

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.

---

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

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