Learning to Use ESP: Do the Calls Match the Targets or Do the Targets Match the Calls?

J. E. Kennedy[1]

(Original publication and copyright: *Journal of the American Society for Psychical Research*, 1980, Volume74, pp. 191-209)

ABSTRACT: An experiment reported by Charles T. Tart as providing evidence that subjects can learn to use ESP is examined for the possibility that the highly significant scoring rates could be due to factors other than ESP by the subjects. Since the procedure involved a screening of experimenters as well as subjects and the highly significant results were obtained by only one experimenter, it is likely that the screening procedure actually served to select a successful experimenter. Analyses were carried out to investigate the possibility that the results could be due, at least in part, to an influence upon the targets such that they matched the subjects' calls. Significant results were found in support of this hypothesis. Determining the mechanism that was dominant in the production of the highly significant overall results for this experimenter will probably be difficult, but hypotheses other than ESP by the subjects must be considered. Several weaknesses in the design of the experiment are discussed so that they can be avoided in follow-up studies.

INTRODUCTION

The recent theoretical and experimental work by Dr. Charles T. Tart (1975a, 1976a, 1977b, 1978a) has aroused considerable interest among parapsychologists. Tart's basic thesis is that subjects in ESP experiments cannot identify the appropriate internal cues and thus learn to use ESP unless immediate feedback of the target is given. In terms of learning theory, feedback is needed for learning to take place and the absence of immediate feedback characterizes an "extinction paradigm" which typically leads to a loss of ability or decline in performance. Tart hypothesizes that the declines in scoring frequently found in card-type ESP experiments are a form of "extinction," due largely to a lack of immediate feedback.

In developing this line of thought, Tart (1975a, 1976a, 1977b, 1978a) has suggested that even with immediate feedback the inability to distinguish between hits due to chance and hits due to ESP severely hinders the learning process and allows learning to take place only if ESP occurs on a relatively large number of trials. Based on this theoretical framework, Tart expects a relationship between a subject's initial ESP ability and the amount of ESP

[1] I wish to thank Rex Stanford for many helpful comments during the preparation of this paper.

learning that takes place. He has characterized the learning situation for three different levels of initial ESP ability: (a) Those subjects with no or very little ESP ability should give no evidence for ESP and neither inclines nor declines should appear in their data. (b) Moderately talented subjects should show extinction curves (i.e., declines) if immediate feedback is not given and ESP performance should be stable (i.e., neither declines nor inclines) if feedback is given, (c) Highly talented subjects, those whose abilities are above an as-yet-unspecified threshold, should show learning or an increase in ESP scoring if immediate feedback is given.

In his review of the experimental literature, Tart (1975a, 1975b, 1976a) found evidence that he interpreted as supporting his theoretical views. He also carried out two studies specifically designed to investigate his learning hypotheses and the results were presented as supporting the hypotheses (Tart, 1975a, 1976a, 1979a). In both the literature survey and his experimental work the primary evidence for the learning theory was an absence of declines in the data of subjects whom Tart considers as moderately talented. There was little evidence of inclines in ESP performance either in the previous experiments or in Tart's own studies, even though some of the highest scoring rates in the recent history of parapsychology were obtained in the first of his studies.

Tart's work is important and deserves careful consideration for several reasons. The severe need for the more reliable psi results that could be produced by a valid application of learning theory is the most obvious one. The fact that, as noted above, some of the most highly significant results in recent years were obtained in Tart's first study is an important event in its own right. The work also requires serious consideration because of the widespread attention it has received as a result of its being extensively published in the parapsychological literature and presented on a popular level as a "research breakthrough" (Tart and Neubert, 1976, p. 12). The experimental design is intriguing in that it employed a screening procedure to select talented subjects, it allowed large-scale testing by students, and it implicitly selected talented experimenters. The testing equipment was described in detail specifically so that others could replicate the equipment. Given all the above factors, we can no doubt expect that others will employ the same design, which adds further to the importance and potential impact of the original work.

An extensive array of theoretical and methodological criticisms of Tart's work has been published by Stanford (1977a, 1977b; reply by Tart, 1977c). Similar though less extensive criticisms were also raised by O'Brien (1976; reply by Tart, 1976b).

My interest in Tart's work stemmed from the potential impor-

tance of the work and from numerous reservations I had about the learning interpretation. My own impression of the previous ex perimental literature had been that any effects of learning were negligible over the time periods involved in laboratory testing,[2] a view also widely held by others (e.g., Rhine and Pratt, 1957, p. 89). In the same vein, like those who reviewed Tart's original mono graph (O'Brien, 1976; Stanford, 1977a), I did not think that the absence of a decline could be construed as evidence for learning.

The hypothesis that ESP learning could occur only for those subjects whose initial ability was above a particular threshold and the idea that declines in psi scoring are a form of extinction as conceptualized in learning theory also seemed (and still seem) very questionable to me. One normally thinks of the extinction of a *learned* response; but Tart's theory, as yet, has not dealt with when or how or even if the initial learning of ESP ability takes place. In the present form of the theory, people must be able to use psi before they can learn to use psi. The only comment on the initial stages of learning is to predict that, at least with laboratory testing, the initial learning will never take place (i.e., subjects with low or moderate psi performances should *not* improve).[3] Tart's applications of learning theory seem to provide an elaborate ex planation for why learning theory basically does not apply to psi experiments and does not deal with the all-important question of how to initially develop or demonstrate psi ability.

Although the evidence for learning strikes me as unconvincing, Tart's work does bring up several other interesting and potentially very useful findings. The suggestion that declines in ESP scoring are less frequent when immediate feedback is given merits further investigation. Perhaps the most valuable aspect of Tart's experimental work is the fact that a large-scale screening procedure was

[2] Various hypotheses can be proposed for why immediate feedback might in fact be detrimental to psi. For example, there is evidence (see Palmer, 1978) that the use of logical response habits or strategies is detrimental to ESP operation. Response habits arise primarily because the subject's calls are influenced by the calls on previous trials. With immediate feedback the calls can also be expected to be influenced by previous targets, thus further introducing response habits and inhibiting ESP. Some subjects may also find feedback for the typical large number of incorrect calls to be frustrating, thus hurting the delicate psychological conditions of the experiment. This frustration would be expected to increase as the probability of getting a hit is reduced.

[3] Tart (1977b, p. 401) has stated that "it is *not* the case that percipients with a low talent level can never hope to do more than stabilize their performance," and notes that factors such as motivation and general learning ability can *probably* compensate for a lower psi talent. However, this statement appears to contradict the main thesis and predictions of his theory, and the fundamental question of how highly (and moderately) talented subjects originally obtain their ability before entering training procedures remains unanswered.

used to select very successful subjects. Follow-up work is clearly called for.

The potential importance of these findings prompted me to spend some time considering the possibility that the results could be due to factors other than the subjects' ESP. In particular, the fact that the highly significant results were obtained by only one experimenter and the indications (Gatlin, 1979) that the targets may not have been random stimulated an investigation into the possibility that some kind of experimenter effect may have entered into the results. The procedural comments and questions which arose during my examination of this possibility and the results of my statistical analyses upon the data will be presented here since they may be of interest to those evaluating this work and particu larly to those planning to replicate and extend it.

Tart's basic experimental design consisted of selecting talented ESP subjects through a two-stage screening process and then testing the selected subjects using a procedure that, according to his theory, might allow them to learn to use ESP. A class of undergraduate students acted as experimenters and carried out large-scale testing of other students. The experimenters recruited their own subjects and each subject worked with his or her experimenter for the duration of the testing. The first time the experiment was carried out (Tart, 1975a, 1976a) two testing machines were used; however, trial-by-trial data were recorded for only one, the ten-choice trainer (TCT). The most significant results were obtained with the TCT and my interest has focused on these data. With the TCT, 10 subjects successfully passed through the screening process into the final stage of the experiment; five were tested by one experimenter (G.T.) and five other experimenters each tested one subject. As will be discussed below, the significant ESP results were obtained only by the subjects tested by G.T.

An improved version of the TCT, called the ADEPT, was used by a different set of students (experimenters and subjects) in a complete replication of the three-stage experiment (Tart, Palmer, and Redington, 1979a). Since the overall ESP results with the ADEPT were not significant, my inquiry has been limited to the work with the TCT.

The TCT included a subject's console with 10 response buttons, corresponding lights for feedback, a rarely used pass button (for trials on which the subject did not feel like making a response), and a "ready" light that signified when a trial began. In another room was located the experimenter/agent's console consisting of 10 switches and lights corresponding to the subject's response but tons. The switches were used by the experimenter to designate the target for each trial. In the study which obtained the highly signifi-

cant results, the experimenter used a 10-outcome electronic random number generator (RNG) to select the target and then entered it into the TCT by manually setting the proper switch. Once the target switch was set, the ready light on the subject's console came on. The experimenter's console also included counters that automatically recorded the number of hits and trials based on the targets entered by the experimenter. The target and response for each trial were written down by the experimenter. A closed circuit TV allowed the experimenter to continuously observe the subject's console "so that the experimenter could watch the subject's hand motions and tell if he was getting closer to or further from the target, when he hesitated, etc." (Tart, 1975a, p. 57). The experimenters became quite involved in trying to telepathically direct the subject's "scanning" hand motions.

COMMENTS ABOUT THE PROCEDURE

The proper evaluation of the evidential value of any psi experiment requires that all procedural ambiguities be cleared up and that any weakness in the design be made explicit. The discussion of the "security" of an experiment must deal with the possibility of sensory cues and unintentional or deliberate errors on the part of either the subject or the experimenter. The question of deliberate manipulations of results has often been of primary concern since procedures that curtail this possibility usually also preclude unintentional mistakes. Traditionally in parapsychology, the integrity of the subject is not assumed and experiments are designed so that the possibility of cheating by the subjects is actually or virtually eliminated. The assumptions concerning the experimenter's integrity have been a more difficult topic. Many researchers today (e.g., Palmer, 1978; Tart, 1977a) believe it is impossible to design an experiment such that the possibility of experimenter fraud is completely eliminated, a position with which I concur. However, some minimal precautions against intentional and unintentional experimenter errors are required and the evaluation of the adequacy of precautions in a particular situation can be a controversial and subjective matter. The "two experimenter, double-blind" design (Rhine, 1974; Rhine and Pratt, 1957), which in principle prevents any *one* person—including an experimenter—from improperly influencing the results, has played a prominent role in the development of parapsychology. Recently, more often than not, strict two-experimenter, double-blind designs have not been employed and, be it good or bad, in practice the guiding principle seems to be that an experimenter is assumed to be honest unless there is specific evidence to the contrary. In the Tart design, however, the

testing was done by undergraduate students as part of their course work and apparently the actual generating of targets and data recording were unobserved. The severe problem of errors, both intentional and unintentional, that can occur with student and/or hired-hand experimenters is well known (see, e.g., Barber, 1976, pp. 57-63; Rosenthal, 1976, pp. 31-32). Since Tart's experimental design involved a screening of unselected undergraduate students for experimenters as well as for subjects and since the only criterion for selection was successful results, the possibility that with future widespread application of this procedure incompetent or unscrupulous experimenters might be selected is a particularly noteworthy concern. The extent to which intentional or unintentional experimenter errors are possible is, more than usual, an inherent weakness with this design.

In the original report of the work, the security measures taken with the TCT were discussed under the heading "Fraud-Proofing" (Tart, 1975a, p. 119). The equipment contained precautions so that the subjects could make only one response per trial. Tart also noted that "a deliberately cheating experimenter, or one with unconscious response patterns, could transmit cues that the subject could pick up. E.g., if the experimenter always hesitated longer between trials when the cards called for Target #1 to be chosen, the subject could learn that this long delay was associated with that target. . . . We found no empirical support for this problem in our main study" (p. 119). While one could question whether the specific analyses that were carried out (reported on p. 54) adequately evaluated this possibility, it is clear that Tart is aware of the problems of error and bias on the part of the student experimenters. It would seem, however, that the design of the TCT also permitted several much less subtle avenues for experimenter error.

The automated hits and trials counters play a key role when considering the possibility of recording or other errors. Yet, the published reports have not mentioned the use of standardized procedures for recording the counters or for verifying the hand-recorded data with the counters. Variations in procedure might also increase the possibility of confusion and errors in the use of the counters. For example, on the last trial of each run G.T. set the equipment so that all feedback lights on the subject's console came on except the actual target, apparently by setting all switches except the true target switch (Tart, 1975a, p. 85). A look at the circuit diagram shows that any number of target switches can be set and the feedback lights will come on corresponding to all the set switches when the subject makes his response. Setting nine switches will also result in a 90% chance that a hit will be recorded

on the counters, so the hand-recorded scores in this case should not exactly match the counters.

Assuming that after correcting for any procedural variations the counters did properly coincide with the hand-recorded scores, there appears to be no way to verify that the experimenter actually used the original targets generated by the RNG. The only record of the output of the RNG is the experimenter's hand-recorded data. A dishonest experimenter could arbitrarily enter targets that are expected to match the subject's response habits, and various types of less overt errors in the course of generating and transferring the targets can also be imagined. In fact, as will be discussed below, Tart has proposed that an unintentional error in generating targets did frequently occur in his experiment and is responsible for some of the nonrandomness in the target sequences. If high scoring rates are obtained through the kinds of errors discussed above, all or most of the targets would have to be modified and the target sequence will then be conspicuously nonrandom.

With the type of equipment Tart used, one could propose methods by which an experimenter could manipulate most targets and still record fairly random target sequences; however, such a set-up might allow the production of high scores by accurately manipulating a smaller number of targets. Besides being able to set any number of target switches, the experimenter can change the target registered on his console any time before the subject responds. Most subjects apparently make "scanning" hand motions and, since an experimenter is observing the subject's hand movements, it would seem possible that on some trials accurate manipulation of hits by changing targets would be a trivial matter. It is not clear to what degree a manipulation of this type would influence the overall randomness of the target sequence, although presumably there would be some residual effect.

In summary, following the published information, the type of equipment and procedure used in Tart's experiment allow several possibilities for unintentional or intentional error on the part of those who do the testing. In any experiment where students are employed to do the testing and where automated equipment is the only control for such errors, these possibilities constitute an essential weakness in the design. The facts that (a) the second generation ten-choice testing machine used in the replication experiment (Tart, Palmer, and Redington, 1979a) had the RNG built into the sender's console, eliminating the need for the experimenter to transfer the target, and (b) automated data-recording equipment was also added for the replication study, indicates that Tart is aware of the weaknesses in the design of the TCT. However, it seems appropriate to explicitly discuss these problems with the

TCT since it played a key role in the significant results that have been reported by Tart and has been described in detail specifically so that others can duplicate and use the equipment. The problems discussed above should be taken into account in further attempts to replicate the study and, of course, must also be kept in mind in interpreting the original results.

RESULTS OF ANALYSES

The fact that all the significant results with the TCT were obtained by one experimenter (G.T.) clearly suggests that some kind of experimenter effect occurred. Besides questions relating to the possibility of experimenter error, the hypothesis of a PK influence by the experimenter must also be considered with this design. I will leave aside untestable speculations about direct PK influences upon the subjects and deal only with identifiable effects. Specifically, the RNG could have been influenced by PK to generate targets that matched the subjects' calls.

Since subjects' ESP calls are characteristically very nonrandom, creating high scores by influencing the targets should lead to nonrandom patterns in the target sequences. The extent to which PK effects could be identified would depend on the number of targets that were influenced. Precise PK effects, in which the targets were influenced to match the next call (a form of precognitive PK or retroactive PK, depending on the point of view), could have achieved high scores by influencing a relatively small number of trials. Less accurate effects (i.e., more general biases of the targets to match call biases) would be more readily identified. The PK hypothesis would be supported by verifying that there are nonrandom patterns in the target sequences *and* that the patterns interact with the subjects' response habits in ways that produce hits. The PK hypothesis would be falsified (or at least deemed untestable) if one of the following situations occurred: (a) the targets were found to be acceptably random, (b) significant biases or patterns were found in the target sequence but shown to be due to non-PK sources, or (c) any significant biases or patterns were such that they would not interact with the calls in a way that would produce hits. Gatlin (1979) has reported significant nonrandom effects in the original target sequences; however, analyses to see how the patterns interact with the calls have not been reported.

Various relevant analyses are presented below. While these analyses may also be pertinent to the question of experimenter errors, it does not seem possible with available information to conclusively distinguish between a nonparanormal versus a

paranormal experimenter effect upon the targets. The concern now is whether such effects exist, not how they could have arisen.

The figures given here are based on a version of the data obtained from Dr. Gatlin and are slightly different from the original results (Tart, 1975a, 1976a). On runs in which the subject made passes, extra trials were collected to complete the run of 25 usable trials (Tart, Palmer, and Redington, 1979b). Tart's analyses discarded trials that were passed while Gatlin used the first 25 targets for each run and generated random numbers to fill in the missing calls. Since I am primarily concerned about the target sequence, as was Gatlin, it seems appropriate to use the targets as they came up rather than to discard those that were passed. Also, in a few runs the experimenter collected fewer than 25 usable trials and Gatlin generated random numbers for both calls and targets to fill out these runs. The results with Gatlin's data are sufficiently close to the original figures reported by Tart that the discrepancies would seem to be negligible. For the results reported here, all analyses of the doublet or triplet level do not overlap runs; i.e., only 24 doublets and 23 triplets were used per run of 25 trials.

The basic results are summarized in Table 1. I will comment briefly on the original ESP findings and present the results of the new analyses.

Direct Hits

The highly significant direct-hit results were due entirely to G.T.'s data. The difference between the direct-hit scores for the five subjects tested by G.T. and the five subjects tested by other experimenters gives $x^2 = 67.9$, 1 *df; p* < 10~[12].

Displacement Effects

Highly significant +1 displacement missing effects were also found in G.T.'s data. This means that a call tended to avoid the target that came up on the next trial. The displacement effects were not found in the data of the other experimenters and the difference between G.T.'s results and those of the other experimenters gives $X^2 = 66.9$, 1 *df; p* < l<r[12].

Tart (1977a, 1978b) has theorized that the displacement missing is an ESP effect, a form of inhibition to enhance direct hits, and presented four statistical analyses as supporting this hypothesis. These analyses need to be discussed in some detail.

First, he reported a significant negative correlation between direct-hit scores and +1 displacement scores (using the subject as

Table 1

SUMMARY OF RESULTS (MCE = 10.00%)

Experimenter 1 = G.T.	Subject	Direct Hits	CR*	+1 Displ.	CR*	Call Doublets	CR*	Target Doublets	CR*
1	1	78/500	4.17	26/480	−3.35	25/480	−3.50	17/480	−4.72
1	2	79/500	4.32	12/480	−5.48	28/480	−3.04	12/480	−5.48
1	3	124/500	11.03	11/480	−5.63	18/480	−4.56	7/480	−6.24
1	4	80/500	4.47	23/480	−3.80	16/480	−4.87	15/480	−5.02
1	5	99/500	7.30	16/480	−4.87	14/480	−5.17	15/480	−5.02
Totals for G.T.		460/2500	14.00	88/2400	−10.34	101/2400	−9.45	66/2400	−11.84
		18.40%		3.67%		4.21%		2.75%	
2	7	54/500	ns	40/480	ns	27/480	−3.20	25/480	−3.50
4	11	39/500	−1.64	48/480	ns	12/480	−5.48	31/480	−2.59
5	14	57/500	ns	60/480	1.83	17/480	−4.72	14/480	−5.17
13	17	59/500	ns	38/480	ns	20/480	−4.26	16/480	−4.87
11	32	46/500	ns	44/480	ns	17/480	−4.72	44/480	ns
Totals for other experimenters		255/2500	ns	230/2400	ns	93/2400	−10.00	130/2400	−7.48
		10.20%		9.58%		3.87%		5.42%	

* The CRs were calculated without using the continuity correction.

the unit of analysis) (Tart, 1977a, 1978b). However, direct- and displaced-hit scores are not independent. Subjects have strong response habits and the nonindependent calls are scored twice against the same targets. For example, since subjects seldom make the same call twice in a row, the presence of a direct hit means that the previous call is likely to be incorrect on +1 displacement. This, of course, produces a negative relationship between direct and displaced scores. 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.[4]

The next two analyses involved correlating a measure that Tart called "strategy boundness" with +1 displacement scores and with direct-hit scores. A significant positive relationship was found on the former correlation and a significant negative relationship on the latter. Tart (1978b, p. 238) subsequently reported that, although three mathematicians found nothing wrong with these correlations, a computer simulation found artifactually significant results. As Stanford (1978, p. 216) has pointed out, correlating strategy boundness with +1 displacement scores is essentially correlating the quantity (R_1 - R_2) with R_1 where R_1 and R_2 are random variables. (Stanford actually looked at the case of correlating R_1 - R_2 with R_2; however, the strategy boundness measure reversed the sign of the quantity R_1 — R_2.) In a derivation that took only a few minutes, Dr. J. A. Greenwood showed me that the expected value for this correlation coefficient is $1/\sqrt{2} = .71$, assuming that R_1 and R_2 are independent (a doubtful assumption in the present case) and have equal variances. Since in the Tart data direct-hit scores are highly correlated with +1 displacement scores, the relationship between direct-hit scores and strategy boundness is to be expected as a result of the artifactual relationship between strategy boundness and +1 displacement scores.

The fourth analysis (Tart, 1978b) involved calculating for each subject the direct-hit (i.e., 0 displacement) CR, and CRs for all of the +1 through +24 displacement scores. A "contrast measure" was obtained for each subject by finding the absolute value of the difference between the largest CR and the smallest CR in the 0 through 4 displacement cells. This measure was compared with a control which was obtained by finding the absolute value of the difference between the largest and smallest CRs in four cells ran-

[4] Those who wish to look into the long history of investigations of displacement effects and become acquainted with the many pitfalls and intricacies of this topic can begin with Pratt, Martin, and Stribic (1974) and then trace backwards through the references to some of the earliest work (e.g., Pratt and Foster, 1950).

domly selected from the +5 to +24 displacement CRs. The contrast measures were compared to the control measures with a t-test and significant results were reported. However, it is quite possible that artifacts played a role here too. The individual control cells were apparently randomly selected and thus in general were not sequential cells. Since the dependence between displacement scores is due to calling habits (assuming random targets), the sequential cells would be expected to show stronger dependence than the more widely separated, randomly-selected cells. The significant results of this analysis may only reflect the different degrees of dependence between the CRs used in the two conditions. It should also be kept in mind that significant results would be expected in this analysis if there are significant direct-hit scores and only chance scores on all displacements.

In discussing these analyses, I have assumed the targets were random. The presence of biases in the target sequences will further enhance the problems with all four analyses.

Furthermore, it seems to me that the rationale for the "inhibition" hypothesis is questionable since, if the targets are independent, any information (whether positive or negative) about past or future targets would be of no value in discerning the current target. In fact, it would seem that responding to other targets could only interfere with ESP performance. Avoiding the next target would restrict the response and be detrimental in the same way that letting a response be restricted by previous calls is detrimental. This is not to say, however, that Tart's displacement results could not be an ESP effect.

In terms of experimenter effects, the +1 displacement missing could indicate that the targets were influenced to avoid the previous calls. Since, as is expected, the subjects significantly avoided making the same call twice in a row, this type of effect would increase the likelihood of getting direct hits. I don't see any way to establish whether the displacement effects were a result of the subjects' ESP or whether they reflect experimenter effects on the targets.

Lack of Target Doublets

A very significant lack of target doublets[5] occurred for all but one subject (see Table 1). This effect was significantly stronger for G.T.'s data than for those of the other experimenters ($\chi^2 = 21.1$, 1

[5] The term "doublet" is used here to refer specifically to the occurrence twice in a row of the same symbol in the target (or in the response) sequence; this type of doublet is usually referred to by Tart as an "XX doublet."

df; p < 10^{-4}). Since, as would also be expected, the subjects tended to avoid calling the previous target (-1 displacement scoring rates of 2.88% for G.T. and 3.58% for the other experimenters), the lack of target doublets would increase the likelihood of getting hits. In accounting for the lack of doublets, Tart (1978b) has proposed an explanation involving the triggering push button on the RNG:

> This push button was not of the type that made a tactically [sic] discernible click when it was depressed, but simply one that got harder to push as you pushed it further in. . . . What apparently happened is that an experimenter would sometimes push and release the button to get the next target, look at the RNG and see that the same number was still in the readout, and so assume that he had not pushed the button in sufficiently to activate the generator. So he would push it again to get a new target" (p. 220).

While this speculation sounds quite plausible, it apparently has not been verified by asking the student experimenters whether they did in fact push the button twice on some trials.

Lack of Matches with the Second Sequential Target

Granted the problem with the RNG button, target T_i should not be the same as T_{i-1}, but Tt should not avoid T_{i-2}. In fact, there should be slightly more than 10.00% matches between T_i and T_{i-2} because the lack of doublets implies that both T_i and T_{i-2} tended to be different from T_{i-1}. With the lack of doublets, the rate of matches between T_i and T_{i-2} should be 10.58% for G.T.'s targets. The actual rate was 8.52%, CR = -3.18, $p < .002$, two-tailed. Since the subjects tended to avoid calling the next to the last target (-2 displacement scoring rates of 5.78% for G.T. and 6.61% for the other experimenters), this effect would also tend to increase the number of hits. For the other experimenters the expected rate of matches between T_i and T_{i-2} is 10.23% and the observed rate is 10.09%. The difference between the rates of such matches for G.T. and the other experimenters is suggestive with $\chi^2 = 3.156$, 1 *df;p* < .08.

Distribution of the Targets Relative to the Previous Target

If the problem with the RNG button were the only source of nonrandomness in the targets, one would expect the targets to be uniformly distributed over the nine other console positions relative to the previous targets. As shown in Table 2, the targets for G.T. are significantly different *(p* < .005) from the expected distribution and have a particular tendency to avoid positions adjacent to the previous target. As would be expected, the calls show a similar

Table 2

NUMBER OF TARGETS AND CALLS IN EACH OF THE NINE COUNTER-CLOCKWISE POSITIONS RELATIVE TO PREVIOUS TARGETS

		1	2	3	4	5	6	7	8	9	
Targets relative to previous targets	G.T.	233	228	263	256	257	285	291	296	225	$\chi^2 = 22.7$, 8 *df*, $p < .005$
	Other Exptrs	248	248	236	293	251	252	240	261	241	$\chi^2 = 9.2$, 8 *df*, p = ns
Calls relative to previous targets	G.T.	209	257	264	274	257	285	275	281	229	
	Other Exptrs	239	275	281	248	274	248	258	270	221	

tendency to avoid positions adjacent to the previous target (see Table 2). The targets for the other experimenters do not show significant clustering and the difference between the target distributions for G.T. compared to those of the other experimenters is again suggestive, $\chi^2 = 14.12$, 8 *df; p* < .08.

Computer Predictor Programs

Gatlin (1979; also see Tart, 1978b) has raised the question whether the subjects could have used the nonrandomness in the target sequences to artifactually obtain hits. One way to investigate this question is to have computer programs predict targets on the basis of patterns in the previous data. Tart (1978b) reported that his "optimal estimator" program can "only get about 30% as many hits above mean chance expectation as the actual percipients achieved" (p. 223). For G.T.'s data this means a scoring rate of about 12.0% with a CR greater than 3. This result is in line with my own experience that various less sophisticated computer strategies utilizing doublet or triplet frequencies of previous targets and/or calls produce about 11.5% to 12.5% scoring rates for G.T.'s targets. These scoring rates are quite significantly below the observed 18.40%, but the fact that they are statistically significant further verifies the nonrandomness of the target sequences. My programs produced only chance results for targets from the other experimenters, which probably indicates that more than a lack of target doublets is involved since the other experimenters also had too few target doublets. The calls for both groups can be predicted by my programs at rates of about 13.00% to 14.00%.[6]

CONCLUSIONS

Various factors which may have entered into the experimental outcome include: (a) ESP by the subjects, (b) the subjects' use of strategies involving patterns in the targets, (c) PK by the subjects, (d) PK by the experimenter, and (e) errors on the part of the experimenter. The presence of nonrandom patterns in the target sequences raises the possibility (likelihood) that factors other than

[6] For two of G.T.'s subjects (1,1 and 1,4), I attempted personally to predict the targets based on my knowledge of the last few original targets and calls (I had no computer help other than counting my hits and displaying the original target and call after I made each guess). I matched the original targets at rates of 13.2% and 15.4%. While these results are basically uninterpretable since I *conceivably* could have used ESP, I doubt that my ESP played a significant role and think it more likely that some people are more accurate predictors than the current computer programs.

ESP by the subjects entered into the results. The computer predictor programs indicate that the patterns in the targets for G.T. can be utilized to obtain very significant results. However, the programs have not been able to achieve results nearly as significant as the original calls and it is not clear to what extent this reflects a failure to employ the right strategies in the programs. Tart (1978b) has provided evidence that his computer program responses have characteristics very different from the original calls. This result could indicate that the subjects did not use inference strategies, or alternatively it could indicate that the strategy in Tart's program is way off base.

Evidence for the hypothesis that there was an influence on the targets by either the subjects or the experimenter requires that nonrandom features be found in the target sequence *and* that these patterns interact with the calls in a way that will produce hits. The mere presence of unusual interactions between the targets and calls, as occurred with the +1 displacement effect, is ambiguous in this regard since it could result either from an influence upon the targets or from a misdirection of ESP. Three analyses (number of target doublets, matching T_i and T_{i-2}, and distribution relative to the previous targets) were carried out specifically because they involved target patterns that would be expected to occur across subjects if the targets were influenced.[7] All three analyses produced significant results in line with the hypothesis that the targets matched the calls.

Although firm evidence has not been presented, the lack of target doublets probably was largely due to experimenter error in operating the RNG. The problem occurred for other experimenters as well as G.T., but was significantly worse for G.T. Whatever the sources, the lack of doublets would be expected to increase the number of hits. The other two patterns (T_i avoiding T_{i-2} and distribution relative to the previous target) cannot be explained by the hypothesized problem with the RNG button and are in line with the hypothesis that the targets were influenced to match the calls. For both analyses, the target patterns for G.T.'s data were significantly different from chance and suggestively different from the targets of the other experimenters. Other than the lack of doublets,

[7] The criticism that these results may have been selected from a large number of analyses does not apply since, to date, the only analyses I have done involving just the targets are (in chronological order): (a) number of doublets, (b) distribution relative to previous target, (c) matching T_i and T_{i-2} and (d) computer programs to predict targets on the basis of previous targets.

I found no evidence for nonrandomness in the target sequences of the other experimenters.

While it is not possible to prove that the influences on the targets were due to the experimenter rather than a PK effect by the subjects, an experimenter effect seems more plausible under the circumstances. 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.

Although there is considerable evidence in line with the hypothesis that in G.T.'s data some targets were influenced to match the calls, it is not clear how much of the overall "ESP" effects can be attributed to this factor (or to the subjects' use of inference strategies about the target sequence patterns). Providing convincing evidence that any of the possible interpretations was the dominant factor will likely be quite difficult. Tart (1978b, p. 219) has commented that he believes there was some PK influence in this experiment, but he believes ESP effects were predominant. It would seem, however, that unless the ambiguities in interpreting these data are resolved, hypotheses involving ESP on the part of the subjects should be held in abeyance. At this point, I am not optimistic that further analyses of these data can resolve the uncertainties; it is likely that evidence for the hypotheses that successful ESP subjects can be selected with this kind of screening procedure and that ESP declines do not occur when immediate feedback is given will have to come from further experimental work under more carefully controlled conditions rather than from continued analysis of these data.

REFERENCES

BARBER, T. X. *Pitfalls in Human Research: Ten Pivotal Points.* New York: Pergamon Press, 1976.

GATLIN, L. L. A new measure of bias in finite sequences with applications to ESP data. *Journal of the American Society for Psychical Research,* 1979, 73, 29-43.

O'BRIEN, D. P. Review *of The Application of Learning Theory to ESP Performance,* by Charles T. Tart. *Journal of Parapsychology,* 1976, 40, 76-81.

PALMER, J. Extrasensory perception: Research findings. In S. Krippner (Ed.), *Advances in Parapsychological Research. 2. Extrasensory Perception.* New York: Plenum Press, 1978. Pp. 59-243.

PRATT, J. G., AND FOSTER, E. B. Displacement in ESP card tests in relation to hits and misses. *Journal of Parapsychology,* 1950, 14, 37-52.

PRATT, J. G., MARTIN, D. R. AND STRIBIC, F. P. Computer studies of the ESP process in card guessing. III. Displacement effects in the C.J. records from the Colorado series. *Journal of the American Society for Psychical Research,* 1974, 68, 357-384.

RHINE, J. B. Security versus deception in parapsychology. *Journal of Parapsychology,* 1974, 38, 99-121.

RHINE, J. B., AND PRATT, J. G. *Parapsychology: Frontier Science of the Mind.* Springfield, 111.: Thomas, 1957.

ROSENTHAL, R. *Experimenter Effects in Behavioral Research.* New York: Irvington Publishers, 1976.

STANFORD, R. G. The application of learning theory to ESP performance: A review of Dr. C. T. Tart's monograph. *Journal of the American Society for Psychical Research,* 1977, 71, 55-80. (a)

STANFORD, R. G. The question is: Good experimentation or not? A reply to Dr. C. T. Tart. *Journal of the American Society for Psychical Research,* 1977, 71, 191-200. (b).

STANFORD, R. G. Review of *Psi: Scientific Studies of the Psychic Realm,* by Charles T. Tart. *Journal of Parapsychology,* 1978, 42, 210-219.

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

TART, C. T. Studies of learning theory application, 1964-1974. *Parapsychology Review,* 1975, 6, 21-28. (b)

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

TART, C. T. Reply to O'Brien. *Journal of Parapsychology,* 1976, 40, 240-246. (b)

TART, C. T. *Psi: Scientific Studies of the Psychic Realm.* New York: Dutton, 1977. (a)

TART, C. T. Toward conscious control of psi through immediate feedback training: Some considerations of internal processes. *Journal of the American Society for Psychical Research,* 1977, 71, 375-407. (b)

TART, C. T. Toward humanistic experimentation in parapsychology: A reply to Dr. Stanford's review. *Journal of the American Society for Psychical Research,* 1977, 71, 81-101. (c)

TART, C. T. Consideration of internal processes in using immediate feedback to teach ESP ability. In W. G. Roll (Ed.), *Research in Parapsychology 1977.* Metuchen, N.J.: Scarecrow Press, 1978. Pp. 91-122. (a).

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. (b)

TART, C. T., AND NEUBERT, R. ESP training. *Psychic,* April 1976, 7, 12-15.

TART, C. T., PALMER, J., AND REDINGTON, D. J. Effects of immediate feedback on ESP performance: A second study. *Journal of the American Society for Psychical Research,* 1979,73, 151-165. (a)

TART, C. T., PALMER, J., AND REDINGTON, D. J. Effects of immediate feedback on ESP performance over short time periods. *Journal of the American Society for Psychical Research,* 1979, 73, 291-301 (b)

Institute for Parapsychology
College Station
Durham, North Carolina 27708

Return to: Paranormal Phenomena Articles