

## ON PUBLICATION POLICY REGARDING NON-SIGNIFICANT RESULTS

Some comments on Dr. J.B. Rhine's article in the comments section of the J.P., 39, No 2, 135-142

Martin Johnson

University of Utrecht

I have read Dr. Rhine's article with the greatest of interest. However, on certain issues in his article, I respectfully disagree. At the most recent Annual Conference of the Parapsychology Foundation, Inc., held in San Francisco in August 1975, Dr. Rhine and I had the opportunity of exchanging our views regarding publication policy.

Normally I would not have felt motivated to criticize Dr. Rhine on parapsychological matters, but since I strongly disagree with him on certain issues, I will be glad to clarify my position. Finally in order to avoid all potential misunderstandings, I would like to make it clear that I have admired Dr. Rhine for his unmatched pioneering work in parapsychology for too long to easily turn critic at this stage. Further, I am indebted to Dr. Rhine in many ways for the inspiring and substantial help he has rendered me over the years. I am not sanguine about my possibilities of changing his opinion, but I hope that our exchange of opinions will be helpful in illuminating problems which I believe are of the utmost importance for our field.

On the strict statistical issue, I agree with Dr. Rhine (see pp 135 - 136 in his article), namely that so far as the test of significance is concerned, one experiment should be considered as independent of another, and the results do not need to be pooled. However, if one accepts this rule, I believe that one should be consequent, e.g. one should refrain from the habit of combining p-values from various, hardly comparable investigations with the intention of demonstrating how well the psi hypothesis is validated by actual findings. This behaviour is not seldom manifested by parapsychologists in their apologetic zest.

Before leaving the statistical issue it should be stressed that all laboratory investigations have both an experimental and a

statistical side. The use of statistics may be more or less proper, but even the most proper use of statistics may lead to spurious correlations or conclusions if there are inadequacies regarding the research process itself. One of these sources of error in the research process is related to selective reporting; another to human limitations with regard to the ability to make reliable observations or evaluations. Dunette (1) says:

"The most common variant is, of course, the tendency to bury negative results. I only recently became aware of the massive size of this great graveyard for dead studies when a colleague expressed gratification that only a third of his studies 'turned out' - as he put it. Recently, a second variant of this secret game was discovered, quite inadvertently, by Wolins, when he wrote to 37 authors to ask for the raw-data on which they had based recent journal articles. Wolins found that of the 32 who replied, 21 reported their data to be either misplaced, lost, or inadvertently destroyed. Finally, after some negotiation, Wolins was able to complete seven re-analyses on the data supplied from five authors. Of the seven, he found gross errors in three - errors so great as to clearly change the outcome of the results already reported."

It should also be stressed that Rosenthal and others have demonstrated that experimenters tend to arrive at results found to be in full agreement with their expectancies, or with the expectancies of those within the scientific establishment in charge of the rewards. Even if some of Rosenthal's results have been questioned the general tendency seems to be unaffected.

I guess we all can agree upon the fact that selective reporting in studies on the reliability and validity, of for instance a personality test, is a bad thing. But what could be the reason for selective reporting? Why does a research worker manipulate his data? Is it only because the research worker has a "weak" mind or does there exist some kind of "steering field" that exerts such an influence that improper behaviour on the part of the research worker occurs?

It seems rather reasonable to assume that the editors of professional journals or research leaders in general could exert a certain harmful influence in this connection. This may be especially true in areas of research where we find school-builders or paradigm-makers, or by and large in areas where a branch of science has to fight extra hard for its survival and recognition. There is no doubt at all in my mind about the "filtering" or "shaping" effect an editor may exert upon the output of his journal. It seems highly probable that he will be prone to champion his own "school of thought" or his cherished ideas. This implies that he prefers and "rewards" studies supporting his own

previously published findings and cherished hypotheses and tends to suppress the reporting of findings leaning in the opposite direction. Such an editorial policy may create a research climate and "system of values" which could account to a great extent for the phenomena described by Dunnette, Rosenthal, and others. I think that one should note that the very existence of a certain publication policy could influence an author's willingness to submit a paper for publication and his strategy in writing it.

To me the issue on how the publication policy may affect the content and output of the research procedure is a burning one. This is a highly relevant issue in most branches of science and I fail to see any good reasons why parapsychology should be exempted. On the contrary, I think it is especially relevant to parapsychology. As I see it, the major risk of selective reporting is not primarily a statistical one, but rather the research climate which the underlying policy create ("you are 'good' if you obtain supporting results; you are 'no-good' if you only arrive at chance results").

Such a policy, implicitly or explicitly expressed, is in my view bound to exert a distorting influence on its end product, the results. Furthermore I am afraid that the risk of this distorting effect is high in a field like parapsychology where we have almost no positions and where most people are badly lacking the security of employment. The situation is likely to become critical, if job security and "promotion" start to depend upon the outcome of one's experiments. I am not primarily thinking here of the probability of the occurrence of deliberate fraud, but rather of a wide range of more subtle ways of "adjusting" and "improving" one's data to meet the requirement of being statistically significant at a pre-determined, but rather arbitrary level. This may involve for instance, re-stating one's hypotheses, small "rationalized" exclusion of data, post hoc analysis until something turns up.

Nevertheless, I do not find it constructive to discuss the phenomenon of unreliable experimenters in terms of "weak" and "strong" minds respectively. I am very well aware that there are educational aspects to the problem, but by and large I believe that the most fruitful way of attacking the problem is to analyze the entire research process and the motivation behind it. To avoid misunderstanding I want to stress that I have never said that a scientist's reliability can not be affected by training in a favourable way, but one should notice that the scientist himself is only one part of a highly complex and dynamic process, the scientific research process.

I have previously made suggestions as to how to improve our understanding of the research process by pin-pointing its risky

links or synapses by the use of the "system approach" (2). The analysis I carried out has had practical implications for the publication policy which we have stated as an ideal for our new journal: the European Journal of Parapsychology. We are aware that the fool-proof experiment does not exist - although much can be done to improve the control of an experiment at different levels. In short, we shall try to avoid selective reporting and yet at the same time try to refrain from making our journal a graveyard for all those studies which did not "turn out". These objectives may be fulfilled by the editorial rule of basing our judgments entirely on our impressions of the quality of design and methodology of the planned study. The acceptance or rejection of a manuscript should take place prior to the carrying out and the evaluation of the results of the study. It also implies that the hypotheses as well as the number of subjects, number of trials, etc., should be carefully specified before the collection of data is made. Another advantage could also be that the researcher would be able to capitalize on criticism before the experiment is firmly outlined.

What I have described as "Model 3 Condition", (that is "optimized automated and intersubjective control of a psi-experiment") in the paper referred to, should relieve the experimenter from the rather dull duty of making registrations, while leaving him free to concentrate on creating a favorable "rapport" with his subjects - a psychological state which is often thought to have a beneficial influence on a subject's psi performance or the psi performance of the interacting experimenter-subject dyad. By the same token it is reasonable to assume that being relieved from the pressure of obtaining "positive" results should not be a disadvantage for producing psi phenomena.

I believe that I have rather good grounds to state that I am not the only person within our field who feels concerned about possible implications of different types of publication policies. Dr. John A. Palmer's idea of establishing a kind of "data bank" is an independently developed idea intended to tackle some of the problems which I have touched upon here. There are several good reasons for supporting Dr. Palmer's idea. Here I will mention just a few aspects.

1) If the proportion of successful experiments is small in relation to unsuccessful ones, this should be made public.

It may well have educational implications. I am under the impression that many an enthusiastic beginner who carries out an "unsuccessful" psi experiment may feel very frustrated and shy away from our field since they may think that they are "no-good" at all as experimenters. They may have got this idea from journals where only the "successful" experiments are published at full

length.

2) On the other hand if it should turn out that a high proportion of experiments were successful this too has a number of implications. How does one reconcile such a finding with the notion that there is an almost non-existent reproducibility in parapsychology? If the ratio of successful experiments is high, that would imply that we could commence with a more constructive and critical approach of theory testing, perhaps in accordance with the famous Popperian scheme:

$$P1 \rightarrow T1 \rightarrow EE \rightarrow P2$$

where "P" stands for "problem"; "T" stands for "tentative theory" and "EE" for (attempted) "error elimination".

3) If similar experiments (jointly planned and carried out with the same high degree of automated control) were carried out at different research centres and yielded strongly contrasting results (statistically significant between different centres) this would also be very important information. A systematic and penetrating analysis of possible causes for those differences, subsequently carefully tested, seems to me to be a fruitful way of enhancing our knowledge of critical factors affecting the outcome of a psi-experiment. The existence of a "data bank" could be a constructive step in making such comparisons possible.

Dr. Rhine may be right with his belief that because of enhanced knowledge and experience, the proportion of successful experiments is today higher, at his Institute and elsewhere, than it was during his early years at Duke University. This appreciation of the situation may be a correct one, but the notion is to the best of my knowledge unproven. A statistical assessment of this issue will be further complicated if a shift in publication policy has taken place since the early Duke days.

There are several other interesting statements and ideas that Dr. Rhine has put forward in his article which I would like to discuss and challenge, but for the time being I will restrict myself to the points I have already made. Readers whether they agree or disagree with the points I have made here are invited to discuss their views in this journal.

#### REFERENCES

- 1) Dunette, M.D. "Fads, Fashions, and Folderols in Psychology" *American Psychologist*, V. 21, 343-352
- 2) Martin Johnson "Models of Control and Control of Bias" *E.J.P.*, V. 1, 36-45