REPLICOMANIA AND THE PURSUIT OF MEANING IN EXPERIMENTAL PARAPSYCHOLOGY

Carl L. Sargent

Views on Replication

It appears that a majority of experimental parapsychologists (e.g., Honorton, 1976, 1977; Palmer, 1977; Sargent, 1981a) and skeptics (e.g., Crumbaugh, 1966; Nicol, 1966; Hansel, 1980; Hyman, 1982) are agreed that experimental parapsychology needs to be able to demonstrate one thing in particular if it is to escape the dreaded label of "pseudo-science," and that is: repeatability. "Parapsychology lacks a repeatable experiment" is a frequently encountered skeptical war-cry, with the word "truly" often used gratuitously to preface "repeatable." Now we can all (probably) agree that a repeatable experiment is a Good Thing. The problems arise when we try to define exactly what a repeatable experiment is and how we recognize it. When is an experiment repeatable, or (to be precise) when is it the case that an experiment gives results which may be said to be reliable or repeatable?

In an earlier paper (Sargent, 1981a) I noted some distinctions within the term "repeatability," which are certainly not original, but which bear reiteration. First, there is intra- versus inter-experimenter (or laboratory) repeatability, intra-experimenter repeatability being the repetition of an experiment by a single experimenter and inter-experimenter repeatability being the repetition of an experiment by a different experimenter to the one who conducted the original experiment. While it is frequently the case that the intra-experimenter repeatability of experimental findings is asserted to be higher than the inter-experimenter repeatability, I noted instances in which this was not obviously true (e.g., response bias effects; Sargent, 1981a, 1982).

The second distinction I made simply reiterated that made by Stanford (1974c) between concrete and conceptual replications. A concrete replication is an attempt faithfully to reproduce the conditions of a first experiment in order to observe the same results (hopefully). A conceptual replication is an attempt to confirm (or refute) the findings of one experiment by conducting another, with a differing design, the design and results of which bear some clear conceptual relationship to those of the first experiment. For example, if one experiment shows that elevation of some variable X (blood glucose, L-DOPA levels in the CNS, anything really) improves psi performance, then reducing the level of X should impair psi performance if X truly influences psi. Yet clearly an experiment which set out to examine the effects of a reduction of X on psi performance would not be a concrete replication of an experiment which investigated the effects of an elevation of X on psi performance; it would be a conceptual replication (see Braud and Braud, 1973, 1974).

I argued in the earlier paper (Sargent, 1981a), following Stanford (1974c), that conceptual replications are superior to concrete replications. Why? First, because concrete replications are impossible. One simply cannot reproduce all the conditions of an original experiment, for various reasons. First, some potentially important factors may not be isolated; if one were, for example, to survey the chronobiological literature one might develop a healthy fear of just how complex determinants of psi performance might be. Second, some factors might not be controllable for practical reasons (e.g., climatic variables) or even in principle (e.g., subtle experimenter personal attributes effects).

Also, concrete replications cannot in principle reveal certain types of psychological artifacts such as non-specific and iatrogenic effects. A classic example of this may be found in the huge pile of studies investigating the role of relaxation in systematic desensitization therapy for phobic disorders (Jacobs and Wolpin, 1971). Many experiments showed that relaxation treatment groups were superior to no-relaxation controls in this treatment; only recently have studies shown that tension groups fare just as well as relaxation groups, thus showing that the original experimental/control difference was just a non-specific effect and had nothing to do with relaxation per se. A continual blind adherence to relaxation vs. control concrete replication studies would never have revealed this. Conceptual replications advance our understanding, concrete replications merely retard it.

Finally, we are now faced in parapsychology with a number of theories and working models of how psi events occur and under which conditions that make predictions and can be tested (e.g., Honorton, 1974, 1977; Mattuck and Walker, 1977; Schmidt, 1979; Stanford, 1974a, 1974b, 1978; Walker, 1975). Now, no theory or model makes just one or two predictions; to be useful, they must make many and only conceptual replications can make progress with studying them, especially in a field like parapsychology where resources are limited.

Now, let us consider some of the very real difficulties involved in trying to assess replication rates, even when only concrete replications are being considered.

What's a Replication Anyway?

Recently, "box-score" analyses have been made of all studies undertaken to investigate effect X in psi research, where X may be RNG testing (Honorton, 1978b), Ganzfeld psi testing (Honorton, 1978a) the relationship between neuroticism and psi test performance (Palmer, 1977), extraversion and psi test performance (Sargent, 1981b), and so on. Here, an attempt is made to include all relevant studies. But just which studies are relevant and which are not? This is no simple matter, even if it is considered desirable to do this.

For example, if considering Ganzfeld/psi research, would one include the studies of Braud (1977) and Schachter and Kelly (1976)? The Braud study used pre-sleep hypnagogic state testing and the Schachter-Kelly study relaxation plus a self-induced hypnagogic-like state. Honorton (e.g., 1978a) does not include these in his box-score analysis. I think a very good case could be made for including the Schachter-Kelly study and a weaker one for the Braud study. Certainly there appears to be no objective answer to this.

On the other hand, should one include studies such as those of Rogo (1976, two studies) in which only partial isolation was used? If so, I cannot see why the Braud study should be excluded; although the subjects were not actually in the laboratory, it seems as good an approximation to the Honorton-Harper study as the Rogo study does. I think it is becoming clear that there is indeed room for negotiating whether particular studies are adequate replications or not. I do not think that we can call any experiment in which subjects have ping-pong balls put over their eyes a Ganzfeld experiment and pool it willy-nilly with others. If the points above do not suffice, let me illustrate my thesis with a reductio ad absurdum.

An experimenter runs a Ganzfeld GESP experiment and each subject is given 30 seconds to make his responses; that's the duration of Ganzfeld exposure. Is this really a viable experiment? Would one say that this should be included in a box-score analysis? Surely not. Yet there are studies in existence which have used as little as seven minutes of Ganzfeld stimulation, and several which have used 15 or less (e.g., Rogo, 1976, 2 studies; Sargent, Harley, Lane and Radcliffe, 1981; Stanford, 1979; Warnock, Dunne and Bisaha, 1977). Where is the dividing line? The answer, of course, is that there is no dividing line. One simply cannot say "17 minutes and 25 seconds," and dichotomize studies into acceptable replication attempts (with a duration of Ganzfeld isolation above this) and unacceptable ones (with a duration of Ganzfeld isolation below this). Experiments cannot be dichotomized in this way any more than they can be dichotomized into "significant" and "insignificant" outcomes.

The natural state of affairs is a continuum. Experiments are more, or less, acceptable as replications. Some of the criteria involved are reasonably "objective." Session duration in Ganzfeld (or progressive relaxation) testing is a good example. Since Honorton's (1977) internal attention states model predicts that sensory habituation is the key to psi-optimization and it is also time-progressive, then it follows from theory that session duration should be positively related to performance. There is evidence that this is so, even though the only direct comparison with subjects randomly allocated to different session durations (Sargent et al., 1981) did not show a main effect of duration. And duration is not dichotomous!

Another variable which should certainly be involved is sample size. Experimental research in psychology indicates that psychologists systematically underestimate the sample size necessary for replicating effects at a significant level (e.g., Kahneman and Tversky, 1971). As an example, the first Ganzfeld study reported (Honorton and Harper, 1974) used N = 30 and obtained P = .017 one-tailed from a 43 percent hit rate. It does not require very high intelligence to work out that, if all attempted replications used N = 30, then around one-half of them should give insignificant results due to sampling error. Rhine (e.g., Rhine and Pratt, 1957) always emphasized the importance of adequate sample size in psi testing and his warnings have rarely been reiterated. If this seems to be stressing a point unduly, the reader should consult the Kahneman-Tversky paper!

It is rare for researchers to compute the N's involved in successful and unsuccessful replication attempts and this practice should be encouraged. One problem is that the effect of N may be obscured by other, e.g., social-psychological, effects—the boredom of running a long experiment may generate experimenter personal attributes effects which depress subject performance. This type of effect will probably be a function of amount of time involved rather than number of trials, so the effect of N should probably be found more clearly with, e.g., fast RNG studies than with, say, relaxation or hypnosis research. Hyman (1982) claims to have found a nearly significant reversal with psi/Ganzfeld experiments (i.e., significance level negatively related to N), but it is clear from his paper that he has been using N for the total design of an experiment and not for the Ganzfeld trials (cf. the Wood, Kirk and Braud paper, 1977, and others).

A third example of an objective reason for being concerned about the validity of an attempted replication would be inadequate variance in independent or dependent variables in cases of correlational experiments. For example, it is in fact quite astonishing that Delanoy (1981) obtained a significant positive correlation between Ganzfeld ESP performance and extraversion considering that all her subjects had extraversion scores above the mean for the general population. Had this study failed to replicate the results reported by myself and my associates (Sargent, 1980; Sargent et al., 1981), it could certainly have been protested that Delanoy's low-variance sample (and low N to boot; N = 6, but there were 12 trials per subject) did not permit a fair test of the effect we had reported. And indeed her sample did not show such an effect; it showed a more powerful one.

I have given some examples of "objective" reasons for being concerned about whether experiments are really "replications" of others or not and I have tried to point out that there are no dichotomies involved. It is not a case of a study being an adequate replication or not. It is a question of how adequate it is and this is not always easy to decide. Where such factors as low N, low variance and so on are concerned, inter-subjective agreement may be high. However, for other issues, such agreement may not be so easy to find.

What's a Replication Anyway?—2. More About Cheshire Cats

In the following section I shall use a particular line of research to illustrate my argument, i.e., psi/Ganzfeld experimentation, since it is one which I have worked within for some time and there is sufficient data base and number of researchers involved to discuss the research in a little detail.

I want to give some examples of variables which may affect the system which is used and where social negotiation is considerable. Some of the variables have been discussed elsewhere (Sargent, 1981c).

Let us consider the target pool used in free-response testing. It is true on trivial logical grounds that this must affect psi performance as it is measured statistically. To give the obvious case, if the pool of pictures presented to the subject or independent judge for judging in free-response trials contains pictures which are insufficiently discriminable, then psi cannot be measured. The system is too insensitive. Now when we move on to factors such as emotive content of targets, again there may be effects from first principles (Sargent, 1981c), but we do not know what these effects actually are. We have a great problem here. We know the target pool must be important, but we do not know what effects target pools have. We do not even have the tools with which to measure the aspects of targets which might be important, despite the pioneer work in this field of Lendell Braud (e.g., Braud and Duke, 1980). Now, recently an American researcher and I acted as judges for Ganzfeld data the other had collected. I found her target pool simplistic, dull and boring, as did the two of my associates to whom I showed some targets. On the other hand, she found my targets over-elaborate, complex and hard to use in judging. These factors may be of crucial importance. We simply do not know. However, given the very boring nature of the targets that researcher uses, I will not be surprised by further null data from her! This belief on my part may be right and may be wrong. It is certainly one which can be debated.

Let us consider another concrete example. At the 1977 Parapsychology Foundation Conference, Jan Ehrenwald asked Honorton why certain experiments using the Ganzfeld procedure have failed to yield significant results. One interesting answer, of course, is sampling error, which would account for the majority of such failures. However, experimenters who do not obtain significant results with the procedure show an understandable, if irrational, resistance to explanations in terms of such error. In any case Honorton (1978a) discussed two studies by Palmer and associates in the context of failed replications. The study by Palmer, Bogart, Jones and Tart (1977) Honorton accepted as "a good study" and accepted that "I do not know how to explain that kind of failure." The study by Palmer and Aued (1975), however, Honorton was not happy about: "the agent . . . didn't even interact with the subjects before the session, or if so, there was only a very quick introduction. If we've learned anything in 90 years, it is that you cannot do psi experiments that way" (Honorton, 1978a, p. 93).

Skeptics would doubtless regard Honorton's argument here as post-hoc, special pleading, etc., despite the fact that there are data which suggest importance of sender-receiver familiarity in Ganzfeld GESP testing (e.g., Sargent, 1980; but see also Sondow, 1979) and, of course, a large literature on interpersonal effects generally (White, 1977), though it must be said that that literature is pretty chaotic. On the other hand, my consideration is that the kind of argument Honorton is offering here is a kind of appeal to background knowledge acquired and shared by the experienced researcher and that this kind of argument is routinely deployed in social science. I have overheard animal psychologists discussing whether such-and-such a batch of rats may have given aberrant data because they were handled by a rather insensitive lab assistant, and in the field of abnormal psychology (which I select only because I happen to be reasonably knowledgeable about it) post-hoc arguments of the variety used by Honorton are very commonplace. This is because experimenters who have worked with particular techniques for long periods or who have worked in particular areas of research for a long time, build up much implicit knowledge about their procedures which is not and in some cases cannot readily be formalized. To be sure, some superstitious conditioning may be involved here, but some of the implicit knowledge is probably accurate. The important thing for the purpose of the argument being advanced here is that, in the terms of Collins and Pinch (1979), "nothing unscientific is happening"; these appeals to implicit experiential background knowledge are part and parcel of ordinary social science. It is, apparently, only in parapsychology that certain social scientists appear to get irate about them.

So, a distinction can be made between inter-subjectively agreeable problems in determining "replication status" (sample size, etc.,) and implicit background-knowledge arguments with a very strong subjective component (e.g., importance of sender-receiver relationship). Needless to say I have not pursued my argument this far to end up with a dichotomy of this kind. There are middle cases; cases in which there are numerous empirical data which support an argument from implicit background knowledge and increase intersubjective agreement; cases in which an extrapolation from theory (cf. the importance of session duration in altered states experiments noted above) rests on too many extra assumptions to be a straightforward affair (cf. Eysenck, 1981), and so on. In effect, there is a continuum between intersubjectively highly agreeable and much more debatable parameters and values.

There are also parameters which are very difficult to find out about. Details are not published. I am reminded of a subject I once tested who had also been tested at a psi-inhibitory laboratory. His ESP performance was superior in Ganzfeld testing at Cambridge, and before he knew the result of his session he noted that he had been more comfortable with us than he had been at the other laboratory, because he had been reclining on a mattress with us and a reclining chair at the other laboratory, and also he had been distracted by corridor noise at the other lab whereas he had been undisturbed with us.

Corridor noise in a Ganzfeld experiment! Although no published report has appeared from the other laboratory, I am certain that no mention of this problem will be contained in any report that does appear. How can one have a Ganzfeld experiment in any realistic sense with such disturbances? The answer is that no room is totally sound-attenuated and everything is a matter of degree; experimental facilities are more or less adequate, not either adequate or inadequate. Just to make the picture even more complex, in this subject's test session at Cambridge there was a loud noise outside the testing area which was audible to the experimenter, very clearly; the subject did not report hearing it, probably because he appeared very relaxed and "dreamy" and in an altered state!

Factors like this may be impossible to find out about. I have chosen the example I have because of the obvious importance of relaxation and sensory habituation, so that the importance of the mattress and noises from the corridor will readily be agreed to by almost everyone. However, many other parameters of less agreed importance will also surely differ systematically across laboratories. These parameters may involve very subtle effects, including how scruffy and homely a laboratory looks, how high-status an experimenter appears, age of subject X age of experimenter interactions and all manner of things. We not only have no idea about most of them, we wouldn't even know how to measure them if we did (cf. Rosenthal, 1976).

I suggest that the status of replication experiments resembles the Cheshire Cat; just as the cat disappears and only the smile is left, on close inspection the construct "replication" appears more and more complex and difficult until finally it begins to disappear in a haze of obscurity and only the appearance remains. But this is by no means the end of the story.

What's a Replication Anyway?—3. Peer Review and Other Random Processes

Other factors influence our perception of the value and importance of replications. Consider the five following scenarios:

1. Professor X and three of his Ph.D. students replicate a particular psi experiment. Four separate researchers, working independently around the globe, all fail to replicate it.
2. Professor X and three of his Ph.D. students replicate a particular psi experiment. Around the globe, four separate and independent student researchers fail to replicate it.
3. Professor X and three of his Ph.D. students replicate a particular psi experiment. Professor Y and three of his Ph.D. students, however, fail to do so.
4. Four separate, independent researchers replicate a psi finding whilst Professor Y and three of his Ph.D. students fail to replicate.
5. Exactly as scenario 4 above, with the additional piquant factor that the four separate researchers are all rabid skeptics—and they actually publish their positive results.

Now I think one would find much agreement that these five scenarios in order provide progressively stronger evidence for the experimental hypothesis being tested. The factors involved here include separation, status and prior disposition. Replications by different researchers are of greater subjective "value" if the researchers are totally independent; if they worked together for many years, one is not quite so impressed. Again, research by real live honest-to-God professors cuts more ice than research by students, and if the professor should happen to be at Oxford, Cambridge, Harvard or Yale, this is much better than being at Newcastle, Stirling, Wyoming or Wisconsin (and heaven help you if you are at the Tri-Valley Center for Human Potential). Not that there is anything wrong with the latter universities, of course, just that a (very strong) subjective bias is being shown up here (for some of the more hilarious consequences of this bias, the reader is urged to consult Peters and Ceci, 1982, and their exposé of the peer review system of refereeing re this bias).

Again, if one knows that a particular researcher was not inclined to have much faith in a particular experimental hypothesis and then does confirm its validity, one will be more impressed with the replication. And still more subtle factors may be involved. If a researcher has experience with the general line of research within which the experimental hypothesis is embedded, his significant results may not be as desirable as those of an equally experienced researcher who usually does quite different things for a living. If one considers Ganzfeld/psi research, a significant first study from a researcher who had previously reported many successes with hypnosis, meditation and progressive relaxation would perhaps not carry as much weight as significant results from a researcher who had always spent his time exploring the finer points of RNG testing. Then again, researchers form opinions about the competence and even the integrity of their fellows and the subjective value of replications will be influenced by these subjective factors.

In short, the "repeatable experiment" seems, at last, to be a mirage. It is simply not possible to state definitively just what actually constitutes a replication attempt; it is not possible to objectify the factors which make some replications more equal than others; and, and here is the final blow, there is no such thing as a repeatable experiment in logic anyway, once a simple admission is made, and that is to deny the 100 percent repeatability criterion demand. Once this is done, and it cannot be avoided, even the cat's smile begins to disappear.

Just how repeatable would an experiment have to be in order to be a "repeatable experiment," pretending for the time being that we actually can decide what constitutes a replication? One criterion which I have never seen explicitly advocated, but which is frequently implicitly used, is the 100 percent replicability criterion; the "repeatability on demand" criterion. One hundred percent repeatability means that any experimenter (or, in the magic phrase, any competent experimenter, although the term "competent" seems almost gratuitous) can replicate an experiment and obtain psi effect X. This criterion is never used in social science. There are no effects that replicate like this in social science. If this criterion were used seriously, all of psychology, sociology, criminology, most of medicine, biochemistry, genetics, pharmacology, and other areas of science would become pseudo-science overnight. It is manifest nonsense to advocate the "any competent experimenter" criterion, and I shall disregard it in further argument.

Thus, a "repeatable experiment" is one that gives a significant result somewhere above 5 percent of the time. But where? The answer is, of course, that there is no such thing as a "repeatable experiment." There are experiments which are more or less significant and emphatically not just two kinds—significant and insignificant (cf. Sargent, 1981a). And just where one puts a subjective cut-off and says "This is good," "This is promising" or "This is inadequate" is totally subjective. This cannot be a rational decision; it is purely subjective. One may be able to compare the replication rates of different lines of experiment (or estimate changes in replication rates over time for a single experiment) although I am dubious even about this; what one cannot do is to state that a particular experiment is, or is not, repeatable with any real confidence.

We might also consider here the possibility of an empirical standard for replication in psi research. Dropping any pretense at a rational criterion, we might become thoroughgoing conventionalists and use some area of research in experimental psychology as an empirical yardstick for assessing replication in parapsychology. Honorton (1976) has suggested this and such possible parallels as subliminal perception (Dixon, 1979; Roney-Dougal, 1981) and olfactory psychology (Sargent, 1981a) have been suggested. There are, however, problems involved with this strategy. The first is that psychologists do not often attend to problems of replication (cf. Barber, 1976, who points out that only 10 percent of introductory texts on psychology even contain the term "repeatability" in the index). The second is that the nonpublication of null data is far more prevalent in psychology, where around 90 percent of experiments involve a rejection of the null hypothesis at the 1 percent level of significance (cf. Barber, 1976, and his references), than in parapsychology, so the dice are rather loaded against psi from the start if such comparisons are made. The third problem, of course, is deciding just which area of psychology is the most appropriate. The more one ponders such an empirical comparison, the more complex things become.

Parapsychology and Mainstream Science: Another 100 Years on the Margins?

The message of this paper seems uncompromisingly negative and even hostile to scientific endeavor. There are two good reasons for this. The first is that the belief that "science" is an undefinable entity and that there are massive chunks of conventionalism in the edifice of science, is rapidly gaining ground among philosophers and especially sociologists of science. The second is a growing realization that Feyerabend (1978) was right to proclaim, "What's so great about science?" These are two very important and all-embracing issues, and obviously in this paper I cannot deal with them; but a few comments can be made.

The problem of the "demarcation criterion" in science has always been a thorny one. Popper (1959) certainly eliminated inductivism as a criterion—to his own satisfaction—and attempted to substitute falsificationism in its place (Popper, 1963, 1972), but even in its methodologically sophisticated versions, falsificationism just does not appeal. Popper's qualifications about premature falsifications, the need for scientists to remain "tenacious" in defence of theories in the face of apparent falsifications, his inability to define exactly what a falsification is and, above all, examples from the history of science of apparent comprehensive falsification of hypotheses now accepted (cf. Feyerabend, 1975) make the falsification criterion unappealing. The real reason why it doesn't cut much ice, though, is illustrated in Popper's reply to the question of why scientists don't behave like the fearless falsificationists of his model: "Well, they should." Popper does not dispute the question—he accepts that scientists just don't behave as his model suggests.

Eysenck (1979, 1981) has suggested that falsificationism (if we pretend that it is a coherent philosophy of science for a while) typifies the highest stage of scientific development. Before that, in earlier stages, justificationism and inductivism are quite sound—and indeed necessary—strategies. To borrow the language of Lakatos (1978), a "research program," when young, needs protecting. A Type II error early on is more dangerous than a Type I error. There is simply no need to apply the same criteria to the whole of science; research programs of different levels of maturity should be evaluated by different criteria. This seems to be common sense; one would not test the intellect of a five-year-old child by setting him university examination papers. He might be a very promising five-year old indeed, but faced with an unreasonable test he won't appear so.

And, indeed, what is so great about science anyway? The excellence of science cannot, as Feyerabend argues in a masterful essay (1978), be established using scientific methods. One cannot justify the methodological supremacy of science by using the criteria of manipulation, control and prediction, since this is in effect using the rules of science to decide the issue. Many things of crucial importance to our lives are things about which science has nothing to say. If one wants to control and manipulate the world, science is a wonderful tool; if one wants to understand it and feel at one with it, then a religious or phenomenological epistemology would be appropriate. Only a fool would turn to science to find out what anything actually *means.* Understanding and acknowledging a plurality of epistemologies—and giving them a fair share of resources as Feyerabend (1978) urges—is essential to an understanding of psi phenomena. If this line of argument appears vague and undisciplined—and, with no training in phenomenology and not being of a religious frame of mind, it probably does—then let me simply point out that, in the rules of social science, there is absolutely nothing which forbids you to bore your subjects to death in experiments. That this might not be psi-conducive is an insight which comes to us not from science, but from common sense. And that inchoate, only partly formalizable body of knowledge is at least as useful to any parapsychologist as the corpus of scientific knowledge. Science is only a tool for exploring some aspects of our universe. It is not the only tool in the tool-chest and not the most appropriate one for many tasks. It is, however, currently the only tool which shows signs of telling the carpenter to throw the rest of the toolbox away!

BIBLIOGRAPHY

Barber, T. X. *Pitfalls in Human Research.* London:Pergamon Press, 1976.

Braud, L. W., and Duke, D. M. "Qualities of free-response targets and their relationship to psi performance." In W. G. Roll (Ed.) *Research in Parapsychology 1979.* Metuchen, NJ: Scarecrow Press, 1980, pp. 74–77.

Braud, L. W., and Braud, W. G. "Further studies of relaxation as a psi-conducive state." *Journal of the American Society for Psychical Research,* 1974, *68,* 3, 229–245.

Braud, W. G. "Long-distance dream and presleep telepathy." In J. D. Morris, W. G. Roll and R. L. Morris (Eds.) *Research in Parapsychology 1976.* Metuchen, NJ: Scarecrow Press, 1977, pp. 154–155.

Braud, W. G., and Braud, L. W. "Preliminary explorations of psi-conducive states: Progressive muscular relaxation." *Journal of American Society for Psychical Research,* 1973, *67,* 1, 26–46.

Collins, H. M., and Pinch, T. J. "The construction of the paranormal: Nothing unscientific is happening." In R. Wallis (Ed.) *On the Margins of Science: The Social Construction of Rejected Knowledge.* Staffordshire: University of Keele, 1979.

Crumbaugh, J. C. "A scientific critique of parapsychology." *International Journal of Neuropsychiatry,* 1966, *2,* 5, 523–531.

Delanoy, D. Paper presented at International Parascience Conference, London, September, 1981.

Dixon, N. F. "Subliminal perception and parapsychology: Points of contact." *Parapsychology Review,* 1979, *10,* 3, 1–6.

Dunne, B., Warnock, M., and Bisaha, J. P. "Ganzfeld techniques with independent rating for measuring GESP and precognition." In J. D. Morris, W. G. Roll and R. L. Morris (Eds.) *Research in Parapsychology 1976.* Metuchen, NJ: Scarecrow Press, 1977.

Eysenck, H. J. *The Structure and Measurement of Intelligence.* New York: Springer, 1979.

Eysenck, H. J. In H. J. Eysenck (Ed.) *A Model for Personality.* New York: Springer, 1981.

Feyerabend, P. K. *Against Method.* London: New Left Books, 1975.

Feyerabend, P. K. *Science in a Free Society.* London: New Left Books, 1978.

Hansel, C. E. M. *ESP and Parapsychology: A Critical Re-evaluation.* Buffalo: Prometheus Books, 1980.

Honorton, C. "ESP and Altered States of Consciousness." In J. Beloff (Ed.) *New Directions in Parapsychology.* London: Elek Science, 1974.

Honorton, C. "Has science developed the competence to confront claims of the paranormal?" In J. D. Morris, W. G. Roll and R. L. Morris (Eds.) *Research in Parapsychology 1975.* Metuchen, NJ: Scarecrow Press, 1976, pp. 199–223.

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

Honorton, C. "Psi and internal attention states: Information retrieval in the Ganzfeld." In B. Shapin & L. Coly (Eds.) *Psi and States of Awareness.* Proceedings of an international conference, held in Paris, France, August 24–26, 1977. New York: Parapsychology Foundation, Inc., 1978, pp. 79–100.

Honorton, C. "Replicability, experimenter influence and parapsychology: An empirical context for the study of mind." Paper presented at an American Association for the Advancement of Science Symposium, held in Washington, D.C., February 12–17, 1978.

Honorton, C., and Harper, S. "Psi-mediated imagery and ideation in an experimental procedure for regulating perceptual input." *Journal of the American Society for Psychical Research,* 1974, *68,* 2, 156–168.

Hyman, R. "Does the Ganzfeld experiment answer the critic's objections?" Paper presented at the Society for Psychical Research/Parapsychological Association Centenary-Jubilee Conference, held at Trinity College, Cambridge, August 16–21, 1982.

Jacobs, A., and Wolpin, M. In A. Jacobs and L. B. Sachs (Eds.) *The Psychology of Private Events.* London: Academic Press, 1971.

Kahneman, D., and Tversky, A. *Psychological Bulletin,* 1971, 2, 105–110.

Lakatos, I. *The Epistemology of Scientific Research Programs.* Cambridge: University Press, 1978.

Mattuck, R., and Walker, E. H. "The action of consciousness on matter: A quantum mechanical theory of psychokinesis." In A. Puharich (Ed.) *The Iceland Papers.* Amherst, WI: Essentia Research Associates, 1979.

Nicol, J. F. "Some difficulties in the way of scientific recognition of extrasensory perception." In G. E. W. Wolstenholm and E. C. P. Miller (Eds.) *Extrasensory Perception.* Boston: Little, Brown and Company, 1956, pp. 24–38.

Palmer, J. "Attitudes and personality traits in experimental ESP research." In B. B. Wolman (Ed.) *Handbook of Parapsychology.* New York: Van Nostrand Reinhold, 1977, pp. 175–201.

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,* 2, 121–145.

Peters, D. P. and Ceci, S. J. "Peer review practices of psychological journals: The fate of published articles submitted again." *Behavioral and Brain Sciences,* 1982, *5,* 187–195.

Popper, K. R. *The Logic of Scientific Discovery.* London: Hutchinson, 1959.

Popper, K. R. *Conjectures and Refutations.* London: Routledge and Kegan Paul, 1963.

Popper, K. R. *Objective Knowledge.* Oxford: University Press, 1972.
Rhine, J. B., and Pratt, J. G. *Parapsychology: Frontier Science of the Mind.* Springfield, IL: Charles C Thomas, 1957.
Roney-Dougal, S. M. "Interface between psi and subliminal perception." *Parapsychology Review,* 1981, *12,* 4, 12–18.
Rosenthal, R. *Experimenter Effects in Behavior Research.* New York: Irvington, 1976.
Sargent, C. L. *Exploring Psi In the Ganzfeld.* New York: Parapsychology Foundation, Inc. 1980.
Sargent, C. L. "The repeatability of significance and the significance of repeatability." *European Journal of Parapsychology,* 1981a, *3,* 4, 423–443.
Sargent, C. L. "Extraversion and performance in 'extrasensory perception' tasks." *Personality and Individual Differences,* 1981b, *2,* 137–143.
Sargent, C. L., "ESP in the twilight zone: State of the art." *Parapsychology Review,* 1981c, *12,* 1, 1–7.
Sargent, C. L. "An unusually powerful response-bias effect." *Parapsychology Review,* 1982, *13,* 3, 8–10.
Sargent, C. L., Harley, T. A., Lane, J., and Radcliffe, K. "Ganzfeld psi-optimization in relation to session duration." In W. G. Roll and J. Beloff (Eds.) *Research in Parapsychology 1980.* Metuchen, NJ: Scarecrow Press, 1981, pp. 82–84.
Schmidt, H. "Evidence for direct interaction between the human mind and external quantum processes." In C. Tart, H. Puthoff and R. Targ (Eds) *Mind at Large.* New York: Praeger, 1979, pp. 207–220.
Stanford, R. G. "An experimentally testable model for spontaneous psi events. I. Extrasensory results." *Journal of the American Society for Psychical Research,* 1974a, *68,* 1, 34–57.
Stanford, R. G. "An experimentally testable model for spontaneous psi events. II. Psychokinetic events." *Journal of the American Society for Psychical Research,* 1974b, *68,* 4, 321–356.
Stanford, R. G. "Concept and psi." In W. G. Roll, R. L. Morris and J. D. Morris (Eds.) *Research in Parapsychology 1973.* Metuchen, NJ: Scarecrow Press, 1974c, pp. 137–162.
Stanford, R. G. "Toward reinterpreting psi events." *Journal of the American Society for Psychical Research,* 1978, *72,* 3, 197–214.
Stanford, R. G. "The influence of auditory Ganzfeld characteristics upon free-response ESP performance." *Journal of the American Society for Psychical Research,* 1979, *73,* 3, 253–272.
Walker, E. H., "Foundations of paraphysical and parapsychological phenomena." In L. Oteri (Ed.) *Quantum Physics and Parapsychology.* Proceedings of an international conference, held in Geneva, Switzerland, August 26–27, 1974. New York: Parapsychology Foundation, Inc., 1975, pp. 1–53.
White, R. A. "The influence of experimenter motivation, attitudes, and methods of handling subjects on psi test results." In B. B. Wolman (Ed.) *Handbook of Parapsychology.* New York: Van Nostrand Reinhold, 1977, pp. 273–301.
Wood, R., Kirk, J., and Braud, W. G. "Free response GESP performance following Ganzfeld stimulation vs. induced relaxation, with verbalized vs. non-verbalized mentation: A failure to replicate." *European Journal of Parapsychology,* 1977, *1,* 4, 80–93.

DISCUSSION

BELOFF: I take your point about the somewhat arbitrary nature of what replication means and that it is impossible to give any hard and fast ruling on this. With reference to my student Deborah Delanoy, the most salient fact about her work is that in spite of her very high motivation to get results with her Ganzfeld work, she's not getting results. Once she was persuaded to run a correlation between extraversion and scoring and she got something that was significant, much to her surprise and incredulity. But the fact is that after having chosen extravert subjects, sheep subjects, all the ones that traditionally were supposed to yield good Ganzfeld results, and after having been coached by you in correct Ganzfeld procedures and so on, with everything going for her she still gets mostly negative results, not significant results, they're all over the place. And this is, of course, a keen disappointment. But generally I would say that the kind of situation that I've found in my experience and the experience of my students to be the most distressing is where one and the same experimenter can't get replication of his own work. I had a very salient case with Richard Broughton, who started getting very promising results on hemisphere difference findings and then found he couldn't repeat his own experiments. Now I mean those are the kinds of experiments that one can most hope to get results with, because, as you said, once you transfer to a different laboratory, a different method you are not really replicating. But once you find you can no longer repeat your own work, then you really are stymied. And I wondered if you have any comments or suggestions about this.

SARGENT: As far as Deborah's work is concerned there is, in fact, one aspect of the procedure that I have great concern about. That was her target pool. When we exchanged target pools I found hers incredibly dull, boring and simplistic and other people in my laboratory considered it a very dull, boring target pool. She found my targets very over-complex, very difficult and so on. There were radical dissimilarities, different views about what kind of target system we should be working with. It's the kind of factor you just don't discover, until one day I agreed to judge her data and a parcel appeared through the post and there were the targets. This is the kind of thing you're never going to find out unless you're actually doing something like that. I've only seen targets used by two other people. I saw some of Chuck Honorton's targets and some of Deborah's and Brian Millar has seen some of ours. So I have no idea about what the target pool is like in most other people's laboratories. Here is a very good example of something which we know must be important. We really don't have details on it. You cannot reproduce in a journal some 200 plates with all our targets. So that's one of the enlarged range of variables that presents problems. On replicating one's own results, I think people have a tendency when they fail once, to stop. At the PA Convention this year I shall be reporting four experiments where in the fourth one, we actually found a real sheep-goat effect. In fact, it wasn't there in the first three. But I think there is a tendency to obtain a finding once or twice, then you try it once or twice more and if it doesn't work you forget it and go home. My guess is that you should always plan for a long series of experiments. You may get three significant findings out of eight; if they happen to be experiments 1, 2 and 3 well that's one thing. But they could turn out to be experiments 5, 6 and 7. Most people have given up by then. I think there is a tendency for people to infer too much from small amounts of data. I think that one trend that was very strongly present in John Palmer's PA Presidential Address a couple of years ago, was the idea that the individual experiments really don't teach you anything. You can only really draw conclusions from pretty wide arrays of data. I think that that's right. I don't think that individuals should be too depressed by one or two studies that don't turn out right.

RONEY-DOUGAL: In your paper you didn't actually talk about where this is going to take us into the next century.

SARGENT: One of the problems we've got here is that we have this criterion that we keep getting thrown at us: you don't have a repeatable experiment. When you start looking at the repeatable experiment you find it has distinctively mirage-like qualities. I'm beginning to wonder whether we should be dismantling this concept a bit. First of all, I'm disputing the criteria of science by saying they're not what they really appear to be. Secondly, I'm not certain that the criteria, even if you could pretend that they existed, would really be worth aiming for. One of the outstanding arguments for trying to persuade skeptics that parapsychology's findings are valid is that parapsychologists need research funds controlled by skeptical individuals. And this may be the wrong way of doing it. Maybe this is entirely erroneous and what is necessary is political action. The people that actually pay the money for the research and the people down there on the street should be simply told how much is spent by the Oxbridge civil servants and their "old boy" links in the research councils every year on a) parapsychology and b) rat food. Then those people could make a decision on whether they like their money being used like that or not. I think we might have the problem the wrong way around. Rather than persuade the scientific community that funds are deserved for working in this kind of area, one needs a better control of the funding of science down to a more democratic and accountable level. I think that's an implication that we should start taking to heart if we begin to examine the criteria we're supposed to be meeting and find they are very dubious when we start looking at them closely.

WEST: I agree entirely with Carl that conceptual replication is more important than literal replication. What is important about replication is that different experimenters at different times should reach the same conclusions. It seems to me, looking at the whole field of psi research, that conclusions on which everyone agrees are remarkably scarce. On the most basic aspects of psi we still don't have agreed answers to questions like which is easier—precognition, telepathy or clairvoyance? Such fundamental questions remain unanswerable because research has not given consistent findings over the years.

Now Carl has been talking about the conclusions of the Ganzfeld experiments which seem to suggest that relaxation is a psi-conducive state. I wonder if everyone really agrees and if the findings are entirely consistent with that conclusion. I would like to ask this question, not strictly concerned with replicability, but related to it. In these experiments, for instance in his own Ganzfeld tests, which have produced successful psi scoring on a number of occasions, Carl has pointed out that there are many intervening variables involved. The nature of the targets, the nature of the experimenter and the circumstances of the test, may affect the results and may affect replicability. Have controlled experiments been done in which the non-relaxed condition has been compared with the relaxed condition with exactly the same set up, same personnel, same sample, same environment, etc?

SARGENT: Yes, that indeed is the case. I suppose the best example would be the experiment by William Braud, Robert Wood and Lendell Braud who compared the Ganzfeld group with a no-Ganzfeld group and also the Braud and Braud experiments which used both relaxation and tension groups. We have the three strata there. Relaxation, control and tension groups. Chuck Honorton has also reported a Ganzfeld study with a control group. Control groups perhaps have not been as popular as they might be, there are certain motivational problems with them. Practically, in my own work, it has been impossible to use them. Since everybody knows about the ping pong balls, if you actually told them to come in and just sit down and make themselves comfortable and talk about what they thought somebody in the next building was looking at, then they'd say, "Gee, we're not going to get the ping pong balls!" The motivation would be shot to bits and we now have an artifact possibility that they might do poorly simply because they had bad motivation. If you wanted to look for literature which has quite a wide range of control groups it would be the hypnosis literature where control groups have been much more frequently used. I think there are something like 23 comparisons of hypnosis and control groups in the literature. I think that on 15 occasions the hypnosis group has shown a significant deviation from the chance level whereas the control groups have never done so. Another point that perhaps I didn't develop as well as I might have is that there certainly are findings that have regularities to them. There are certainly many findings where if you look at the statistically significant results they're clearly all of one kind. For example, if you look at the sheep-goat effect, the results always show sheep better than goats when they turn out significantly. You don't get it the other way around. And the same is true with the extravert and introvert difference. There are known to me 21 significant findings in which, in 20 cases, the extraverts were superior to the introverts; only once has the reverse been found. And this was the only case where an automated test was used; perhaps the introverts felt more at home with it. It is also true in the case of positive versus negative deviations of progressive relaxation, dream states and Ganzfeld testing. There is the relationship between neuroticism and ESP. There are many occasions in which the data show clear regularities; all the significant results go one way. That's a slightly different matter from the repeatability rate, though. It's a related matter, but it is rather different, because you now are saying let's forget the null data, let's have a look at the stuff that shows you that a real effect is there. I think this is actually a criterion where you can say a real effect is definitely present, without having to invoke repeatability, strangely enough, because you are not actually making that many comments about repeatability.

ROUSSEL: If I ever get funds to create a lab I will be looking for that kind of information to orient students towards noting all the details. But if we need that much information about the physical nature of things, shape, size, light and personality variables and also about the experimenter, it gets very complicated. A few years ago, in an article in the *Parapsychology Review*, entitled "Holistic Methodology", you advocated simplification. So my question is where did you get with this idea of simplification, because yesterday you mentioned many judges and how you now use the idea of simplification?

SARGENT: Oh my goodness, it's always pleasant to find people with a long memory who dig up what you said years ago and hit you over the head with it. We started actually researching using the simplification postulate. There is a paper in the 1982 *European Journal of Parapsychology*, by Trevor Harley and myself where we both did parallel experiments in which we played all the roles—experimenter, subject, randomizer. You get rid of everybody if you have problems with the experimenter effect. The easiest thing to do is to get rid of the subjects. You can't get rid of the experimenter, but get rid of the subjects and there is just you. For political reasons it's probably sensible to have two of you doing this experiment at the same time. We really did do this and we got results so similar that if you'd known the performance rate of one subject you could have predicted the results of the other with perfect accuracy. Neither of our results were individually significant, but when you put them together you got statistical significance, which is rather amusing, but it was exactly what we wanted to get. But the only problem is that when you really think about it you know those are very good results from Sargent and Harley. Now there is only one Sargent and there is only one Harley. So we can go on doing that, but the problem is that it's just those two individuals. The problem with simplification, when one starts to think about it deeply, is that you really only start getting increasingly idiosyncratic effects and I'm hopeful that you could do that with other people and they might get stable results. You might start getting the kind of private science that some of the observational theories drive us towards, if you do this kind of thing. That is, I think, a valid direction for research to go and it would certainly be a new methodology. What I was trying to present in this paper was how we tend to discuss experimental data of the kind that we have been collecting for the last quarter of a century and how it's been discussed in that time and how, if we want to continue doing this kind of thing, we should go about doing it. My argument is that we may have been chasing a mirage when we've been aiming for the repeatable experiment, because you start trying to figure out what a repeatable experiment is and it vaporizes before your eyes. It is very difficult to determine what it is. So I think this paper was a summary of what we've been doing, the way we've been tending to approach regularities in research findings in the last 25 to 30 years.