

Comment

Persi Diaconis

In my experience, parapsychologists use statistics extremely carefully. The plethora of widely significant p-values in the many thousands of published parapsychological studies must give us pause for thought. Either something spooky is going on, or it is possible for a field to exist on error and artifact for over 100 years. The present paper offers a useful review by an expert and a glimpse at some tantalizing new studies.

My reaction is that the studies are crucially flawed. Since my reasons are somewhat unusual, I will try to spell them out.

I have found it impossible to usefully judge what actually went on in a parapsychology trial from their published record. Time after time, skeptics have gone to watch trials and found subtle and not-so-subtle errors. Since the field has so far failed to produce a replicable phenomena, it seems to me that any trial that asks us to take its findings seriously should include full participation by qualified skeptics. Without a magician and/or knowledgeable psychologist skilled at running experiments with human subjects, I don't think a serious effort is being made.

I recognize that this is an unorthodox set of requirements. In fact, one cannot judge what "really goes on" in studies in most areas, and it is

Persi Diaconis is Professor of Mathematics at Harvard University, Science Center, 1 Oxford Street, Cambridge, Massachusetts 02138.

Comment: Parapsychology — On the Margins of Science?

Joel B. Greenhouse

Professor Utts reviews and synthesizes a large body of experimental literature as well as the scientific controversy involved in the attempt to estab-

Joel B. Greenhouse is Associate Professor of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213-3890.

impossible to demand wide replicability in others. Finally, defining "qualified skeptic" is difficult. In defense, most areas have many easily replicable experiments and many have their findings explained and connected by unifying theories. It simply seems clear that when making claims at such extraordinary variance with our daily experience, claims that have been made and washed away so often in the past, such extraordinary measures are mandatory before one has the right to ask outsiders to spend their time in review. The papers cited in Section 5 do not actively involve qualified skeptics, and I do not feel they have earned the right to our serious attention.

The points I have made above are not new. Many appear in the present article. This does not diminish their utility nor applicability to the most recent studies.

Parapsychology is worth serious study. First, there may be something there, and I marvel at the patience and drive of people like Jessica Utts and Ray Hyman. Second, if it is wrong, it offers a truly alarming massive case study of how statistics can mislead and be misused. Third, it offers marvelous combinatorial and inferential problems. Chung, Diaconis, Graham and Mallows (1981), Diaconis and Graham (1981) and Samaniego and Utts (1983) offer examples not cited in the text. Finally, our budding statistics students are fascinated by its claims; the present paper gives a responsible overview providing background for a spectacular classroom presentation.

lish the existence of paranormal phenomena. The organization and clarity of her presentation are noteworthy. Although I do not believe that this paper will necessarily change anyone's views regarding the existence of paranormal phenomena, it does raise very interesting questions about the process by which new ideas are either accepted or rejected by the scientific community. As students of science, we believe that scientific discovery

advances methodically and objectively through the accumulation of knowledge (or the rejection of false knowledge) derived from the implementation of the scientific method. But, as we will see, there is more to the acceptance of new scientific discoveries than the systematic accumulation and evaluation of facts. The recognition that there is a social process involved with the acceptance or rejection of scientific knowledge has been the subject of study of sociologists for some time. The scientific community's rejection of the existence of paranormal phenomena is an excellent case study of this process (Allison, 1979; Collins and Pinch, 1979).

Implicit in Professor Utts' presentation and paramount to the acceptance of parapsychology as a legitimate science are the description and documentation of the professionalization of the field of parapsychology. It is true that many researchers in the field have university appointments; there are organized professional societies for the advancement of parapsychology; there are journals with rigorous standards for published research; the field has received funding from federal agencies; and parapsychology has received recognition from other professional societies, such as the IMS and the American Association for the Advancement of Science (Collins and Pinch, 1979). Nevertheless, most readers of *Statistical Science* would agree that parapsychology is not accepted as part of orthodox science and is considered by most of the scientific community to be on the margins of science, at best (Allison, 1979; Collins and Pinch, 1979). Why is this the case? Professor Utts believes that it is because people have not examined the data. She states that "Strong beliefs tend to be resistant to change even in the face of data, and many people, scientists included, seem to have made up their minds on the question without examining any empirical data at all."

The history of science is replete with examples of resistance by the established scientific community to new discoveries. A challenging problem for science is to understand the process by which a new theory or discovery becomes accepted by the community of scientists and, likewise, to characterize the nature of the resistance to new ideas. Barber (1961) suggests that there are many different sources of resistance to scientific discovery. In 1900, for example, Karl Pearson met resistance to his use of statistics in applications to biological problems, illustrating a source of resistance due to the use of a particular methodology. The Royal Society informed Pearson that future papers submitted to the Society for publication must keep the mathematics separate from the biological applications.

Another obvious source of resistance to new sci-

entific ideas, and the one referred to by Professor Utts above, is the prevailing substantive beliefs and theories held by scientists at any given time. Barber offers the opposition to Copernicus and his heliocentric theory and to Mendel's theory of genetic inheritance as examples of how, because of preconceived ideas, theories and values, scientists are not as open-minded to new advances as one might think they should be. It was R. A. Fisher who said that each generation seems to have found in Mendel's paper only what it expected to find and ignored what did not conform to its own expectations (Fisher, 1936).

Pearson's response to the antimathematical prejudice expressed by the Royal Society was to establish with Galton's support a new journal, *Biometrika*, to encourage the use of mathematics in biology. Galton (1901) wrote an article for the first issue of the journal, explaining the need for this new voice of "mutual encouragement and support" for mathematics in biology and saying that "a new science cannot depend on a welcome from the followers of the older ones, and [therefore]... it is advisable to establish a special Journal for Biometry." Lavoisier understood the role of preconceived beliefs as a source of resistance when he wrote in 1785,

I do not expect my ideas to be adopted all at once. The human mind gets creased into a way of seeing things. Those who have envisaged nature according to a certain point of view during much of their career, rise only with difficulty to new ideas. (Barber, 1961.)

I suspect that this paper by Professor Utts synthesizing the accumulation of research results supporting the existence of paranormal phenomena will continue to be received with skepticism by the orthodox scientific community "even after examining the data." In part, this resistance is due to the popular perception of the association between parapsychology and the occult (Allison, 1979) and due to the continued suspicion and documentation of fraud in parapsychology (Diaconis, 1978). An additional and important source of resistance to the evidence presented by Professor Utts, however, is the lack of a model to explain the phenomena. Psychic phenomena are unexplainable by any current scientific theory and, furthermore, directly contradict the laws of physics. Acceptance of psi implies the rejection of a large body of accumulated evidence explaining the physical and biological world as we know it. Thus, even though the effect size for a relationship between aspirin and the prevention of heart attacks is three times smaller than the effect size observed in the ganzfeld data

base, it is the existence of a biological mechanism to explain the effectiveness of aspirin that accounts, in part, for acceptance of this relationship.

In evaluating the evidence in favor of the existence of paranormal phenomena, it is necessary to consider alternative explanations or hypotheses for the results and, as noted by Cornfield (1959), "If important alternative hypotheses are compatible with available evidence, then the question is unsettled, even if the evidence is experimental" (see also Platt, 1964). Many of the experimental results reported by Professor Utts need to be considered in the context of explanations other than the existence of paranormal phenomena. Consider the following examples:

(1) In the various psi experiments that Professor Utts discusses, the null hypothesis is a simple chance model. However, as noted by Diaconis (1978) in a critique of parapsychological research, "In complex, badly controlled experiments simple chance models cannot be seriously considered as tenable explanations: hence, rejection of such models is not of particular interest." Diaconis shows that the underlying probabilistic model in many of these experiments (even those that are well controlled) is much more complicated than chance.

(2) The role that experimenter expectancy plays in the reporting and interpreting of results cannot be underestimated. Rosenthal (1966), based on a meta-analysis of the effects of experimenters' expectancies on the results of their research, found that experimenters tended to get the results they expected to get. Clearly this is an important potential confounder in parapsychological research. Professor Utts comments on a debate between Honorton and Hyman, parapsychologist and critic, respectively, regarding evidence for psi abilities, and, although not necessarily a result of experimenter expectancy, describes how "...each analyzed the results of all known psi ganzfeld experiments to date, and reached strikingly different conclusions."

(3) What is an acceptable response in these experiments? What constitutes a direct hit? What if the response is close, who decides whether or not that constitutes a hit (see (2) above)? In an example of a response of a Receiver in an automated ganzfeld procedure, Professor Utts describes the "dream-like quality of the mentation." Someone must evaluate these stream-of-consciousness responses to determine what is a hit. An important methodological question is: How sensitive are the results to different definitions of a hit?

(4) In describing the results of different meta-analyses, Professor Utts is careful to raise ques-

tions about the role of publication bias. Publication bias or "the file-drawer problem" arises when only statistically significant findings get published, while statistically nonsignificant studies sit unreported in investigators' file drawers. Typically, Rosenthal's method (1979) is used to calculate the "fail-safe N ," that is, the number of unreported studies that would have to be sitting in file-drawers in order to negate the significant effect. Iyengar and Greenhouse (1988) describe a modification of Rosenthal's method, however, that gives a fail-safe N that is often an order of magnitude smaller than Rosenthal's method, suggesting that the sensitivity of the results of meta-analyses of psi experiments to unpublished negative studies is greater than is currently believed.

Even if parapsychology is thought to be on the margins of science by the scientific community, parapsychologists should not be held to a different standard of evidence to support their findings than orthodox scientists, but like other scientists they must be concerned with spurious effects and the effects of extraneous variables. The experimental results summarized by Professor Utts appear to be sensitive to the effect of alternative hypotheses like the ones described above. Sensitivity analyses, which question, for example, how large of an effect due to experimenter expectancy there would have to be to account for the effect sizes being reported in the psi experiments, are not addressed here. Again, the ability to account for and eliminate the role of alternative hypotheses in explaining the observed relationship between aspirin and the prevention of heart attacks is another reason for the acceptance of these results.

A major new technology discussed by Professor Utts in synthesizing the experimental parapsychology literature is meta-analysis. Until recently, the quantitative review and synthesis of a research literature, that is, meta-analysis, was considered by many to be a questionable research tool (Wachter, 1988). Resistance by statisticians to meta-analysis is interesting because, historically, many prominent statisticians found the combining of information from independent studies to be an important and useful methodology (see, e.g., Fisher, 1932; Cochran, 1954; Mosteller and Bush, 1954; Mantel and Haenszel, 1959). Perhaps the more recent skepticism about meta-analysis is because of its use as a tool to advance discoveries that themselves were the objects of resistance, such as the efficacy of psychotherapy (Smith and Glass, 1977) and now the existence of paranormal phenomena. It is an interesting problem for the history of science to explore why and when in the development of a

of a discipline it turns to meta-analysis to answer research questions or to resolve controversy (e.g., Greenhouse et al., 1990).

One argument for combining information from different studies is that a more powerful result can be obtained than from a single study. This objective is implicit in the use of meta-analysis in parapsychology and is the force behind Professor Utts' paper. The issue is that by combining many small studies consisting of small effects there is a gain in power to find an overall statistically significant effect. It is true that the meta-analyses reported by Professor Utts find extremely small p -values, but the estimate of the overall effect size is still small. As noted earlier, because of the small magnitude of the overall effect size, the possibility that other extraneous variables might account for the relationship remains.

Professor Utts, however, also illustrates the use of meta-analysis to investigate how studies differ and to characterize the influence of difficult covariates or moderating variables on the combined estimate of effect size. For example, she compares the mean effect size of studies where subjects were selected on the basis of good past performance to studies where the subjects were unselected, and she compares the mean effect size of studies with feedback to studies without feedback. To me, this latter use of meta-analysis highlights the more valuable and important contribution of the methodology. Specifically, the value of quantitative methods for

research synthesis is in assessing the potential effects of study characteristics and to quantify the sources of heterogeneity in a research domain, that is, to study systematically the effects of extraneous variables. Tom Chalmers and his group at Harvard have used meta-analysis in just this way not only to advance the understanding of the effectiveness of medical therapies but also to study the characteristics of good research in medicine, in particular, the randomized controlled clinical trial. (See Mosteller and Chalmers, 1991, for a review of this work.)

Professor Utts should be congratulated for her courage in contributing her time and statistical expertise to a field struggling on the margins of science, and for her skill in synthesizing a large body of experimental literature. I have found her paper to be quite stimulating, raising many interesting issues about how science progresses or does not progress.

ACKNOWLEDGMENT

This work was supported in part by MHCRC grant MH30915 and MH15758 from the National Institute of Mental Health, and CA54852 from the National Cancer Institute. I would like to acknowledge stimulating discussions with Professors Larry Hedges, Michael Meyer, Ingram Olkin, Teddy Seidenfeld and Larry Wasserman, and thank them for their patience and encouragement while preparing this discussion.

Comment

Ray Hyman

Utts concludes that "there is an anomaly that needs explanation." She bases this conclusion on the ganzfeld experiments and four meta-analyses of parapsychological studies. She argues that both Honorton and Rosenthal have successfully refuted my critique of the ganzfeld experiments. The meta-analyses apparently show effects that cannot be explained away by unreported experiments nor over-analysis of the data. Furthermore, effect size does not correlate with the rated quality of the experiment.

Neither time nor space is available to respond in detail to her argument. Instead, I will point to some of my concerns. I will do so by focusing on those parts of Utts' discussion that involve me. Understandably, I disagree with her assertions that both Honorton and Rosenthal successfully refuted my criticisms of the ganzfeld experiments.

Her treatment of both the ganzfeld debate and the National Research Council's report suggests that Utts has relied on second-hand reports of the data. Some of her statements are simply inaccurate. Others suggest that she has not carefully read what my critics and I have written. This remoteness from the actual experiments and details of the arguments may partially account for her optimistic assessment of the results. Her paper takes

Ray Hyman is Professor of Psychology, University of Oregon, Eugene, Oregon 97403.