Parapsychology[1]

Monica J. Harris Robert Rosenthal

Harvard University

2024, Vol. 4, No. 1, pp. 18–33

Abstract: Monica J. Harris and Robert Rosenthal were commissioned by the National Research Council to conduct meta-analyses and review five areas of potential human enhancement. Despite finding that ganzfeld psi research followed the most rigorous protocols they were pressured to withdraw their supportive evaluation of psi, but refused to do so. This is their original report, with only very minor formatting changes. Among other things, Harris and Rosenthal concluded that "it would be implausible to entertain the null given the combined *p* from these 28 studies… when the accuracy rate expected under the null is 1/4, we estimate the obtained accuracy rate to be about 1/3." They were then asked to analyze the effect of potential design and procedure flaws and, after doing so, they concluded that: "Our analysis of the effects of flaws on study outcome lends no support to the hypothesis that ganzfeld research results are a significant function of the set of flaw variables."

Keywords: Ganzfeld; parapsychology; anomalous cognition; meta-analysis; Robert Rosenthal

There are two transition points in the recent history of parapsychology. At each point parapsychology advanced to a new level of more rigorous research and scientific respectability, though neither point earned for it full acceptance as a respectable field of scientific inquiry (Boring, 1962; Murphy, 1962, 1963; Truzzi, 1981). The first point was in 1882 when the Society for psychical Research was founded in London by a group primarily from Cambridge University. Among the distinguished presidents of this Society were William James, Henri Bergson, Lord Rayleigh, and C.D. Broad (Schmeidler, 1968). The second point was in 1927 when William McDougall, newly arrived at Duke University, was joined by J. B. Rhine (Boring, 1950; Schmeidler, 1968). It was Rhine who established the basic procedures of parapsychological research that are still employed

1 The authors are deceased but had earlier granted the editor of ***JAEX*** permission to publish this contribution. Because their original paper did not have an abstract, the journal editor added it for indexing purposes and left the paper, formatting and all, almost exactly as originally written.

https://doi.org/10.31156/jaex.26222

today. His best known method required the subject to guess which one of five designs was the “target” stimulus. Since the probability of guessing the correct design was .20, any subject’s “psi” ability could be evaluated for statistical significance by comparing the obtained success rate with the .20 expected under the null hypothesis.

Parapsychological investigations cover a wide variety of phenomena including: telepathy (e.g., guessing a design being viewed by another); clairvoyance (e.g., guessing a design not being viewed by another); precognition (e.g., guessing a design not yet selected); psychokinesis (e.g., trying to influence the fall of a pair of dice); survival after death (e.g., reincarnation). The. first three of these are often referred to generically as ESP (extrasensory perception). Because the types of research subsumed under the topic of parapsychology range so widely, and because of the sheer number of parapsychological investigations, we have confined our discussion to a focused domain of parapsychological inquiry: the ganzfeld experiments.

Ganzfeld Experiments

In these experiments subjects typically are asked to guess which of four stimuli had been “transmitted” by an agent or sender with these guesses made under conditions of sensory restriction (usually white noise and an unpatterned visual field). There were several strong reasons for the selection of this domain of parapsychological research:

1. The domain is of recent origin so that even the earliest studies managed to avoid some of the older problems found in parapsychological research (Hansen & Lehmann, 1895; Kennedy, 1938, 1939; Moll, 1898; Rosenthal, 1965, 1966; Warner & Raible, 1937).
2. Because of the recency of the research, access to original data was more likely than for some of the older areas (Rao, 1985).
3. The domain is considered an especially promising area of parapsychological inquiry (Hyman, 1985; Rao, 1985).
4. Investigations in this area have been carried out by respected researchers (Hyman, 1985).
5. The area has been the subject of recent sophisticated public debate by eminent investigators and critics of the area (Honorton, 1985; Hyman, 1985; Rao, 1985).
6. As an outgrowth of this debate, two formal meta-analyses of this area have become available (Honorton, 1985; Hyman, 1985).

Journal of Anomalous Experience and Cognition (JAEX) 2024, Vol. 4, No. 1, pp. 18–33

Meta-Analytic Results

Five indices of "psi" success have been employed in ganzfeld research (Honorton, 1985). One criticism of research in this area is that some investigators employed several such indices in their studies and failed to adjust their reported levels of significance (*p*) for the fact that they had made multiple tests (Hyman, 1985). Since most studies employed a particular one of these five methods, the method of direct hits, Honorton focused his meta-analysis on just those 28 studies (of a total of 42) for which direct hit data were available.

The method of direct hits scores a success only when the single correct target is chosen out of a set of t total targets. Thus the probability of success on a single trial is $1/t$ with t usually = 4 but sometimes 5 or 6. The other methods, employing some form of partial credit, appear to be more precise in that they use more of the information available. Although they differ in their interpretation of the results, Honorton (1985) and Hyman (1985) agree quite well on the basic quantitative results of the meta-analysis of these 28 studies. This agreement holds both for the estimation of statistical significance (Honorton, 1985, p.58) and of effect size (Hyman, 1985, p. 13).

Stem-and-Leaf Display

Table 3 shows a stem-and-leaf display of the 28 effect slze estimates based on the direct hits studies summarized by Honorton (1985, p. 84). The effect size estimates shown in Table 3 are in units of Cohen's *h* which is the difference between (a) the arcsine transformed proportion of direct hits obtained and (b) the arcsine transformed proportion of direct hits expected under the null hypothesis (i.e., $1/t$). The advantage of *h* over *j*, the difference between raw proportions, is that all *h* values that are identical are identically detectable while all *j* values that are identical (e.g., .65-.45 and .25-.05) are not equally detectable (Cohen, 1977, p. 181).

The stem-and-leaf display of Cohen's *h* values is shown on the left and the display is summarized on the right. Tukey (1977) developed the stem-and-leaf plot as a special form of frequency distribution to facilitate the inspection of a batch of data. Each number in the data batch is made up of one stem and one leaf, but each stem may serve several leaves. Thus, the stem .1 is followed by leaves of 3, 8, 8 representing the numbers .13, .18, .18. The first digit is the stem; the next digit is the leaf. The stem-and-leaf display functions as any other frequency distribution but the original data are retained precisely.

TABLE 3 Stem-and-Leaf Plot and Statistical Summary of Ganzfeld Studies Employing Criterion of Direct Hits

Cohen's *h*

Stem	Leaf
1.4	4
1.3	3
1.2	
1.1	
1.0	
.9	
.8	
.7	3
.6	
.5	8
.4	0 2 2 2 4
.3	1 2 2 4 4 7 8
.2	2
.1	3 8 8
.0	7 7 9
-.0	5
-.1	0
-.2	
-.3	2
-.4	0
-.5	
-.6	
-.7	
-.8	
-.9	3

Summary Statistics	
Maximum	1.44
Quartile$_3$ (Q_3)	.42
Median (Q_2)	.32
Quartile$_1$ (Q_1)	.08
Minimum	-.93
$Q_3 - Q_1$	.34
σ: [.75($Q_3 - Q_1$)]	.26
S	.45
Mean[a]	.28
N	28
Proportion positive sign	.82
Z of proportion positive	3.40
Combined Stouffer *Z*	6.60
t test of mean *Z*	3.23
Correlation between *h* and *Z*	.86
Correlation between *h* and raw *j*	.98
Number of studies with mean *Z* = 0.00 required to bring combined results to *p* > .05	423

Confidence Intervals*

	from	to
80%	.17	.39
95%	.11	.45
99%	.04	.52
99.9%	-.03	.59

* where *N* = 28 studies

[a] Unweighted; weighted mean = .23

Distribution. From Table 3 we see that the distribution of effect sizes· is unimodal with the bulk of the results (80%) falling between -.10 and .58. The distribution is nicely symmetrical with the skewness index ($g1$ = .17) only 24% of that required for significance at $p < .05$ (Snedecor & Cochran, 1980, pp. 78-79, 492). The tails of the distribution, however, are too long for normality with kurtosis index g_2=.2.04, p=.02. Relative to what we would expect from a normal distribution, we have studies that show larger positive and larger negative effect sizes than would be reasonable. Indeed, the two largest positive effect sizes are significant outliers at $p < .05$, and the largest negative effect size approaches significance with a Dixon index of .37 compared to one of .40 for the largest positive effect size (Snedecor & Cochran, 1980, pp. 279-280, 490). The total sample of studies is still small; however, if a much larger sample showed the same result, that would be a pattern consistent with the idea that both strong positive results ("psi") and strong negative results ("psi-missing") might be more likely to find their way into print or at least to be more available to a meta-analyst.

Effect size. The bulk of the results (82%) show a positive effect size where 50%

would be expected under the null (p = .0004). The mean effect size, h, of .28 is equivalent to having a direct hit rate of .38 when .25 was expected under the null. The 95% confidence interval suggests the likely range of effect sizes to be from .11 to .45, equivalent to accuracy of guessing rates of .30 to .46 when .25 was expected under the null hypothesis.

Significance testing. The overall probability that obtained accuracy was better than the accuracy expected under the null was a p of $3.37/10^{11}$ associated with a Stouffer z of 6.60 (Mosteller & Bush, 1954; Rosenthal, 1978a, 1984).

File drawer analysis. A combined p as low as that obtained can be used as a guide to the tolerance level for null results that never found their way into the meta-analytic data base (Rosenthal, 1979, 1984). It has long been believed that studies failing to reach statistical significance may be less likely to be published (Sterling, 1959; Rosenthal, 1966). Thus it may be that there is a residual of nonsignificant studies languishing in the investigators' file drawers. Employing simple calculations, it can be shown that, for the current studies summarized, there would have to be 423 studies with mean p =.50, one-tailed, or z = 0.00 in those file drawers before the overall combined p would become just > .05.

That many studies unretrieved seems unlikely for this specialized area of parapsychology (Hyman, 1985; Honorton, 1985). Based on experience with meta-analyses in other domains of research (e.g., interpersonal expectancy effects) the mean or effect size for nonsignificant studies is not 0.00 but a value pulled strongly from 0.00 toward the mean Z or mean effect size of the obtained studies (Rosenthal & Rubin, 1978).

Comparison to an Earlier Meta-Analysis

We felt it would be instructive to compare the results of the ganzfeld research meta-analysis by Honorton (1985) to the results of an older and larger meta-analysis of another controversial research domain --that of interpersonal expectancy effects (Rosenthal & Rubin, 1978). In that analysis, eight areas of expectancy effects were summarized; effect sizes (Cohen's d roughly equivalent to Cohen's h) ranged from .14 to 1.73 with a grand mean d of .70. Honorton's mean effect size (h =-.28) exceeds the mean d of two of the eight areas (reaction time experiments [d =.17]; and studies employing laboratory interviews [d =.14]).

The earlier meta-analysis displayed the distribution of the Z's associated with the obtained p levels. Table 4 shows a comparison of the two meta-analyses' distri-

Journal of Anomalous Experience and Cognition (JAEX) 2024, Vol. 4, No. 1, pp. 18–33

butions of *Z*s. It is interesting to note the high degree of similarity in the distributions of significance levels. The total proportion of significant results is somewhat higher for the ganzfe1d studies but not significantly so ((1) =1.07, *N* = 373, *p* = .30, =.05).

Interpretation of Meta-Analytic Results

Although the results of the meta-analysis are clear, the meaning of these results is open to various interpretations (Truzzi, 1981). The most obvious interpretation might be that at a very low *p*. and with a fairly impressive effect size, the ganzfeld psi phenomenon has been demonstrated. However, there are rival hypotheses that will need to be considered. many of them put forward in the recent detailed evaluation of the ganzfe1d research area by Hyman (1985).

Procedural Rival Hypotheses

Sensory leakage. A standard rival hypothesis to the hypothesis of ESP is that sensory leakage occurred and that the receiver was knowingly or unknowingly cued by the sender or by an intermediary between the sender and receiver. As early as 1895, Hansen and Lehmann (1895) had described "unconscious whispering" in the laboratory and Kennedy (1938, 1939) was able to show that senders in telepathy experiments could give auditory cues to their receivers quite unwittingly. Ingenious use of parabolic sound reflectors made this demonstration possible. Moll (1898), Stratton (1921), and Warner and Raible (1937) all gave early warnings on the dangers of unintentional cueing (for summaries see Rosenthal, 1965a, 1966). The subtle kinds of cues described by these early workers were just the kind we have come to look for in searching for cues given off by experimenters that might serve to mediate the experimenter expectancy effects found in laboratory settings (Rosenthal, 1966, 1985).

By their nature, ganzfeld studies tend to minimize problems of sensory cueing. An exception occurs when the subject is asked to choose which of four (or more) stimuli had been "sent" by another person or agent. When the same stimuli held originally by the sender are shown to the receiver, finger smudges or other marks may serve as cues. Honorton has shown, however, that studies controlling for this type of cue yield at least as many significant effects as do the studies not controlling for this type of cue.

Recording errors. A second rival hypothesis has nearly as long a history. Kennedy and Uphoff (1939) and Sheffield and Kaufman (1952) both found biased errors of recording the data of parapsychological experiments. In a meta-analysis of 139,000 recorded observations in 21 studies, it was found that about 1% of all observations were in error and, that of the errors committed, twice as many favored the hypothesis as opposed it (Rosenthal, 1978b). While it is difficult to rule recording error out of ganzfeld studies (or any other kind of research) their magnitude is such that they could probably have only a small biasing effect on the estimated average effect size (Rosenthal, 1978b, p. 1007).

Intentional error. The very recent history of science has reminded us that while fraud in science is not quite of epidemic proportion it must be given close attention (Broad & Wade, 1982; Zuckerman, 1977). Fraud in parapsychological research has been a constant concern, a concern found justified by periodic flagrant examples (Rhine, 1975). In the analyses of Hyman (1985) and Honorton (1985), in any case, there appeared to be no relationship between degree of monitoring of participants and the results of the study.

Statistical Rival Hypotheses

File drawer issues. The problem of biased retrieval of studies for any meta-analysis was described earlier. Part of this problem is addressed by the 10 year old norm of the Parapsychological Association of reporting negative results at its meetings and in its journals (Honorton, 1985). Part of this problem is addressed also by Blackmore who conducted a survey to retrieve unreported ganzfeld studies. She found that 7 of her total of 19 studies (37%) were judged significant overall by the investigators. This proportion of significant results was not significantly (or appreciably) lower than the proportion of published studies found significant. A problem that seems to be a special case of the file drawer problem was pointed out by Hyman (1985). That was a possible tendency to report the results of pilot studies along with subsequent significant results when the pilot data were significant. At the same time it is possible that pilot studies were conducted without promising results, pilot studies that then found their way into the file drawers. In any case, it is nearly impossible to have an accurate estimate of the number of unretrieved studies or pilot studies actually conducted. Chances seem good, however, that there would be fewer than the 423 results of mean $z = 0.00$ required to bring the overall combined g to > .05.

Multiple testing. Each ganzfeld study may have more than one dependent variable for scoring a success. If investigators employ these dependent variables sequentially until they find one significant at $p < .05$ the true p will be higher than .05 (Hyman, 1985). Although a simple Bonferroni procedure can be used to adjust for this problem

Journal of Anomalous Experience and Cognition (JAEX) 2024, Vol. 4, No. 1, pp. 18–33

(e.g., by multiplying the lowest obtained by the number of dependent variables tested) this adjustment is quite conservative (Rosenthal & Rubin, 1983). The adjustment can be made with greater power if the investigators are willing to order or to rate their dependent variables on a dimension of importance (Rosenthal & Rubin, 1984, 1985). Most useful, however is a procedure that uses all the data from all the dependent variables with each one weighted as desired so long as the weighting is done before the data are collected (Rosenthal & Rubin, 1986).

Randomization. Hyman (1985) has noted that the target stimulus may not have been selected in a truly random way from the pool of potential targets. To the extent that this is the case the *p* values calculated will be in error. Hyman (1985) and Honorton (1985) disagree over the frequency in this sample of studies of improper randomization. In addition, they disagree over the magnitude of the relationship between inadequate randomization and study outcome. Hyman felt this relationship to be significant and positive; Honorton felt this relationship to be nonsignificant and negative. Since the median *g* level of just those 16 studies employing random number tables or generators (z=.94) was essentially identical to that found for all 28 studies it seems unlikely that poor randomization procedures were associated with much of an increase in significance level (Honorton, 1985, p. 71).

Statistical errors. Hyman (1985) and Honorton agree that six of the 28 studies contained statistical errors. However, the median effect size of these studies (h=.33) was very similar to the overall median (h=.32) so that it seems unlikely that these errors had a major effect on the overall effect size estimate. Omitting these six studies from the analysis decreases the mean h from .28 to .26. Such a drop is equivalent to a drop of the mean accuracy rate from .38 to .37 when .25 is the expected value under the null.

Independence of studies. Because the 28 studies were conducted by only 10 investigators or laboratories, the 28 studies may not be independent in some sense. While under some data analytic assumptions such a lack of independence would have implications for significance testing, it does not in the ganzfeld domain because of the use of trials rather than subjects as the independent sampled unit of analysis. The overall significance level, then, depends on the results of all trials, not the number of studies, of subjects or investigators (any of which may be viewed as fixed rather than random).

However, the lack of independence of the studies could have implications for the estimation of effect sizes if a small proportion of the investigators were responsible for all the nonzero effects. In that case the average of the investigators' obtained effects would be much smaller than the average of the studies' obtained effects.

In an extreme example the median effect size of a sample of studies could be .50 while the median effect size of a sample of investigators could be zero because very few investigators obtained any nonzero effect. That did not turn out to be the case for the ganzfeld domain. The median effect size (*h*) was identical (.32) for the 28 studies and the 10 investigators or laboratories. The mean effect sizes, however, did differ somewhat with a lower mean for labs (.23) than for studies (.28). The proportions of results in the positive direction were very close: .82 for studies and .80 for labs.

It is of interest to note that investigators did differ significantly from one another in the magnitude of the effects they obtained with $F(9, 18) = 3.81$, $p < .01$, intra-class $p = .63$. There was little evidence to suggest, however, that those investigators tending to conduct more studies obtained higher mean effect sizes; the f(I, 18) testing that contrast was 0.38, $p = .54$, $r^2 = .14$.

Conclusion

On the basis of our summary and the very valuable meta-analytic evaluations of Honorton (1985) and Hyman (1985), what are we to believe? The situation for the ganzfeld domain seems reasonably clear. We feel it would be implausible to entertain the null given the combined *p* from these 28 studies. Given the various problems or flaws pointed out by Hyman and Honorton, the true effect size is almost surely smaller than the mean *h* of .28 equivalent to a mean accuracy of 38% when 25% is expected under the null. We are persuaded that the net result of statistical errors was a biased increase in estimated effect size of at least a full percentage point (from 37% to 38%). Furthermore, we are persuaded that file drawer problems are such that some of the smaller effect size results have probably been kept off the market. If pressed to esti-

2 After preparation of this paper we learned of a possible problem in the randomization procedures employed by the investigator contributing the largest number (9) of Ganzfeld studies to the set of 28 summarized in this section. Accordingly we constructed Table 4a to investigate the effect on the mean and median effect sizes of omitting all the studies conducted by this investigator. The top half of Table 4a shows this effect when we consider the 28 studies, as the units of analysis. Omitting the 9 questioned studies lovers the mean effect size from .28 to .26 and does not change the median effect size which remains at .32. The lower half of Table 4a shows this effect when we consider the 10 investigators as the units o£ analysis. Omitting the investigator in question lowers the mean effect size from .23 to .22 but raises the median effect size from .32 to .34. It seems clear that the questioned randomization of the 9 studies of this investigator cannot have contributed substantially to an inflation of the overall effect size.

Journal of Anomalous Experience and Cognition (JAEX) 2024, Vol. 4, No. 1, pp. 18–33

mate a more accurate effect size we might think in terms of a shrinkage of *h* from the obtained value of .28 to perhaps an *h* of .18. Thus, when the accuracy rate expected under the null is 1/4, we estimate the obtained accuracy rate to be about 1/3.[2]

Postscript

We have been asked to respond to a letter from the committee listing questions about the presence and consequences of methodological flaws in the ganzfeld studies discussed by Honorton (985), Hyman (1985), Hyman and Honorton (1986), Rosenthal (1986), and by the present authors of this paper. Our response is in two parts. In part 1, we examine the likely effects of flaws on the meta-analytic results of the ganzfeld studies. In part 2, we examine the results of a series of new studies designed to address the flaws discussed by Hyman and Honorton in their individual and joint papers.

Flaw Effects

The committee has called our attention to possible flaws in the randomization procedure employed by Sargent and his colleagues. In its letter it noted that Honorton agreed with Hyman about the assignment of these randomizations flaws to the Sargent study. However, Honorton states in two letters that this agreement was not reached (personal communications of November 25, 1987, and January 5, 1988). Apparently, experts on the ganzfeld research disagree on whether the Sargent studies' randomization procedures are flawed given all the evidence available to both, evidence which is summarized in papers by Blackmore (1987), Harley and Matthews (1987), and Sargent (1987).

For purposes of this postscript and the following data analyses, we are going to assume that Hyman is correct in his assignment of randomization flaws and all other flaws he assigned in his 1985 paper. The heart of the matter is the relationship of flaws to research results and that is what our analyses are designed to investigate. In a 1986 manuscript, Hyman suggested that the relationship of flaws to study outcomes should be examined in a multivariate manner. Accordingly, that is the nature of our analyses in our first pass effort to examine the likelihood that methodological flaws are driving the results of the ganzfeld studies to an appreciable degree.

Canonical analysis. The most general of the multivariate procedures examines the maximum relationship that can be found between two sets of variables, for example, a set of predictor variables and a set of outcome variables. In our analysis the predictor variables were Hyman's (1985) flaw variables of documentation (DOC), feedback (FB), randomization (R), security (SEC), single target (ST), and statistical analysis (STAT), all coded as 0 if adequate or l if not adequately done or not adequately specified. The outcome variables were the significance level ~ and the effect size Cohen's *h*. The adjusted canonical correlation was only .46, a magnitude that for two-predicted-from-six could have arisen under the null hypothesis 54 times out of 100 $F(12,40) = 0.91$). Interestingly, three of the six flaw variables correlated positively with the flaw canonical variable and with the outcome canonical variable (DOC, FB, R) but three correlated negatively (SEC, ST, STAT). Thus, the canonical analysis gives no support to the hypothesis that the research results are a significant function of the set of flaw variables.

Regression analysis. Separate analyses were also done for each of the outcome variables *Z* and *h*. The battery of predictor variables correlated only .44 with Cohen's *h* (F(6,2l) = 0.84, $p = .56$) and .57 with *z* (F(6,2l) = 1.65, $p = .18$). For neither of the outcome variables did any of the six predictors account for a significant proportion of variance either in zero-order form or after partialing. Since there were two methods of partialing employed, a total of 36 (3 methods x 6 predictors x 2 outcome variables) *t*'s were computed, none of which reached the .05 level. Regression analyses, therefore, gave no more support than did the canonical analysis to the hypothesis that ganzfeld research results are a significant function of the set of flaw variables.

New Evidence

Hyman (1985) and Honorton (1985) were agreed (Hyman and Honorton, 1986) that new studies were needed that would take account of the flaws they had found in their critiques of earlier research. Since our present paper was completed we have learned of a series of 10 new studies conducted by Honorton, one of the four investigators singled out by the Committee on Techniques for the Enhancement of Human Performance as among the best in the country (Druckman and Swets l 1988, p. 22).

The series of 10 ganzfeld studies yielded a combined *z* of 2.791, $p = .0026$ and a mean *h* of .23. This effect size, based on 10 studies, is only slightly smaller than the mean effect size of Sargent's nine studies ($h = .30$) and is very close to the mean effect size of the remaining 19 studies ($h = .26$; see Table 4a). For the original 28 studies plus the 10 new ones from Honorton's lab l the combined *z* is now 7.10 and the mean effect size is now an *h* of .27. Omitting Sargent's nine studies changes matters very little-- *z* is now 5.74 and $h = .25$. In short, the new evidence based on studies designed

Journal of Anomalous Experience and Cognition (JAEX) 2024, Vol. 4, No. 1, pp. 18–33

to meet earlier methodological objections is very consistent with the earlier evidence and makes the null hypothesis still more implausible.

Conclusion

Our analysis of the effects of flaws on study outcome lends no support to the hypothesis that ganzfeld research results are a significant function of the set of flaw variables. In addition, a series of 10 new studies designed to control for earlier presumed flaws yielded results quite consistent with the original set of 28 studies.

References

Blackmore, S. (1987). A report of a visit to Carl Sargent's laboratory. *Journal of the Society for Psychical Research, 54*, 186-198.

Druckman, D., & Swets, J. A. (Eds.) (1988). *Enhancing human performance. Issues, theories, and techniques*. Washington, DC: National Academy Press.

Harley, T., & Matthews, G. (1987). Cheating, psi, and the appliance of science: A reply to Blackmore. *Journal of the Society for Psychical Research, 54*, 199-207.

Hyman, R. (1986). *To conclude or not to conclude*: *A reply to the commentators.* Unpublished manuscript, University of Oregon.

Hyman, R., & Honorton, C. (1986). A joint communiqué: The psi ganzfeld controversy. *Journal of Parapsychology, 50*, 351-364.

Rosenthal, R. (1986). Meta-analytic procedures and the nature of replication: The ganzfeld debate. *Journal of Parapsychology*, *50*, 315-336.

Sargent, C. (1987). Sceptical fairytales from Bristol. *Journal of the Society for Psychical Research*, *54*, 208-218.

Table 4.

Proportion of Studies Reaching Critical Levels of Significance for Two Research Areas

Interval for *Z*	Expected Proportion	Expectancy Research[a]	Ganzfeld Research[b]	Difference
Unpredicted Direction				
-1.65 and below	.05	.03	.07	+.04
NotSignificant				
-1.64 to + 1.64	.90	.60	.50	-.10
Predicted Direction				
+1.65 and above	.05	.36	.43	+.07
+2.33 and above	.01	.19	.25	+.06
+3.09 and above	.001	.12	.18	+-06
+3.72 and above	.0001	.07	.04	-.03

[a] ***N = 345*** studies; from Rosenthal & Rubin (1978). [b] ***N*** = 28 studies; from Honorton (1985).

Table 4a.

Effects on Effect Size (h) of Removing Studies by Sargent

	Mean	*Median*
Sargent's Studies (N = 9)	.30	.37
Analysis by Studies		
Including Sargent (*N* = 28)	.28	.32
Omitting Sargent (*N* = 19)	*.26*	*.32*
Difference	.02	.00
Analysis by Investigators		
Including Sargent (*N* = 10)	.23	.32
Omitting Sargent (*N* = 9)	*.22*	*.34*
Difference	.01	-.02

Parapsychologie

Monica J. Harris Robert Rosenthal

Résumé: Monica J. Harris et Robert Rosenthal ont été chargés par le National Research Council (« Conseil National de la Recherche ») de mener des méta-analyses et d'examiner cinq domaines d'amélioration potentielle de l'être humain. Alors que les auteurs ont constaté que les recherches psi utilisant le Ganzfeld avaient la rigueur méthodologique la plus importante, ils ont subi des pressions pour retirer leur évaluation favorable à la psi. Ce que les auteurs ont refusé de faire. Le rapport publié est le rapport original, qui n'a subi que des modifications mineures de mise en forme. Entre autres choses, Harris et Rosenthal ont conclu que «il serait peu plausible d'accepter le résultat nul étant donné les p combinés de ces 28 études... lorsque le taux de réussite attendu pour le résultat nul est de 1/4, nous estimons que le taux de réussite obtenu est d'environ 1/3». Il leur a ensuite été demandé d'analyser l'effet des défauts potentiels de conception et de procédure et, après l'avoir fait, ils ont conclu que: «Notre analyse de l'effet des défauts sur les résultats obtenus ne confirme pas l'hypothèse selon laquelle les résultats issus de la méthode ganzfeld constituent en soi une catégorie parmi l'ensemble des variables possibles de défauts.»

Translation into French by Antoine Bioy, Ph. D.

Journal of Anomalous Experience and Cognition (JAEX) 2024, Vol. 4, No. 1, pp. 18-33

Parapsychologie

Monica J. Harris Robert Rosenthal

Zusammenfassung: Monica J. Harris und Robert Rosenthal von der Harvard University wurden vom Nationalen Forschungsrat beauftragt, Meta-Analysen durchzuführen und fünf Bereiche des potenziellen Human Enhancement zu überprüfen. Obwohl sie feststellten, dass die Ganzfeld-Psi-Forschung den strengsten Protokollen folgte, wurden sie unter Druck gesetzt, ihre positive Bewertung von Psi zurückzuziehen, was sie jedoch ablehnten. Dies ist ihr Originalbericht mit nur sehr geringfügigen Änderungen in der Formatierung. Unter anderem kamen Harris und Rosenthal zu dem Schluss, dass "es unplausibel wäre, die Null-Hyothese anzunehmen, wenn man die p-Werte aus diesen 28 Studien zusammennimmt... wenn die unter der Null-Hypothese erwartete Genauigkeitsrate 1/4 beträgt, schätzen wir die erhaltene Genauigkeitsrate auf etwa 1/3." Sie wurden dann gebeten, die Auswirkungen möglicher Design- und Verfahrensfehler zu analysieren, und kamen zu folgendem Schluss: "Unsere Analyse der Auswirkungen von Fehlern auf das Studienergebnis unterstützt nicht die Hypothese, dass die Ganzfeld-Forschungsergebnisse eine signifikante Funktion einer Anzahl von Fehlervariablen sind."

Translation into German by Eberhard Bauer, Ph. D.

Parapsicologia

Monica J. Harris Robert Rosenthal

Resumo: Monica J. Harris e Robert Rosenthal foram designados pelo Conselho Nacional de Pesquisa para realizar meta-análises e revisar cinco áreas de potencial aprimoramento humano. Apesar de terem constatado que a pesquisa Psi Ganzfeld seguia os protocolos mais rigorosos, foram pressionados retirar sua avaliação favorável ao psi, mas se recusaram a fazê-lo. Este é seu relatório original, com apenas mudanças de formatação muito pequenas. Entre outras coisas, Harris e Rosenthal concluíram que "seria implausível considerar a hipótese nula dada a combinação de p desses 28 estudos... quando a taxa de precisão esperada sob a hipótese nula é de 1/4, estimamos que a taxa de precisão obtida seja cerca de 1/3." Eles foram, então, solicitados a analisar o efeito de possíveis falhas de design e procedimentos e, depois de o fazê-lo, concluíram que: "Nossa análise sobre os efeitos das falhas no resultado do estudo não dá suporte à hipótese de que os resultados da pesquisa ganzfeld sejam uma função significativa do conjunto de variáveis de falha."

Translation into Portuguese by Antônio Lima, Ph. D.

Parapsicología

Monica J. Harris Robert Rosenthal

Resumen: Monica J. Harris y Robert Rosenthal recibieron el encargo por parte del National Research Council de realizar meta-análisis y revisar cinco áreas de potencial mejora humana. A pesar de descubrir que la investigación psi ganzfeld seguía los protocolos más rigurosos, se les presionó para que retiraran su evaluación de apoyo a psi, pero se negaron a hacerlo. Este es su informe original, con sólo unos pequeños cambios de formato. Entre otras cosas, Harris y Rosenthal concluyeron que "sería inverosímil considerar la nulidad dada la *p* combinada de estos 28 estudios... cuando la tasa de precisión esperada bajo la nulidad es de 1/4, estimamos que la tasa de precisión obtenida es de aproximadamente 1/3". A continuación, se les pidió que analizaran el efecto de posibles defectos de diseño y procedimiento y, tras hacerlo, concluyeron que: "Nuestro análisis de los efectos de los defectos sobre el resultado del estudio no apoya la hipótesis de que los resultados de la investigación ganzfeld sean una función significativa del conjunto de variables de los defectos".

Translation into Spanish by Etzel Cardeña, Ph. D.