

# A meta-analysis of mass-media tests of extrasensory perception

**Julie Milton**

*Department of Psychology, University of Edinburgh, UK*

**Richard Wiseman**

*Department of Psychology, University of Hertfordshire, UK*

Some scientists have argued that recent meta-analyses of many different types of parapsychological study suggest that extrasensory perception (ESP) might exist, albeit as a small effect. Large-scale ESP experiments conducted via newspapers, magazines, radio and television can generate a huge number of guesses and offer researchers a way of quickly obtaining enough data to discover reliably whether such small effects actually exist. The experimental conditions used in mass-media ESP studies are almost identical to most national lotteries (i.e. large numbers of people sitting at home attempting to guess the identity of a distant target) and so positive results from such studies would challenge the notion that lotteries are unpredictable. Meta-analysis of eight ESP studies conducted via the mass media, representing over 1.5 million individual trials, indicate a very low, negative effect size ( $\zeta/N^{1/2} = -.0046$ ) whose overall cumulative outcome did not differ significantly from chance expectation (Stouffer  $\zeta = -1.60$ ). The paper discusses the implications of these results for the debate about the existence of ESP and its practical implications for lottery organizers.

September 1937 saw the beginning of one of the largest tests of extrasensory perception (ESP) ever conducted (Goodfellow, 1938). Every Sunday evening for 15 weeks, the Zenith Radio Corporation made a nationwide broadcast across the USA, inviting listeners to test their ESP ability by trying to guess the sequence of a short series of randomly ordered 'target' stimuli (e.g. dark or light, hot or cold) sequestered in the studio. Over 46000 members of the public responded, making more than one million individual guesses. Over the past 60 years other researchers have carried out similar ESP experiments via newspapers, magazines, radio and television. The results of these experiments have both theoretical and practical importance.

Meta-analyses of many different types of parapsychological study have suggested that ESP might exist, albeit as a small effect (Bem & Honorton, 1994; Honorton, 1985; Honorton & Ferrari, 1989; Milton, 1997; Stanford & Stein, 1994; see, however, Milton & Wiseman, in press). Highly statistically significant cumulated

\* Requests for reprints should be addressed to Dr Julie Milton, Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK.

results, strong enough to be unlikely to be due to publication bias, have been found, for example, in forced-choice studies of precognition, in which participants attempted to identify target stimuli to be randomly chosen at some future time (Honorton & Ferrari, 1989). Study quality in these meta-analyses has not generally shown any obvious relationship to effect size but problems inherent in all meta-analyses, such as the confounding of variables and imperfect reliability of assigning quality credit to studies that often give only partial information of interest to the meta-analyst (e.g. Eysenck, 1994; Orwin & Cordray, 1985; Stanford & Stein, 1994; Wanous, Sullivan & Malinak, 1989), stand in the way of these meta-analyses acting as conclusive proof of the existence of ESP. Confirmation of their findings is therefore crucial. Given the small size of the suggested effects, they would only be expected to be reliably detected in statistically powerful studies containing large numbers of trials. Mass-media ESP experiments generate an enormous number of guesses and thus present researchers with a quick way of obtaining enough data to assess reliably whether such effects actually exist.

However, mass-media ESP studies involve people sitting at home attempting to guess the identity of a distant target and not discovering whether they were correct until days, or even weeks, later. Such studies are quite unlike the vast majority of parapsychological studies. In most ESP studies, participants are tested in a laboratory, are perhaps a room or two away from the target and receive feedback within minutes of making their guesses. It is not clear, therefore, whether these studies would be expected to confirm the findings of the other parapsychological studies, assuming that those findings were due to a genuine ESP effect. It is important, therefore, to establish whether mass-media studies yield an effect. However, this is not only important from a scientific point of view but also from a very practical one that affects many people's pockets. Mass-media ESP studies are almost identical in their procedures to most national lotteries. Lottery organizers go to great lengths to ensure that the system used for drawing winning numbers is properly random and thus unpredictable. Positive results from mass-media ESP studies would suggest that lotteries may be far from unpredictable and would alter the way in which the public perceived and attempted to win national lotteries.<sup>1</sup>

To date, there has been no meta-analytic evaluation of mass-media ESP studies and so the question of whether such studies demonstrate an ESP effect remains unanswered. This paper presents such an analysis, first outlining some of the rather unusual methodological and statistical issues involved in evaluating certain mass-media studies, and then discussing the theoretical and practical implications of the results.

### **Retrieval of studies**

Study retrieval was based on the published literature. Three methods of retrieval were used.

<sup>1</sup> It would, of course, be very interesting to obtain data from national lotteries themselves to address both the general question of whether mass-media recruitment ESP studies work and the specific issue of whether ESP can be used to perform better than expected by chance in national lotteries. Unfortunately, Camelot (personal communication to first author, 17 July 1998) and other nations' lottery companies will not release data on how many people call each number each week in the lottery, which are necessary to correct each lottery's outcome for non-independence of calls on a shared target sequence (see the discussion below of the 'stacking effect').

First, studies were traced through *Parapsychology Abstracts International* (now *Exceptional Human Experience*; keywords: magazine, mass-media, newspaper, radio, television), a specialized abstracts journal that covers parapsychological material published both within and outside the parapsychology literature from before the turn of the century to the present day.

Second, papers in the four main peer-reviewed English-language parapsychology journals (*European Journal of Parapsychology*, *Journal of the American Society for Psychical Research*, *Journal of Parapsychology* and *Journal of the Society for Psychical Research*) were examined individually as far back as 1960, as were abstracts of peer-reviewed papers presented at the annual convention of the Parapsychological Association published in *Research in Parapsychology* since its first issue in 1972.

Third, the studies identified by the first two methods were examined for references to other mass-media recruitment studies.

### *Criteria for inclusion*

A mass-media recruitment study was defined as one in which participants were directly invited to volunteer themselves via magazines, newspapers, television or radio. Only studies that reported an outcome measure based on the number of correct guesses for all participants were included.

It has long been known that in studies which involve a single-target sequence that all participants attempt to guess, participants' responses cannot be assumed to be independent of each other; preference factors may lead people to call the same target choice as each other on any given trial, particularly when there are few trials (e.g. Marks & Kammann, 1980; Woolley, 1926). The potential strength of such factors was demonstrated by Lund (1939) in a study of response bias in which 596 participants each guessed a short (6-trial) random sequence of the five symbols used on ESP cards (star, square, wavy lines, cross, circle). He found that the most frequently chosen symbol, the star, was chosen 32% of the time, almost twice as often as the next most frequently chosen symbol, the circle, and well above the 20% predicted if participants responded at random.

This 'stacking' of responses on certain symbols can lead to spuriously high or low ESP study outcomes if experimenters assume independence of calls and use the theoretical variance rather than the observed variance of the distribution of responses. The difference between theoretical and observed variance can be extreme. In a study by Nash (1964), the observed variance of the responses, calculated by the Greville (1944) method (see Pratt, 1954 and Burdick & Kelly, 1977, for a less technical exposition), was 895 times the expected theoretical value. Not applying the Greville correction would have led to a 30-fold overestimate of the study outcome's standard normal deviate and effect size. Because of the necessity of applying the Greville correction for an accurate measure of a study's outcome, studies in which participants shared a single target sequence were not included unless their outcome measure had been calculated using the Greville correction, or unless they provided sufficient data for the present authors to apply the correction.

In addition to exclusions because of these general criteria, two further studies were excluded because they were severely problematic for meta-analytic purposes. First, a

**Table 1.** Quality, significance levels and effect sizes of studies in the database

Study	$z$	Effect size ( $z/N^{1/2}$ )
Green, Eastman & Adams (1963)	-0.97	-.0227
Brier (1967), confirmatory experiment	-1.64	-.0172
Nash (1964)	-0.87	-.0058
Rhine (1962)	-2.53	-.0046
Schmeidler (1980) (i) <sup>a</sup>	-0.83	-.0026
Michie & West (1957)	-0.04	-.0002
Goodfellow (1938) <sup>b</sup>	1.07	.0010
Schmeidler (1980) (ii) <sup>a</sup>	1.28	.0156

<sup>a</sup> Schmeidler provides data concerning the overall number of correct guesses only in a fuller version of this paper reported in the convention proceedings (Schmeidler, 1979); Table 4 of the paper provides data for two subgroups of participants who, combined, include all those in the study who fully completed the ESP test. The two groups are (i) those who provided complete questionnaire data, and (ii) those who did not.

<sup>b</sup> The Greville correction was applied to the data published by Goodfellow in Table 1, p. 603. Data from 2 of the 15 broadcasts are not included (October 3 and December 26) because not enough information was given to calculate the correction for those dates.

study by Medhurst, Stark & Thompson (1965) was excluded because participants were not tested in the usual manner but were sent packs of cards to shuffle and test themselves (this informal method was used for the purposes of screening participants for later research, and was not considered by the authors themselves to be a stringent test of ESP). Because any results could have so very easily have been attributed to cheating or unintentional error, this study was not included in the meta-analysis. Second, a study by Wiklund & Jacobson (1976) was also excluded because it is not clear on which data to base an effect size calculation. Due to a misunderstanding between the two authors, two target sequences, and therefore two possible outcomes, were generated. In addition, each participant's two targets were not independent, and therefore should not be analysed as such. However, there is no basis on which to choose one target or the other as the single target for analysis.

#### *Characteristics of the database*

Eight studies, conducted by seven different authors, fulfilled the inclusion criteria. Altogether, at least 86000 unselected participants (range in study from 91 to at least 46433) took part in a total of over 1.5 million individual trials (range in study from 1830 to 1204400). The number of trials per participant ranged from 6 to 100.

#### *Overall cumulation*

The effect size measure used was  $z/N^{1/2}$ , where  $z$  is the standard normal deviate associated with the number of correct guesses observed, and  $N$  is the number of trials

in a study. The mean effect size was  $-0.0046$  ( $SD = 0.011$ ). The overall combined Stouffer  $z$  (obtained by summing each study's standard normal deviate and dividing by the square root of the number of studies: see e.g. Mosteller & Bush, 1954) of the eight studies was  $-1.60$ , which is non-significantly below chance at  $\alpha = .05$ . The studies' individual significance levels and effect sizes are presented in Table 1, in order of increasing effect size.<sup>2</sup>

### Discussion

Despite comprising only eight studies, these experiments constitute an enormous database of over 1.5 million individual trials. However, despite having large statistical power, the cumulated effect size does not significantly differ from chance and thus does not provide any evidence to support the existence of ESP.

Meta-analyses of other groups of ESP studies have produced highly statistically significant outcomes. The dramatic difference between the outcome of these analyses and the results reported in this paper could be resolved in one of two ways. It is possible that the other meta-analyses reflect a genuine ESP effect and that the mass-media studies failed to find this effect because they were carried out under very different conditions. Alternatively, it may be that ESP does not exist and that the mass-media studies accurately estimate its effect size as indistinguishable from zero. In this scenario, the positive results of the apparently successful meta-analyses would be due to methodological flaws.

Either way, the present meta-analysis does not provide evidence of psychic functioning under the conditions used in mass-media studies. Whether this finding is seen as positive or negative depends upon one's point of view. The result will no doubt disappoint those who hoped that this unusual methodology might provide a quick and easy way of proving the existence of ESP. At the same time, the finding will provide considerable relief to lottery organizers worldwide.

### Acknowledgements

The authors are grateful to Robert L. Morris and Gertrude R. Schmeidler for their helpful comments on an earlier version of this paper.

### References

- Barrington, M. R. & Markwick, B. (1995). A television ESP experiment. *Journal of the Society for Psychical Research*, **60**, 267–269.
- Bem, D. J. & Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin*, **115**, 4–18.
- Brier, B. (1967). A correspondence ESP experiment with high-I.Q. subjects. *Journal of Parapsychology*, **31**, 143–148.

<sup>2</sup> Updating and extending the literature search to include *PsycLit* (1887–1998, using the keywords extrasensory perception, precognition, clairvoyance or telepathy in combination with mass media, television, radio, newspapers or magazines) as suggested by an anonymous referee yields only one additional eligible study (Barrington & Markwick, 1995). Its outcome was non-significant (unconservatively  $p(\text{one-tailed}) = .10$ ,  $z = 1.26$ , effect size =  $0.03$ ). If it had been included in the meta-analysis it would have raised the overall Stouffer  $z$  slightly closer to chance to  $-1.09$ .

- Burdick, D. S. & Kelly, E. F. (1977). Statistical methods in parapsychological research. In B. B. Wolman (Ed.), *Handbook of parapsychology*, pp. 81–130. New York: Van Nostrand Reinhold.
- Eysenck, H. J. (1994). Meta-analysis and its problems. *British Medical Journal*, **309**, 789–792.
- Goodfellow, L. D. (1938). A psychological interpretation of the results of the Zenith radio experiments in telepathy. *Journal of Experimental Psychology*, **23**, 601–632.
- Green, C. E., Eastman, M. & Adams, S. (1963). A new use for mass media in parapsychology. *Journal of the Society of Psychical Research*, **42**, 114–124.
- Greville, T. N. E. (1944). On multiple matching with one variable deck. *Annals of Mathematical Statistics*, **15**, 432–434.
- Honorton, C. (1985). Meta-analysis of psi ganzfeld research: A response to Hyman. *Journal of Parapsychology*, **49**, 51–91.
- Honorton, C. & Ferrari, D. C. (1989). 'Future telling': A meta-analysis of forced-choice precognition experiments, 1935–1987. *Journal of Parapsychology*, **53**, 281–308.
- Lund, F. H. (1939). Extra-sensory perception another name for free-association? *Journal of General Psychology*, **20**, 235–238.
- Marks, D. & Kammann, R. (1980). *The psychology of the psychic*. Buffalo, NY: Prometheus.
- Medhurst, R. G., Stark, H. & Thompson, G. T. (1965). A large-scale experiment to discover agents and percipients for ESP testing. *Journal of the Society for Psychical Research*, **43**, 109–135.
- Michie, D. & West, D. J. (1957). A mass ESP test using television. *Journal of the Society for Psychical Research*, **39**, 113–133.
- Milton, J. (1997). Meta-analysis of free-response ESP studies without altered states of consciousness. *Journal of Parapsychology*, **61**, 279–319.
- Milton, J. & Wiseman, R. (in press). Does psi exist? Lack of replication of an anomalous process of information transfer. *Psychological Bulletin*.
- Mosteller, F. M. & Bush, R. R. (1954). Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology: Vol. 1. Theory and method*, pp. 289–334. Cambridge, MA: Addison-Wesley.
- Nash, C. B. (1964). A television test on ESP. *International Journal of Parapsychology*, **6**, 139–142.
- Orwin, R. G. & Cordray, D. S. (1985). Effects of deficient reporting on meta-analysis: A conceptual framework and reanalysis. *Psychological Bulletin*, **97**, 134–147.
- Pratt, J. G. (1954). The variance for multiple-calling ESP data. *Journal of Parapsychology*, **18**, 37–40.
- Rhine, J. B. (1962). The precognition of computer numbers in a public test. *Journal of Parapsychology*, **26**, 244–251.
- Schmeidler, G. R. (1979). A mass ESP contest: Demographic data and stimulus-response bias. In *Parapsychological Association 22nd Annual Convention: Proceedings of presented papers*. Toronto: Parapsychological Association.
- Schmeidler, G. R. (1980). A mass ESP contest: Demographic data and stimulus-response bias. In W. G. Roll (Ed.), *Research in parapsychology 1979*, pp. 104–107. Metuchen, NJ: Scarecrow Press.
- Stanford, R. G. & Stein, A. G. (1994). A meta-analysis of ESP studies contrasting hypnosis and a comparison condition. *Journal of Parapsychology*, **58**, 235–269.
- Wanous, J. P., Sullivan, S. E. & Malinak, J. (1989). The role of judgement calls in meta-analysis. *Journal of Applied Psychology*, **74**, 259–264.
- Wiklund, N. & Jacobson, N. O. (1976). A public experiment with precognition. *Journal of the Society for Psychical Research*, **48**, 293–300.
- Woolley, V. J. (1926). The broadcasting experiment in mass-telepathy. *Proceedings of the Society for Psychical Research*, **38**, 1–9.

Received 13 March 1998; revised version received 12 August 1998