Journal of Parapsychology, Vol. 50, December 1986

COMMENTS ON THE "JOINT COMMUNIQUÉ"

BY JOHN PALMER

I am confident that I share the sentiments of many other followers of the "Psi Ganzfeld Controversy" in applauding Ray Hyman and Charles Honorton for the constructive approach they have adopted in their "Joint Communiqué." Their recommendations are generally sound and provide a solid foundation for future interactions between parapsychologists and outside critics.

STATUS OF THE EVIDENCE

I wish to begin my own commentary by evaluating the authors' "General Areas of Agreement" with reference to two papers I have recently published on how the psi controversy should be conceptualized (Palmer, 1986; Palmer, in press). The former article is cited twice by Hyman and Honorton in their paper.

The two authors agree on three important points with which I, too, agree:

1. There is an overall significant effect in this data base that cannot reasonably be explained by selective reporting or multiple analysis. (p. 351, abstract)
2. The present data base does not support any firm conclusion about the relationship between flaws and study outcome. (p. 353)
3. Whether the [yet to be demonstrated?] anomaly is ultimately to be considered "paranormal" will . . . depend on further developments such as the extent to which the findings can be brought under lawful control and the construction of a positive theory of the "paranormal." (p. 354)

In other words, the current data base does *not* demonstrate paranormality.

The remaining issue of this type, which is particularly relevant to my own ruminations, is whether the data base demonstrates a genuine communications anomaly. It is not clear if the authors see eye to eye on this point. In Footnote 3, they state, "The term *psi* in this paper simply denotes a communications anomaly" (p. 353). On p. 354, they say that "*if* a variety of parapsychologists and other investigators continue to obtain significant results . . . *then* the existence of a genuine communications anomaly will have been demonstrated." (My emphasis.) In other words, they seem to be *agreeing* that the current data base does *not* demonstrate such an anomaly. Yet on p. 351 (abstract) they state, "We continue to differ over the degree to which the effect constitutes evidence for psi," i.e., for a communications anomaly. A few such inconsistencies are to be expected in a joint paper written by two authors with fundamentally different orientations to the evidence, and my purpose is not to criticize them. However, such inconsistencies do pinpoint issues that may be in need of further consideration.

My own position, as developed in my published papers referred to above, requires me to conclude that the ganzfeld data base *does* demonstrate a genuine communications anomaly. If one accepts my definition of such anomalies as cases of ostensible communication that, when taken at face value, transcend the basic limiting principles of conventional science and for which no adequate explanation yet exists, the above conclusion follows logically and necessarily from the three statements quoted at the beginning of my commentary and to which the authors have already agreed. Thus, it is not clear to me on what basis one or both of them deny the ganzfeld data base the status of a genuine anomaly.

Does it fail because the possibility that the results are attributable to some combination of Hyman's flaws has not been conclusively ruled out? I happen to agree that such artifactual hypotheses cannot be conclusively rejected for this data base, but what does that have to do with the status of the data as anomalous? *Of course* an anomaly may ultimately prove to be an artifact. If we are to refuse to call a finding anomalous simply because we can concoct an artifactual explanation for it, the term would be useless because there would never be any instances of it. If only certain artifacts, such as those cited by Hyman, trigger the denial mechanism, a plausible rationale must be presented for singling them out.

I think more is involved here than a semantic quibble. The real issue is that denying the status of anomaly to the ganzfeld data base denies scientific legitimacy to the research behind it. By seeming to provide a justification for ignoring the research, such a denial of legitimacy serves to obscure the fundamental challenge that the ganzfeld data (and many other psi data) provide: there are repeatedly demonstrated statistical effects that when taken at face value are incompatible with the basic limiting principles of nature as defined by conventional science and that have, as yet, no adequate explanation. *This fundamental fact must not be lost amid the fine points of the debate.* As much as I sympathize with the desire to not dignify "suboptimal" methodology, our primary task is not to assign grade points to research reports but to interpret as objectively as possible an admittedly imperfect but potentially important body of data. The alternative explanations implied by Hyman's flaws lack plausibility, and their empirical support, even if one accepts Hyman's flaw classifications (which I do not), is tentative and circumstantial. To suggest that some other, unspecified flaws account for the results is pure speculation. I see no rational basis for denying scientific legitimacy to this research, so long as the claim made on its behalf is one of anomaly.

The practical consequences of this issue of legitimacy should not be underestimated. All else being equal, the likelihood that competent researchers will be attracted to psi research or that funding sources will support it is greatly enhanced if the demonstration of a scientifically legitimate anomaly is generally acknowledged. In other words, the likelihood of our having the resources to solve the puzzle is to some degree dependent on our willingness to acknowledge that the puzzle is real. Since the two authors "agree that further research in this area is important, not only for parapsychology, but for science generally" (p. 354), it is a shame that they could not agree to grant such acknowledgment to the ganzfeld data base.

THE VALUE OF REPLICATIONS

Another sentence that concerns me is the following: "If psi is responsible for the outcomes obtained in this data base, then the ganzfeld experiment should continue to produce successful outcomes when the various problems that Hyman pointed out are eliminated" (p. 353). Maybe, maybe not. In the absence of a good understanding of the factors that are necessary for success in the ganzfeld, it is by no means a sure bet that successful results would continue to occur even if the flaws were retained. Subsequent failures to replicate would *not* logically compel the conclusion that the original findings are attributable to Hyman's flaws. This hypothesis can only be confirmed by designs that systematically compare the results in two or more conditions, with proper blinds, that are identical except for the presence of the specific flaws being addressed.

This is not to deny the importance of simple replication attempts. Successful replications that take steps to eliminate Hyman's flaws would significantly reduce the likelihood that such flaws were responsible not only for those replications but also for the findings in the original data base. Moreover, the robustness of the ganzfeld procedure would be further confirmed, thereby providing a solid basis for process-oriented research intended to explain the findings.

My concern is that *unsuccessful* replications might be used to argue that the puzzle has been resolved in favor of an artifactual interpretation and that we can all go home. On the contrary, it is my view that under such circumstances the puzzle would remain unresolved, and the appropriate response would be additional research to try to uncover the variables responsible for the divergent results. A valid explanation of the anomaly, whether artifactual or paranormal, can come only from positive results, not from negative results.

Multiple Analyses

As I indicated at the beginning of my commentary, I found the authors' "Recommendations for Future Psi Ganzfeld Experiments" to be very valuable. However, there is one recommendation that I cannot accept without qualification. It is best represented in their text by the following sentence: "When multiple tests are planned, appropriate adjustments should be made to keep the total overall error rate within the commonly accepted region" (p. 358).

Adjustments such as the Bonferroni inequality, which the authors cite as a concrete example, are appropriate if the purpose is to make a claim of statistical significance in the context of an isolated experiment. As a general rule, they are not appropriate if the claim of statistical significance is ultimately to be based on a series of experiments.

In parapsychology, it is becoming increasingly apparent that the claims being made are predominantly of the latter type. Parapsychologists' embrace of meta-analytic techniques is a clear example of this trend. The authors themselves agree that "the outcome of a single experiment rarely, if ever, determines the acceptance or rejection of laws and theories" (p. 361).

The problem with using adjusted p values to draw conclusions in cross-experiment analyses can be illustrated by the following extreme-case example. Assume that five hypotheses, A through E, are tested in ten identical replications. Assume further that A is confirmed in each of the ten experiments at $p = .05$, but none of the other hypotheses is ever confirmed. The evidence for A is obviously strong. However, if each of the ten confirmations of A were to be corrected for the four nonsignificant analyses using the Bonferroni adjustment, they would all be classified as nonsignificant. This in turn would lead to the absurd conclusion that A is false because it failed to be confirmed ten times in a row.

This problem would not arise in strict meta-analyses that use effect sizes rather than p values as the unit of analysis, but one still is likely to find in the literature formal or informal "box score" analyses that do use p values. In fact, the Hyman-Honorton debate started out using this approach.

My own recommendation would be to use adjusted p values only when a *within*-experiment claim of significance is clearly intended. Even in these cases, it often will be difficult to determine objectively the total number of analyses to be assumed for the adjustment. I would suggest further that in such cases the unadjusted p values be cited as well. It is true that to some extent the logic of meta-analysis argues against citing p values at all for individual effects, relying strictly on effect sizes. However, I am not yet ready to endorse such a radical step. Unadjusted p values still provide a convenient guideline, or at least a partial guideline, for deciding which outcomes in a study are worthy of discussion or replication. Admittedly, such a guideline is arbitrary, but so is the .05 rejection criterion itself. However, and this perhaps is the most crucial point, unadjusted p values should always be accompanied by a statement affirming their contextual or tentative status. If researchers in the past are to be faulted, I think it should be for failing to include such disclaimers, not for using unadjusted p values.

References

PALMER, J. (1986) Progressive skepticism: A critical approach to the psi controversy. *Journal of Parapsychology,* **50,** 29–42.

PALMER, J. (in press). Have we established psi? *Journal of The American Society for Psychical Research.*

Institute for Parapsychology
Box 6847, College Station
Durham, NC 27708