject. What makes such findings relevant to the EE is that the output also has positive or negative consequences for other people involved in the experiment, most notably the experimenter. In the experiment of Stanford et al. (1975), it is at least intriguing in this connection that the more successful of the two experimenters was an extrovert, and there is considerable evidence from other lines of research that extroverts score better on ESP tests, at least, than introverts (Sargent, 1981). The main point, however, is simply that there is as much reason to consider the experimenter the psi source as any of the designated subjects. I am certainly not the first person to point this out (e.g., Kennedy & Taddonio, 1976; Stanford et al., 1975; White, 1976), but I still do not think its importance has fully sunk in among parapsychologists. If one thinks in terms of burden of proof, the burden should fall on those who maintain that the experimenter is not making some paranormal contribution to significant psi outcomes that are otherwise deemed of paranormal origin.

It is not as easy to conceptualize how the experimenter could be a psi source in ESP experiments as in PK experiments, because in the former the subject's conscious experience is by definition an essential ingredient in the exercise. Recently, the favored solution to this problem has been the observational theories (Millar, 1978), which treat all psi, in effect, as PK. In other words, observation of the data by the experimenter, which must occur at one level or another at some point in any psi experiment, retroactively causes the brain of the subject to create a mental image related to the target. However, the OTs are quite controversial (Braude, 1979; May, Radin, Hubbard, Humphrey, & Utts, 1985; Varvoglis, 1986; Walker, 1984, 1987). Indeed, the retro-PK evidence that serves as the main empirical foundation of the OTs can be accounted for (in my opinion, more parsimoniously) as real-time PK on the part of the experimenter or some other interested party.

However, one need not resort to the OTs to account for experimenter psi in ESP experiments. One alternative is Stanford's (1978) conformance behavior model; one would simply make the reasonable assumption that the experimenter is a "disposed system." Those of a more traditional bent can fall back on simple, albeit unconscious, ESP and PK. The experimenter, for example, might acquire knowledge of a target by clairvoyance and communicate it to the subject by active telepathy, or Stanford's (1974) MOBIA. In a personality-ESP study, he or she might also learn by clairvoyance who, say, the extroverts are and send the target information exclusively to them. More exotic mechanisms are also possible. To give just one example, an experimenter might unconsciously bias the timing of a ganzfeld trial so that the subject's (nonparanormal) stream of consciousness is intersected at a point when it happens to coincide reasonably well with the target. (This particular mechanism would not be plausible if the hit were striking, however.)

The distribution of psi. I find a tendency among parapsychologists to view experimenter psi as an all-or-none affair: either all the psi in a given experiment comes from the subject(s) or it all comes from the experimenter. Although this could ultimately prove to be the case, I think our best guess at this point is that the subject(s) and the experimenter contribute jointly to most outcomes.

The notion that psi ability is restricted to a small percentage of the population (Millar, 1979) has never appealed to me. Although there are isolated exceptions (e.g., idiot savants), most human abilities are distributed in the population in a manner roughly approximating the normal curve, whether or not they correspond to it exactly. Therefore, I think the odds strongly favor psi ability falling into this category. This does not mean, however, that in a given experiment the results may not be attributable largely if not entirely to one individual. First, even the normal-curve model allows for great individual differences in psi ability. Second, situational factors may dictate whether, or to what degree, this ability manifests in a particular experimental session.

Is there any way to tell what the relative strength of the experimenter's input is likely to be? One approach would be to examine the research literature to glean what variables seem to influence psi by subjects. The problem, of course, is how to be sure that the results of *these* experiments are not due to experimenter psi. For example, earlier in this paper I alluded to the review of extroversion-ESP research by Sargent (1981) to support the conjecture of experimenter psi in the Stanford et al. (1975) REG experiment. It is not entirely unreasonable to speculate that many of the studies Sargent reviewed themselves involved experimenter psi; certainly many of them were conducted by experimenters who had excellent track records in other types of experiments, most notably Sargent himself.

This, of course, illustrates the quintessential Catch-22 that the experimenter psi hypothesis confronts us with. There is no sure way out, but if we are willing to grant that subjects do contribute something, at least, to psi outcomes (which I think is reasonable), then it is also reasonable to suppose that some of the more replicable findings of process-oriented psi research are telling us something real about the factors that mediate psi success.

What, then, should we (tentatively) conclude from this research? The early observations of Rhine and Pratt (1957) and the covert psi research strongly suggest that needs and motivations are important. This factor clearly implies relatively strong psi input by the experimenter, because (at least with unselected subjects) the need of the experimenter to get positive results (or to confirm a hypothesis) is on the average likely to be greater than whatever needs the subjects bring to the task.

However, the experimenter may have other needs as well that may or may not support a positive result. If the psi process is really unconscious, some of the most important needs may themselves be of an unconscious, psychodynamic nature. If psi ability is fear-evoking at a deep level as some have suggested (e.g., Tart, 1984), then the experimenter's chronic or episodic level of ego-strength may be relevant. For example, experimenters may be prone to unconsciously inhibit their own psi at times in their lives when their self-concept is vulnerable to some personal crisis, even if consciously they may feel highly motivated to succeed. This same argument, of course, applies to subjects. The point is that assessment of needs may be more complicated and require more in-depth analysis than is often supposed. In any case, needs must be more precisely defined for experimental purposes (Weiner & Geller, 1984).

As far as the personality literature is concerned, we might expect experimenter psi to be more likely if the experimenter is an extrovert and accepts the reality of psi. As for situational variables, we might expect experimenter psi to occur most strongly when the experimenter is in a spontaneous, labile mental state, especially if this follows attention or concentration on the task (Stanford, 1974). I would argue further that this need not involve a conscious intent to use psi to influence the task outcome. It is not unlikely that such conditions are frequently met by experimenters in psi research (Kennedy & Taddonio, 1976).

The chaos model. If one carries the above thesis to its logical conclusion, the result can be both mind-boggling and depressing. This fact was vividly brought to our attention in an article written 25 years ago by Eisenbud (1963):

> Experiments are conducted on the curious assumption that the subjects in them will not use the very faculties they are being tested for (and over which, presumably, they have as little control as they have over the weather) until they step across the threshold of the laboratory and hear the starting gong, and that then they will use those faculties only within the confines of their designated roles in the particular design employed. . . .
>
> By the same token it seems implicitly to be taken for granted that experimenters (or "independent" judges or checkers or raters, for their part) will not, for whatever obscure reason, use any psi faculties *they* may have to muddy the field . . . Everyone behaves, in short, as if there were some sort of gentleman's agreement committing subjects, experimenters, judges and other participating personnel to

> stick faithfully to their assigned roles in the experiment as scripted and to neither take any notice of nor infringe upon what any of the others are doing. (p. 258)

Eisenbud's remarks bear directly on the source-of-psi problem and reveal that it is even more far-reaching than I have implied so far. Up to this point I have been writing as if the only people we had to contend with were the subject(s) and (one) experimenter. In reality, there are other persons who may be as likely if not more likely psi sources than the experimenter; one prominent candidate is a principal investigator who is not actually testing the subjects, but is psychologically involved in the experiment.

Two other factors also add to the chaos:

1. *Space-time independence.* Although this matter remains controversial (Vassy, 1988), there is still little good evidence that the psi process is constrained by physical space or time. Even if there are such constraints, it is unlikely that they are so severe as to preclude a wide range of opportunities for various individuals to inject psi into an experiment. It is certainly unrealistic to suppose that the input would have to come at the time of a given experimental trial.

2. *Goal-directedness.* Several parapsychologists have made the point that psi seems to be goal-directed, in the sense that the complexity of the task seems to have little bearing on the likelihood of its accomplishment (e.g., Kennedy & Taddonio, 1976; Schmidt, 1975; Stanford, 1978). An important implication of this conclusion for the experimenter psi hypothesis is that single "bursts" of psi could have far-reaching effects on an entire experiment.

An example is the possibility that psi could determine an entry point in a random number table that produces a target sequence corresponding well enough with the response biases of a group of ESP subjects to yield a significant result. If one were to consider the amount of information one would have to process to achieve such an outcome by logical methods, the task would be staggering. But the assumption of goal-directedness precludes such considerations. Using complicated methods to determine the entry point helps not at all in light of this assumption. Although in the classic experiment on this problem, Morris (1968) found that the dice-throwing method obviated the bias, one could argue reasonably that Morris and his experimenter simply got the outcome they wanted or expected. It does not even stretch the principle too far to assume that the bias could be effected to some degree at least when a single entry point is selected for a series of experiments, as recommended by Stanford (1981), at least when effect sizes are small. This is not to suggest that we should not do all we can to frustrate such biases when the costs are not otherwise debilitating, but let's not be too sanguine about the success of such efforts.

The above considerations, if applied in full, lead one to predict total chaos in psi research. Countless individuals with all kinds of conscious and unconscious needs and motives would be competing to affect the outcome of an experiment, either positively or negatively. Subjects could use their psi ability not only to influence their own results, but also the results of other subjects, for example, by telepathically beaming incorrect target information (which they would not necessarily need to do at the time of the trial); in other words, the inputs of individual participants could be interactive instead of additive. The experimenter, principal investigator, lab assistants, secretaries, outsiders with interest in the outcome, all would add their input to the mix. Whatever psychological or physical factors mediate psi could not be anywhere near effectively controlled on such a broad scale. Chaos, indeed.

This scenario, however, points to the problem with the chaos model. Our experimental results in parapsychology, elusive as they often seem to be, are just not *that* chaotic. Various meta-analyses and more traditional literature reviews on topics such as the ganzfeld (Honorton, 1985), REG research (Nelson & Radin, 1988), and personality correlates (Sargent, 1981) reveal at least a rudimentary lawfulness in psi data. Ironically, perhaps the best single line of evidence contradicting the chaos model is the EE itself! In other words, the chaos model simply does not square with the data, which means, to put it bluntly, that the chaos model is wrong.

This implies that we must put some constraints on the scope of the various causal factors—number of psi sources, space-time independence and/or goal directedness—that uphold the chaos model. Unfortunately, we have no sound theoretical or empirical basis for deciding what those constraints should be. I would therefore suggest that we make our choice on pragmatic grounds. In other words, we should postulate the minimum constraints that would realistically allow empirical research on the source-of-psi problem, including tests of the experimenter psi hypothesis.

In this spirit, I would propose the following two constraints: (a) potential psi sources are restricted to those who are psychologically involved in the experiment, i.e., persons who are aware of it, think about it, and consider it meaningful or important; (b) psi sources can only affect an outcome (trial, target generation, etc.) within one hour before or during this outcome. These proposed constraints are obviously somewhat arbitrary and perhaps could be improved upon. Observational theorists, for example, would surely want to add a post-outcome time period for persons who observe the data in some form. However, we must agree on some set of such constraints if we hope to tackle this problem empirically.

Testing the experimenter psi hypothesis. The imposition of pragmatic constraints on the chaos model renders the experimenter psi hypothesis testable. Either a correlational or experimental approach could be adopted. The correlational approach in its broadest sense would require that experimenters observe and record during the appropriate time frame aspects of their mental states that could be relevant to the production of paranormal effects, including their levels of need or desire for a successful outcome, moods, general level of arousal, and activities that might reflect cognitive lability and opportunity for release-of-effort effects. Manipulating these variables, while desirable in principle, could sometimes prove disruptive and interfere with the experimenter's other duties. Also, some of these variables, like moods, would be hard to manipulate. Thus a correlational approach might generally make the most sense at the early stages of inquiry.

On the other hand, experimenters, particularly successful ones, might profitably consider manipulating their own mental states systematically to see if this affects the results of their experiments. Schlitz (1987) has noted that some successful psi experimenters intentionally put themselves in mental states that could be considered psi-conducive prior to or during their experimental sessions. It would be interesting to see what would happen if these experimenters treated such states as an experimental variable. Other predictor variables associated with psi success could also serve as the basis for experimental manipulations and testable hypotheses.

A related strategy is more holistic: to simply treat the experimenter as another subject. In a PK experiment, for example, the experimenter might actually join the subject in attempting to influence the REG, perhaps even acknowledging this participation openly. The experimenter would also fill out the same psychological tests as the subject and undergo the same experimental treatments. I am not suggesting here that conscious intent by the experimenter is necessary after all; what this approach accomplishes is to facilitate comparison between the experimenter and subjects by making their experiences and activities more uniform. This admittedly radical strategy has the added advantage from the point of view of the experimenter psychology hypothesis of helping to break down the barrier between the roles of "subject" and "experimenter."

But is such an approach desirable? Stanford (1981), for example, has warned us about the need to make up our minds whether we want to be "shamans" or "scientists". Indeed, should successful experimenters whose interests lie elsewhere worry about experimenter psi at all? There is something to be said for the recently overused adage: "If it ain't broke, don't fix it." If experimenter psi is a factor in the success of some experimenters, preoccupation with it might even create a psi-inhibitory mental set or experimental atmosphere that could undermine that success.

This attitude might be short-sighted, however. If experimenter psi is a factor in psi outcomes—and I believe there are good reasons to presume that it is (as I argued above) —failure to identify and deal with it could adversely affect the rate of replication at other laboratories, which (as I also argued above) is what ultimately will vindicate the experimenter's own success.

Interactions and Confounds

Although I have devoted more space to experimenter psi than to the other interpretations of the EE, they all deserve serious and equal attention. My guess is that all three have some validity. They also interact with each other: for instance, if subjects are selected who have relatively great psi potential and are treated in such a way as to maximize their motivation, confidence, and ease, the relative psi input of the subjects vis-à-vis the experimenter or other interested parties is likely to be enhanced, as is the overall level of psi.

Second, it is not always easy to tease these interpretations apart experimentally. As an example, consider once again the experiment of Stanford et al. (1975) for which I had raised the possibility of experimenter psi by the extroverted experimenter. The authors, while acknowledging the possible role of experimenter psi in their study, interpreted this particular effect according to the experimenter psychology hypothesis: the extroverted experimenter may have facilitated a more psi-conducive social interaction with the subject. We have no way of knowing which interpretation is best. As a second example, consider my suggestion that experimenters might manipulate or control their own mental states to influence their own psi input. The problem here is that such manipulations carry the risk of affecting the experimenter's social interaction with the subject, thus providing a potential confounding of the experimenter psi and experimenter psychology interpretations of the EE. Any research program that seeks to deal incisively with the EE must keep these potential confounds and interactions in mind.

Conclusion

As I look back over this paper, I question if I have said much that is really new. My main purpose in writing it, however, was not so much to break new ground as to sensitize my colleagues to what I see as one of the most important problems facing modern parapsychology. Although parapsychologists are certainly aware of the problem, and almost every discussion section of an experimental report nowadays seems to allude to the EE at least in passing, I have seen little evidence that most of us are yet ready to confront it head on. My main thesis is that until this attitude changes, we are not likely to see the breakthrough that many of us have devoted the better part of our professional lives to achieving.

REFERENCES

Braude, S. E. (1979). The observational theories in parapsychology: A critique. *Journal of the American Society for Psychical Research*, *73*, 349–366.

Edge, H., & Farkash, M. (1982). Further support for the psi-distributed hypothesis. *Research in parapsychology 1981* (171–172). Metuchen, NJ: Scarecrow Press.

Edge, H. L., Morris, R. L., Palmer, J., & Rush, J. H. (1986). *Foundation of parapsychology*. London: Routledge.

Eisenbud, J. (1963). Psi and the nature of things. *International Journal of Parapsychology*, *5*, 245–273.

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

Honorton, C., & Schechter, E. I. (1986). Ganzfeld target retrieval with an automated testing system: A model for initial ganzfeld success. *Proceedings of the Parapsychological Association 29th Annual Convention*, Rohnert Park, CA, 401–414.

Kennedy, J. E., & Taddonio, J. L. (1976). Experimenter effects in parapsychological research. *Journal of Parapsychology*, *40*, 1–33.

May, E. C., Radin, D. I., Hubbard, G. S., Humphrey, B. S., & Utts, J. M. (1985). Psi experiments with random number generators: An informational model. *Proceedings of the Parapsychological Association 28th Annual Convention*, Medford, MA, 235–266.

Millar, B. (1978). The observational theories: A primer. *European Journal of Parapsychology*, *2*, 304–332.

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

Morris, R. L. (1968). Obtaining non-random entry points: A complex psi task. In J. B. Rhine & R. Brier (Eds.), *Parapsychology today* (pp. 75–86). New York: Citadel.

Nelson, R. D., & Radin, D. I. (1988). Statistically robust anomalous effects: Replication in random event generation experiments. *Proceedings of the Parapsychological Association 31st Annual Convention*, Montreal, Canada, 75–86.

Rhine, J. B., & Pratt, J. G.. (1957). *Parapsychology: Frontier science of the mind*. Springfield, IL: Thomas.

Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts.

Rosenthal, R. (1978). Combining results of independent studies. *Psychological Bulletin*, *85*, 185–193.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*, 638–641.

Sargent, C. L. (1981). Extraversion and performance in "extra-sensory" perception tasks. *Personality and Individual Differences*, *2*, 137–143.

Schechter, E. I. (1977). Nonintentional ESP: A review and replication. *Journal of the American Society for Psychical Research, 71*, 337–374.

Schlitz, M. (1987). An ethnographic approach to the study of psi: Methodology and preliminary data. *Research in parapsychology 1986* (103–106). Metuchen, NJ: Scarecrow Press.

Schmeidler, G. R., & Maher, M. (1982). Judges' responses to the nonverbal behavior of psi-conducive and psi-inhibitory experimenters. *Journal of the American Society for Psychical Research, 75*, 241–257.

Schmidt, H. (1975). Toward a mathematical theory of psi. *Journal of the American Society for Psychical Research, 69*, 301–319.

Schmidt, H., Morris, R., & Rudolph, L. (1986). Channeling evidence for a PK effect to independent observers. *Journal of Parapsychology, 50*, 1–15.

Schmidt, H., & Schlitz, M. (1988). A large scale pilot PK experiment with pre-recorded random events. *Proceedings of the Parapsychological Association 31st Annual Convention*, Montreal, Canada, 19–35.

Shapin, B., & Coly, L. (Eds.) (1985). *The repeatability problem in parapsychology.* New York: Parapsychology Foundation.

Stanford, R. G. (1974). An experimentally testable model for spontaneous psi events. II. Psychokinetic events. *Journal of the American Society for Psychical Research, 68*, 321–356.

Stanford, R. G. (1977). Conceptual frameworks of contemporary psi research. In B. B. Wolman (Ed.), *Handbook of parapsychology* (pp. 823–850). New York: Van Nostrand Reinhold.

Stanford, R. G. (1978). Toward reinterpreting psi events. *Journal of the American Society for Psychical research, 72*, 197–214.

Stanford, R. G. (1981). Are we shamans or scientists? *Journal of the American Society for Psychical Research, 75*, 61–70.

Stanford, R. G. (1985). Toward the enhancement of inter-laboratory and inter-experimenter replicability in psi research. In B. Shapin & L. Coly (Eds.), *The repeatability problem in parapsychology* (pp. 212–237). New York: Parapsychology Foundation.

Stanford, R. G., & Palmer, J. (1972). Some statistical considerations concerning process-oriented research in parapsychology. *Journal of the American Society for Psychical Research, 66*, 166–179.

Stanford, R. G., Zenhausern, R., Taylor, A., & Dwyer, M. A. (1975). Psychokinesis as psi-mediated instrumental response. *Journal of the American Society for Psychical Research, 69*, 127–133.

Tart, C. T. (1984). Acknowledging and dealing with the fear of psi. *Journal of the American Society for Psychical Research, 78*, 133–143.

Van de Castle, R. L. (1977). Sleep and dreams. In B. B. Wolman (Ed.), *Handbook of parapsychology* (pp. 471–499). New York: Van Nostrand Reinhold.

Varvoglis, M. P. (1986). Goal-directed and observer-dependent PK: An evaluation of the conformance-behavior model and the observational theories. *Journal of the American Society for Psychical Research, 80*, 137–162.

Vassy, Z. (1988). Distance, ESP, and ideology. *Behavioral and Brain Sciences, 10*, 616–617.

Walker, E. H. (1984). A review of criticisms of the quantum mechanical theory of psi phenomena. *Journal of Parapsychology, 48*, 277–332.

Walker, E. H. (1987). Intuitive data sorting vs. quantum mechanical observer theories. *Journal of Parapsychology, 51*, 217–227.

Weiner, D. H., & Geller, J. (1984). Motivation as the universal container: Conceptual problems in parapsychology. *Journal of Parapsychology, 48*, 27–37.

White, R. A. (1976). The limits of experimenter influence on psi test results: Can any be set? *Journal of the American Society for Psychical Research, 70*, 333–369.

White, R. A. (1977). The influence of experimenter motivation, attitudes and methods of handling subjects in psi test results. In B. B. Wolman (Ed.), *Handbook of parapsychology* (pp. 273–301). New York: Van Nostrand Reinhold.

DISCUSSION

HONORTON: John, I have a couple of comments. My own position has been that the absolutely terrible way in which we have tended to characterize our subjects in reports could alone be the source of experimenter differences. I doubt that that is the entire source of experimenter differences. I am inclined to believe that conventional experimenter effects account for a great deal of the variability. But I think we certainly need to enforce a requirement that experimental studies characterize their subject populations to a much greater, more precise degree than has been done so far. The experimenter psi hypothesis is a hypothesis that I believe has very little direct evidence to support it. It is a hypothesis that, as you yourself mentioned, has been attractive because it is plausible given the lack of known boundary conditions. It is also the lazy way out. You say, "I did not get results so it must be your psi rather than your subject's psi." And I think it has had generally a stultifying effect, one that has inhibited research for many years. I would like to see a more focused approach as you yourself do. Finally, on the Stouffer Z t-test difference that you raised in the beginning of your paper, certainly it is true that the t test is a more appropriate way of looking at replicability across investigators rather than just summing the Z scores by the Stouffer method. But I would argue that you should not be comparing how significant it is. You should be looking at effect size once again. Your t of 2.58 with 9 degrees of freedom which is significant with a p of .015 is equivalent to an average effect per experimenter of .82 standard deviations above the expected mean. Cohen's D and an effect size of .82 in psychology is considered a very strong effect. Furthermore, you say that the mean component Zs of three of the ten investigators are negative. That is true. That is, however, like saying that the water glass in front of you is a third empty rather than that it is two-thirds full. Seventy percent of the investigators have positive Zs and I think that is the point to be drawn from that particular assessment. The main thing that I want to communicate here is that you should not compare these things on the basis of p values, but convert them into effect sizes because that gives you much more information about what is going on.

PALMER: As far as the descriptions of subjects are concerned, I certainly agree that we need better ones. I do happen to think that part of the experimenter effect may well be attributable to different characteristics among the subjects tested by different experimenters. I think I made that point in my paper.

As for there being very little direct evidence for the experimenter psi effect, there are certainly some supportive studies, such as the West and Fisk experiment, but there are not an overwhelming number of them. However, I think it is more than simply an *ad hoc* suggestion, as I attempted to point out in my paper. That is particularly true when you look at the results from the covert psi experiments, which I think provide indirect empirical support for the experimenter psi effect. But, again, I think that needs to be tested directly. I certainly agree that experimenter psi has been used in the past as a kind of a *post hoc* excuse, but that is not the way we have to use it. I would rather see us treat experimenter psi as an experimental hypothesis and test it directly, again, because I do believe there is a strong *a priori* likelihood that it is partly responsible for the experimenter effect.

With regard to the Stouffer Z, I think it is important to make a distinction between the magnitude of an effect, or an effect size, and its distribution. The issue in terms of replicability is not the magnitude of the average effect size. That could either be due, say to either four effect sizes of one, or two effect sizes of zero plus two effect sizes of two. These two outcomes have much different implications for the replicability issue. What is crucial in terms of replicability is not magnitude, but consistency. It is the estimate of consistency that I do not think Stouffer Z gives us very effectively. In terms of the three out of ten or seven out of ten, I agree that it's an example of the glass half empty or the glass half full, but I think part of the problem is that you have ten investigators. Unless you have very strong consistency, it is really hard to conclude anything from that analysis. A seven-three split with a sample size of ten is really not that much different from a fifty-fifty split in terms of statistical significance. I do think, by the way, that within the ganzfeld database, in particular, even when you do use what I consider the more appropriate statistics, there is still, good evidence of consistency. My point was not so much to question the conclusions about the ganzfeld, but to question the appropriateness of the Stouffer Z in drawing those conclusions.

SCHOUTEN: John, your talk focused on two important issues, that is the experimenter factor and its possible relevance for research and what you so nicely called the chaos model. Now as regards the second, you say, that everything points to chaos, but fortunately we see some order in that data and therefore the situation is not as bad as it seems. But it sounds a bit contradictory to me, because if you take the experimenter effect seriously, then in fact you are saying you can explain the order in chaos because some experimenters have some prejudice-isms but then indeed chaos reigns. Once experimenters get other ideas they will find other things, so I think that is not the reason to say the chaos model does not apply. I would say the chaos model does not apply because generally in science when you start, you start with a mess. In physics it took centuries before some really smart scientists found out what the regularities were. If you look at so called simple physical laws you can understand why. If there is a third power involved or four variables what an unbelievably difficult problem it must have been to sort it all out, considering also the noise. So I would suggest that you drop the chaos model for that reason. You just cannot say at the moment that there is chaos. We still have to look for regularities behind it. Your other point is the experimenter effect. I really think in the first place that there is weak evidence for it and, secondly, it is the easy way out. I think there are two other relevant points here. People tend to paint black and white pictures; some labs are successful, some labs are not. But I think the reality is much more complicated. In the first place, I think the labs who do have success have undoubtedly also run experiments which failed. I do not think there are 100% success rates. I know by experience that labs that are failing in this respect occasionally have experimenters or studies who turn out to be successful. Secondly, what bothers me a bit about the whole thing is that in practice most experiments are not carried out by the few experimenters we have in this field, but in reality are run by assistants, students and so on, who have most often different opinions than the experimenter who supervises the study. Unless you have these data you can not make a fair comparison. A third point is that I find it a bit overdone. If a subject is watching a screen in the RNG study or if you have a ganzfeld study where an agent is studying a target picture, we hope to see a little bit of psi in this situation. But then to assume that the experimenter who is outside that room will be equally able to influence the situation and the results is again assuming that a doubly difficult task will be as easy as a simple task. If the study is properly done the experimenter should be blind as to all the targets. To assume that he still has the same influence in the trial is to my taste a bit overdone. So I really think there are many reasons to be careful about the experimenter psi hypothesis.

PALMER: I concluded my paper denying that the chaos model applies. I do think that we need to put some constraints on it, and indeed it is those constraints that allow not only the experimenter effect but other parapsychological hypotheses to be tested. Ironically, as I pointed out in my paper, I think the best evidence against the chaos model is the experimenter effect, because it is an excellent example of order within parapsychology. It suggests, that the experimenter is having some kind of influence on the outcome, whether it be through his own psi or the way he treats subjects or the way he selects subjects or whatever. As far as the easy way out is concerned, I have to reiterate what I said to Chuck. I am not saying that we should use experimenter psi as a *post hoc* catch-all for anything that we cannot explain. I am simply saying that there is enough reason for us to take it seriously that we should investigate it more directly. The point was made that there is very little experimental evidence for the experimenter psi effect. To the extent that is true, it is not because, at least in my reading of the literature, that there has been a high failure rate among experimenters who have attempted to test the experimenter psi effect; rather it is because very few such experiments have been done, and that is my complaint. I am not trying to say that we should accept experimenter psi without evidence; I am simply saying that we should try to collect the evidence that will allow us to assess it, and that we haven't been doing enough of that.

As far as lab assistants are concerned, I think there is what might be called a lab ambiance that is different in some laboratories than in others. I am not sure who is responsible for that; perhaps if you had to assign responsibility to someone it would be to the principle investigator. But I think that ambiance does filter down to the experimental assistants or to other people who run the studies. I think there may be more uniformity there than perhaps you are assuming, even if the experimenters have somewhat different beliefs about psi. As for the implausibility of influences by experimenters who are blind or outside the room, if you are right, then the assumption of the goal-directed nature of psi for which there is some empirical support, is wrong. I see little reason, from what we know either from theory in parapsychology or from the experimental literature, to preclude someone in the next room who did not know the target, or even the experimental condition, from influencing the data. This is the kind of question that we need to investigate and not simply assume, as I think we have too often in the past, that if there is an effect, either it has to be the subject or, if it is the experimenter, then he has to be in the same room as the subject. Those are all assumptions that we have very little basis for making.

MORRIS: I was very glad to see you note that it is more that just other experimenters who may need to be taken into account. In fact, what you may have here is an entire experimental system all of which needs to be described much more thoroughly if we are going to do any kind of systematic hypothesis testing. It may be that what we are really talking about is the success or failure rate of an entire experimental system. We could try to erect as psi-conducive a system as we can, working with experimenters, trying to learn as much as we can about having them contribute positively to the system's overall results. The experimenter may well have contributed quite a bit in ways that we will gradually learn to be able to describe, so that when someone else tries to replicate, their experimenters will be in similar circumstances. An alternative would be to try to get the experimenters quite uninvolved, perhaps to train them to "keep their psychic hands off the data."

PALMER: I like very much your general point of looking at this with a systems approach. I did not get to it in my talk, but in my paper I put forth the radical suggestion that experimenters might actually participate in their own experiments as additional subjects. So instead of having an experimenter tell the subject to influence the RNG, both experimenter and subject attempt to influence it together. This might make the experimenter's behavior in the experiment more like that of the subject and make it easier to draw comparisons. The other approach is, of course, to try to isolate the experimenter from the situation as much as possible. I think this would be a very good thing to manipulate systematically and see what happens. It comes back to the point I was making before, that we need to take these possibilities seriously and start testing them.

ADAMENKO: Maybe the problem is how to control for the experimenter effect, how to decrease it. This is my first point. The second is that the replicability of psi is not important. I think it is very important to incorporate psi into mainstream science. We have a good example in nuclear physics. At the end of the last century Charles Richet suggested the use of statistical methods to investigate psi. Nuclear physicists took this idea because it is impossible to replicate "behavior" of elementary particles and they succeeded. Dr. Rhine used statistical methods to prove that ESP and PK exist. So my second question is, why do you believe that replicability is important to incorporate psi into mainstream science?

PALMER: I am not really acquainted with the situation in physics. I think parapsychology tends to be judged more by the standards that are applied to psychology, at least in the United States, in which field I think replicability is perhaps more of a factor. But my main reasons are two-fold. One is that when you have effects that are not extremely robust nor very closely linked to already existing theory, replicability is important simply to assure that there are no errors in the methodology. In other words, when you have different people getting the same effect it is somewhat less likely that any one of those outcomes is due to some kind of an artifact. You can always argue that maybe if different people replicate a finding they are all making the same mistake, but I think the problem of artifacts does get minimized to some degree with replication. My second reason is sociological. Given the particular situation we face in parapsychology, I think that other scientists are going to expect that at least some of them should be able to get these results themselves. I am not necessarily saying that is this fair or justified; but I think it is a reality that we simply have to live with.

With respect to your first point, controlling the experimenter effect is like controlling psi in general. Can we turn psi on or off? I think the evidence indicates that we can. Even if you assume that the subject is the source of the effect, we seem to have certain manipulations that facilitate psi missing as opposed to psi hitting. There are certain personality correlates that seem to suggest that some types of people score above chance, while other types of people score below chance. I think these same factors apply to the experimenter, simply as another psi source.