Ambiguous Data Result in Ambiguous Conclusions: A Reply to Charles T. Tart

J. E. KENNEDY[1]

INTRODUCTION

In a recent paper in *the Journal* (Kennedy, 1980), I discussed the evidence for experimenter effects in an ESP learning experiment reported by Charles Tart (1975, 1976). My discussion centered around the existence of nonrandom target sequences that would match the subjects' calling habits. In a response to my paper, Tart (1980) raised numerous issues that need further comment. While Tart's response inaccurately represented my position on several points, and brought up some peripheral matters, my comments here will be limited to only the most important issues.

Tart described his learning theory as a potential breakthrough in parapsychology that has been mishandled by the parapsychological community. However, even if it were true, his theory appears to me to be primarily of academic value since it makes the paradoxical prediction that subjects with highly developed ESP abilities will show learning with immediate feedback while those who do not already possess well developed abilities will not be able to learn to use ESP—at least not with laboratory testing. The crucial questions of how highly talented subjects originally obtained their ESP abilities and how others without such talent can develop their latent abilities to a high level have not been dealt with. Parapsychologists must still *discover* rather than develop talented subjects. Thus, as noted in my paper, Tart's experimental screening procedure for finding talented subjects is, to my mind, the most important part of his work and the feature most in need of careful evaluation.

Tart construed my paper as primarily questioning the competence and honesty of the highly successful experimenter, G.T. In fact, my discussion of the types of errors that could occur with the procedure Tart used in his first learning experiment *emphasized avoiding these problems in future work*. The possibility of errors, however, must be kept in mind when assessing the results of Tart's

---

experiments and his explanation that the observed lack of XX target doublets was due to experimenter error clearly indicates the legitimacy of such questions. As noted in my paper (pp. 198-199), I do not see any way with available information to distinguish between experimenter error and experimenter PK upon the targets. However, there is a question that is more important than making this distinction.

If the highly significant results were primarily due to experimenter effects upon the targets, be the effects PK or errors, the idea that talented subjects were found, and thus, the most important aspect of Tart's results, comes into question. To my mind, this more general topic of experimenter effects was the central issue in my paper. A second point that is also very relevant is that Tart's interpretation of these data as evidence for his *ESP learning theory* becomes doubtful if the results were actually due to some kind of experimenter effect on the targets.

Three general topics in Tart's response to my paper need to be considered: (a) the uniqueness of experimenter G.T., (b) the evidence that the targets were influenced to match the subjects' responses, and (c) the interpretation of the strong displacement missing effects that were found.

UNIQUENESS OF G.T. AS AN EXPERIMENTER

I took the fact that G.T. was a unique experimenter as being self-evident. With the ten-choice machine, G.T. had five subjects in the final stage of the experiment and each of them obtained highly significant results. None of the subjects tested on ten-choice machines by other experimenters have shown convincing evidence for nonchance results in the final stage of the experiments. Throughout his writings Tart has underplayed or completely ignored the obvious experimenter effects in his data, and in his response to my paper he specifically argued that G.T. was not unique. However, the fact is that the difference between G.T.'s results and those of the other experimenters is extremely significant and will remain so even after correcting for any number of multiple comparisons one can reasonably imagine for these data. Tart's argument that G.T. was not unique did not deal with this central issue. I will briefly comment on his main points, although they are for the most part peripheral to the matter of the obvious experimenter effects in the data.

Tart took strong issue with my statement, "Other than G. T.'s subjects, no one in the final stage of either of Tart's two experiments obtained convincing evidence for psi with one of the ten-choice machines" (p. 207). Tart felt that this conclusion misrepresented the actual situation because it did not consider (a) the

significant results obtained in other experiments with the ten-choice machines, (b) the significant results obtained by several other experimenters using four-choice machines, and (c) two subjects who obtained .05 level effects with ten-choice machines in the second training study. With regard to (a), since I viewed the selection of talented subjects as the most important aspect of Tart's findings and the feature most in need of replication, I limited my remarks to the two studies which used screening procedures, i.e., the "training studies." Pooling in extraneous data from other experiments, as Tart did in his response, is inappropriate in this context since it confounds the evaluation of the training studies. For point (b), the facts that (1) the four-choice results in the first experiment could not be investigated in detail because trial-by-trial data were not recorded and (2) the subsequent publications about the first training study have dealt primarily with the ten-choice data, are the reasons my paper was limited to the ten-choice work. The more general success with the four-choice machines does not nullify the need to investigate the possibility of experimenter effects—particularly with the ten-choice data. Concerning (c), of the seven subjects tested with ten-choice machines in the second training study, one gave hitting at the .05 level and another had .05 level missing. The overall results for the seven subjects did not approach significance with the planned analyses and, as noted previously, for this reason the second study was not discussed in my paper. I do not find the two subjects selected from the overall chance results to provide *convincing evidence* for ESP.

Tart also argued that the CR or *p* value can be a misleading measure for comparing psi performance since "the same frequency of psi functioning on a ten-choice machine will yield much higher significance levels than on a four-choice machine" (p. 213). He then noted that there is considerable overlap when the psi coefficients (Timm, 1973) for G.T.'s subjects are compared with those of the significant subjects for the four-choice task. While this argument is basically irrelevant to the question of possible experimenter effects with the ten-choice machines, it needs to be discussed so that no readers will be led to accept a dubious concept.

The idea that the estimated frequency of ESP hits independent of the probability of a hit is an appropriate measure for psi effects is a very dubious assumption. Thouless (1935) and later Rhine (1951) noted that the approximately equal deviations in high- and low-aim conditions indicate that the frequency of trials with complete ESP information is different for the two conditions. While the evidence for equal deviations is not yet compelling, that effect would imply that the rate of hits is *not* independent *of P.* The role of P is a fundamental aspect of psi operation; however, this topic has not

been systematically investigated. (For a discussion of the probability of a hit factor in PK, see Kennedy, 1978.) The available evidence indicates that the information content of the trial (i.e., theP) is an important factor in the net frequency of psi hits.[2] Timm (1973) noted that the amount of *transmitted information* is closely related to the CR^2. Thus, the transmitted information, with which approximately equal deviations would be expected in high- and low-aim conditions, may be a more appropriate measure for comparison than the estimated frequency of ESP hits (i.e., the psi coefficient). With such a measure, there is no overlap at all between G.T.'s results and those of any of the other experimenters in Tart's training studies.

There can be little doubt that G.T. was unique and that some type of experimenter effect occurred. All interpretations and generalizations of these data must be done with this fact in mind. The next question is whether G.T. created an experimental environment that brought out the subjects' ESP abilities or whether the targets were influenced to match the subjects' calls.

EVIDENCE THAT THE TARGETS WERE INFLUENCED TO MATCH THE RESPONSES

The central part of my paper comprised the analyses of the target sequences for patterns that would match the subjects' response habits. Significant results were obtained for G.T.'s data on each of the three analyses. While Tart discussed various aspects of these analyses, he did not directly dispute my findings. He did question (p. 216) the extent to which it is possible to identify response habits independently of the actual target sequence. Yet later, when developing a computer model of response habits, he stated he had found two response characteristics that were common to almost all of the subjects tested to date (p. 218). These response habits were avoidance of calling the previous (-1) target and also avoidance of the second-to-last (-2) target. Since two of the analyses I reported were based on exactly these response habits, there would seem to be little doubt as to the applicability of two of the analyses. Tart suggested that a PK effect upon the targets to take advantage of the subjects' response habits is a likely, though somewhat speculative, explanation for the third analysis (distribution of targets relative to

_____

[2] That many researchers interpret the available evidence this way is indicated by the fact that in a questionnaire distributed at the 1971 convention of the Parapsychological Association, approximately 70% of the respondents agreed with the statement, "The absolute number of extrachance hits in a given number of calls increases as the probability of the target increases (in the middle range of P's)" (Schmeidler, 1971, pp. 213, 217).

the previous targets) presented in my paper. Thus, my conclusion that the results of all three analyses support the hypothesis of experimenter influences on the targets has not been brought into question.

DISPLACEMENT EFFECTS AND THEIR INTERPRETATION

While it does not directly provide evidence relevant to the question of experimenter effects, the interpretation of the displacement effects is a matter that needs further clarification. Tart stated (p. 216) that I incorrectly understood his theory of trans-temporal inhibition. I indicated that he had hypothesized the highly significant displacement missing reflected a mechanism to enhance direct hits, but it appeared to me that such displacement effects could only interfere with and be detrimental to direct hits. In response, Tart first noted that "the rationale for trans-temporal inhibition is not that percipients desire to respond to future and past targets . . ."—a point unrelated to anything in my paper—and then commented, "It [trans-temporal inhibition] does indeed interfere with responding to the present-time target, and that is the whole point of the theory. Readers interested in the theory should refer back to my original publications for clarification" (pp. 216-217).

Those readers who examine his original papers will find statements such as the following:

> What I am postulating, then, is an *active inhibition* of precognitively and postcognitively acquired information about the immediately future and immediately past targets, which serves to enhance the detectability of ESP information with respect to the desired real-time target (Tart, 1978, p. 233).

The only interpretation I can see for Tart's recent comments is that he has reversed his original position and now apparently agrees that displacement missing will interfere with rather than enhance detectability of the real-time targets.

According to Tart, I also speculated that the "lack of XX [target] doublets interacting with the response biases of the percipients would produce higher levels of hits and the displacement effects . . ." (p. 217). While I did note the obvious fact that the interaction between the lack of XX target doublets and the subjects' bias against calling the previous targets "would increase the likelihood of getting hits" (p. 203), I did not suggest that these factors would "produce" the displacement effects. Tart apparently was referring to my discussions of the correlations he reported between direct and +1 displaced hit scores (Tart, 1978). As an example of why these two scores are not independent, I pointed out that the ten-

dency to not make the same call twice in a row would lead to a negative relationship between direct and displaced hits. My original statement was, "This example is not given as necessarily explaining the strong relationships Tart found, but rather to indicate the dependence problem which invalidates the statistical significance he reported for this correlation" (p. 201). The point was that the usual procedure for calculating the significance of a correlation cannot be legitimately applied under these circumstances; thus, we do not know whether or not the correlations are significant.

In response, Tart reported the results of some computer simulations carried out to investigate possible artifacts due to target patterns interacting with response biases. The programs simulated the lack of target doublets and the response biases of avoidance of calling the -1 and -2 targets. Neither significant direct hits nor significant correlations between direct and displaced hits were found in the simulations, an outcome which Tart interpreted as indicating that the possible artifacts "have no real empirical consequences in these data" (p. 219). This conclusion, however, is unsound for several reasons.

First, Tart's conclusion is based on only 50 iterations of the program while a much larger number is normally used to obtain reliable results under such circumstances. Even accepting these simulation results as representative, we still do not have an estimate of the significance levels of the observed correlations. The simulations suggest that with these sample sizes the dependence does not lead to expected correlations that are significantly different from zero; *but, this situation does not establish that the observed values are significantly different from the expected values.* Further, even if the simulations had followed the usual Monte Carlo procedure of generating *p*-values for the observed correlations, the outcome would be unconvincing because of the inadequacy of the underlying model.

While Tart's computer program tried to correct for two types of response habits (avoidance of calling the -1 and the -2 *targets),* it apparently did not consider the tendency of the subjects to avoid calling the previous *call.* This is one of the strongest response biases and was given as an example of a response characteristic that would obviously create dependence between direct and displaced hit scores. Given the apparent failure to consider this very pertinent factor, any results from Tart's simulation programs are of questionable value.

On a more fundamental level, we must consider whether our understanding of human response habits is sufficient to allow adequate computer models to be made. The fact that a few response habits can be identified across subjects does not mean that simple

computer models based on these habits can be used to evaluate the overall effect of response biases. A similar situation arises with Tart and Dronek's (1979) Probabilistic Predictor Program (PPP). To reiterate my previous comments, the facts that the PPP does not predict the targets as accurately as the original subjects and does not produce the displacement effects may only indicate a failure to use an appropriate strategy in the program. The development of computer models of human capabilities and behavior is an interesting area but, given the current level of development, *negative results with such models simply cannot be taken as compelling evidence when interpreting ambiguous effects in psi experiments.*

In summary, the theory of trans-temporal inhibition is based on a relationship between direct and displaced hit scores. Since the two measures are not independent, the usual correlation is not an appropriate statistic for hypothesis testing. The data are not easily treated by Monte Carlo methods; thus, assigning a significance level to the observed correlation coefficients is a difficult matter that has not yet been adequately treated. Further, as noted above, the underlying concept that displacement missing reflects an ESP mechanism for enhancing direct hits, appears to be logically untenable. The occurrence of displacement missing by means of ESP would (it appears to me) tend to *lower* the direct hit scores. An influence on the selection of the targets to avoid the previous call would interact with the subjects' response habits (avoidance of calling the same symbol twice in a row) in a way that would *increase* the likelihood of direct hits and produce displacement missing. However, I see no way of determining from the data whether the displacement effects were a result of ESP by the subjects or experimenter influence on the targets.

CONCLUSIONS

Dr. Tart and I are in agreement about several basic points. *We agree that there is evidence for nonrandomness in the target sequences and that some increase in scoring is likely a result of this nonrandomness. We also agree that it is very difficult, if not impossible, to establish the magnitudes of the scoring due to influences on the targets versus scoring due to ESP.* Further, we agree that at this point new experimental work will probably be more useful than continued analyses of these data.

The primary differences between the views of Tart and myself center around the interpretation of the results given the above points. As I understand him, Tart's basic position is that since the available evidence for influences on or patterns in the target sequences can *compellingly* account for only part of the high scores,

it can be safely concluded that the results were predominantly produced by ESP. On the other hand, my position is that the uncertainty in establishing the magnitude of non-ESP effects carries through to the conclusions about the results; since there is evidence that influences on the targets did occur and since a strong influence making the targets match the calls might be difficult to detect by post-hoc, global analyses, the interpretation of these results is ambiguous and cannot be assumed to be ESP. Readers must decide for themselves which position is more tenable.

REFERENCES

KENNEDY, J. E. The role of task complexity in PK: A review. *Journal of Parapsychology,* 1978, 42, 89-122.

KENNEDY, J. E. Learning to use ESP: Do the calls match the targets or do the targets match the calls? *Journal of the American Society for Psychical Research,* 1980, 74, 191-209.

RHINE, J. B. The outlook in parapsychology. *Journal of Parapsychology,* 1951, 15, 151-63.

SCHMEIDLER, G. R. Parapsychologists' opinions about parapsychology, 1971. *Journal of Parapsychology,* 1971, 35, 208-218.

TART, C. T. *The Application of Learning Theory to ESP Performance.* New York: Parapsychology Foundation, 1975.

TART, T. C. *Learning to Use Extrasensory Perception.* Chicago: University of Chicago Press, 1976.

TART, C. T. Space, time and mind. In W. G. Roll (Ed.), *Research in Parapsychology 1977.* Metuchen, N.J.: Scarecrow Press, 1978. Pp. 197-249.

TART, C. T. Are we interested in making ESP function strongly and reliably? A response to J. E. Kennedy. *Journal of the American Society for Psychical Research,* 1980, 74, 210-222.

TART, C. T., AND DRONEK, E. Trying to profit from nonrandom-icity in ESP target sequences: Initial explorations with the probabilistic predictor program. Paper presented at the Twenty-Second Annual Convention of the Parapsychological Association, Moraga, California, August, 1979.

THOULESS, R. H. Dr. Rhine's recent experiments on telepathy and clairvoyance and a reconsideration of J. E. Coover's conclusions on telepathy. *Proceedings of the Society for Psychical Research,* 1935, 43, 24-37.

TIMM, U. The measurement of psi. *Journal of the American Society for Psychical Research,* 1973, 67, 282-294.

Institute for Parapsychology
College Station
Durham, North Carolina 27708

Other Methodology Articles