Presented at PA (1978)

SLIDE ORDER

S1
S2
S3
S4
~~S5~~
S6
S7
S8
T-6
T-4
T-5

# Randomicity, Predictability, and Mathematical Inference Strategies in ESP Feedback Training Situations

Charles T. Tart

Psychology Department

University of California

Davis, CA 95616

I have long believed that the most pressing problem in parapsychology is how to get strong and reliable manifestations of psi in the lab, so we can profitably get on to the important questions of what psi is and how it works. As a result, some years ago I rather innocently set out to determine if providing immediate feedback of results to percipients would allow them to at least stabilize, if not actually strengthen, their ESP abilities. Little did I realize what a hornets' nest of controversy I would stir up! The initial research, published in my Parapsychology Foundation monograph (Tart, 1975), and more widely in my Learning to Use Extrasensory Perception book (Tart, 1976a), has stood up rather well to lengthy questionings and attacks from O'Brien (1976), Stanford (1977), Gardner (1977), but today I find myself under fire from Gatlin's guns. I shall try to address myself to some useful questions that arise when immediate feedback of results is provided to percipients.

When the vast majority of ESP studies were done without immediate feedback to the percipients, the question of possible biases in target sequences, significant departures from equiprobability and serial independence, artifactually inflating the results was rather easily handled. Unless the global biases of the percipients just happened to match the global biases of the target sequence, a matter that could be checked by control matchings, such biases were not important unless of very large magnitude. With the increasing use of immediate feedback about target identity, in an effort to stabilize

and improve ESP performance (Tart, 1966; 1977e), questions about possible effects of biases in the target sequences are more important, for we can conceive of a percipient gradually learning what the biases of the target sequences are and then altering his own response strategies to take advantage of them, thus creating artifactual "hits" which might tell us something about mathematical inference strategies, but little or nothing about ESP.

In the brief time allotted to me I shall try to outline some new perspectives on this issue that I have developed, such as finding that the standard Chi-square measures of bias are not very useful measures of the predictability of a biased sequence, and present a brief description of the results of a powerful mathematical inference predictor program developed by Eugene Dronek and me that attacks the problem of prediction of biased sequences directly. Unfortunately, while it is useful that Dr. Gatlin raised this issue here today, her own solution to it is gravely flawed in a variety of ways, and the particular conclusions she has reached are invalid. It will be my unpleasant task today to demonstrate this. These flaws include such things as a persistent failure to understand the difference between valid prediction and trivial postdiction, the classical error of equating correlation with causation, confusing potential and proof, claiming enhanced sensitivity for her statistical procedures when they probably lack validity, making at least one major claim for which she presents no supporting data at all, and interpreting invalid statistical abstractions in ways which would have revealed themselves as obviously false if she had looked at the raw data they were based on. I shall detail these problems below, as they apply to Dr. Gatlin's analyses of my feedback training data. Dr. Pratt will comment on Dr. Gatlin's analysis of the Martin-Stribic data.

Gatlin's Main Assertions:

As a number of colleagues have remarked to me that they have had difficulty following the written version of Dr. Gatlin's paper (Gatlin, 1978c),

let me briefly summarize her main assertions. She asserts that:

(1). Given two numerical sequences that not only are each biased, but have matching patterns of biases, when you compare these sequences with each other you will get a higher number of identical numbers in the same positions, hits, than you would expect if you mistakenly assume that the sequences are random and unbiased.

(2). Both some target and many response sequences in my first feedback Training Study show significant degrees of bias, at the singlet, doublet, and triplet levels.

(3). Because of immediate feedback about the identity of each target, data is available to percipients from which they might calculate characteristics of possible biases in their individual target sequences.

(4). The human mind has "fantastic" (her term) capabilities, presumably unconscious, for pattern recognition in numerical sequences that allow use of the information obtained through feedback: as Dr. Gatlin puts it, ". . . extremely subtle biases at high n-tuple levels in finite sequences can be utilized by the human mind."

(5). The percipients in my study not only had the potential to utilize such bias patterns, they did use them to obtain most or all of the hit scores above chance expectation. Therefore, Dr. Gatlin asserts that:

(6). All of the above-chance scoring in my first Training Study can be explained by percipients figuring out and utilizing the target sequence biases with an unconscious mathematical inference strategy, so there is no need to postulate ESP as an explanation of the data.

There are other miscallaneous assertions in Dr. Gatlin's paper, but I believe I have adequately outlined her main argument here.

What Are the Biases in the Target Sequences?

Although I am probably more familiar with Dr. Gatlin's D-measures of bias

than most of you, I still find them difficult to follow, so in order to look at potential biases* in the target sequences of my first Training Study I shall present them in more familiar Chi-square measures, to which Gatlin's D-measures are so strongly related** that for practical purposes they are equivalent, in spite of her stress on their uniqueness. Table 1 presents both Chi square measures I have computed and the few D-measures Dr. Gatlin presented in her paper (she presented only those reaching statistical significance) for the singlet and doublet levels.

- - - - - - - - - - - - - - - - - - - -

Insert Table 1 about here

SLIDE # S-1

- - - - - - - - - - - - - - - - - - - -

Two notes on the values in Table 1 should be made. My Chi-squares (and other calculations) will be close to but sometimes not equivalent to any calculated from Dr. Gatlin's analyses, as she treated the data I provided her in a slightly less accurate manner by filling in target data associated with Passes (no response) by the percipients with a response digit from her computer

---

*I shall use the term "bias" in a general way in this paper to describe even the slightest deviation from an equiprobability and serial independence model, with the question of whether such bias is only a random fluctuation or is statistically or practically significant handled separately.

**I correlated for sets of singlet and doublet D-measures and Chi-squares and $\Psi$-Chi-squares (Davis & Akers, 1974), calculated by Dr. Gatlin, working from a computer and printout she provided me while she was still honoring her commitment to provide me with copies of all analyses she carried out on my data, and found the correlations to range from .91 to 1.00, for an average correlation of .97.

pseudo-random* number generator program, whereas I deleted these trials, since the percipients did not receive feedback when they passed. Second, the Chi-square values I have calculated for the singlet level in Table 1 are based on a model of equiprobability of all singlets (p = .10), but since there is some singlet bias, the doublet level Chi-square calculations are corrected for singlet bias by being based on marginal totals, rather than theoretical values. Without this correction, significant Chi-square values at the doublet level might be only reflections of singlet bias, rather than validly indicating a higher order bias. Insofar as I understand Dr. Gatlin's $D_2'(T)$ measure it also calculates doublet bias independent of singlet bias.

The fact shown in Table 1, that seven of Dr. Gatlin's doublet level bias measures are significant, when only three of mine are, is an interesting discrepancy. Dr. Gatlin claims greater "sensitivity" for her D-measures than for conventional Chi-square measures. Whether this claim of sensitivity of her D-measures is actually valid, or just represents an arbitrary lowering of standards for significance is a point I will leave for the more statistically erudite to work out, but what we should note here is that the "significant" departures from the model that Dr. Gatlin claims to have detected with her D-measures are even tinier than those detected with the Chi-square and, as we shall see later, tiny departures from equiprobability and serial independence

---

*Dr. Gatlin describes using a random number generator program in her paper, but the computer at UC Berkeley where she carried out her analyses uses a pseudo-random generator with an algorhythm, as practically all computer random generator programs do. It is probably satisfactorily random for short sequences.

may not be practically useful for making hits with a mathematical inference strategy.

As the Chi-square measures in Table 1 show, two target sequences were significantly biased at the singlet level and three at the doublet level. Why might this have occurred? Two of 10 NS, p , 2-tail, exact binomial.

In reporting on this bias in earlier publications (Tart, 1977a; 1977b), I pointed out that prior to collecting data in the first study, my colleagues and I were aware of the many studies which showed that subjects' desires to alter the output could significantly affect electronic random number generators (RNGs). Although we wanted our percipients to be only percipients and not agents, i.e., to use ESP but not PK, they nevertheless could score well by unconsciously PKing our electronic RNG to make its output fit their response preferences. Our instructions to the percipients to try a variety of strategies may have further enhanced the PK possibility. For this reason, we made a decision before starting the experiment to let the satisfactoriness of our RNG rest on two samples of 1,000 targets each, taken before we introduced our percipients to the equipment and after the last percipient had finished the study. These checks showed satisfactory randomicity at the singlet and doublet levels. Dr. Gatlin herself reports (Gatlin, 1978b) that the entire target sequence for the Training Study (5,000 trials) shows satisfactory randomicity at the singlet and doublet levels. The finding of significant non-randomicity in some individual target sequences is thus not unexpected, although the current issues would not have arisen if all the target sequences had shown randomicity by the standard Chi-square tests. In retrospect, the appearance of bias in some sequences has been an advantage, serving as a stimulus to clarify some important issues.

An observation supporting the idea of PK as the responsible factor for the singlet biases is that the high number in the two significant sequences

is not the same, a finding sensible in light of psychological number preferences which would probably differ between individuals, but which would not seem likely to arise for electronic reasons.

My suggestion of PK as a possible explanation for lack of randomicity in two of the ten target sequences is concerned mainly with the singlet level of bias, as there is a more prosaic explanation for the three target sequences which show significant doublet bias. As explained elsewhere (Tart, 1977b), the RNG was built with a pushbutton switch to activate it for each trial. This pushbutton was not of the type that makes a discernible "snap" or "click" when it is depressed, but a cheaper type in which resistance to being pushed steadily increases with button travel. In interviewing some experimenters after the significant doublet bias was found, they pointed out that sometimes they would push the pushbutton to obtain the next target number, notice that the number on the electronic display had not changed, and then assume that they had not pushed the button hard enough to make contact and activate the RNG, so they would push it again! This would produce a great deficiency of XX doublets (1,1s, 2,2s, etc.), and, indeed, this lack of XX doublets is the major contribution to the significant doublet results. If the ten XX doublet cells are left out of the Chi-square calculations, two of the three significant doublet tests fall to insignificance and the third one is just significant at the .05 level.

In summary, we have two of ten target sequences which are significantly biased at the singlet level, one that is significantly biased at the doublet level independent of the XX doublet lack, and two that are significantly biased at the doublet level due to the experimenter error which systematically depleted XX doublets from the target sequences. The possibility of inflated scoring through matching of this lack of XX doublets with similar biases on the percipients' part has already been discussed in the literature (Tart, 1977a;

1977b; 1977f), and shown to be trivial, given the very high level of scoring in the study.

Triplet Level Biases?

What are conspicuous by their absence in Table 1 are any triplet level tests, although Dr. Gatlin reports her $D_3'$ (T) measures and makes a number of interpretive statements about their significance. Unless her $D_3'$ measure has vastly different properties than a triplet level Chi-square, however, it seems certain that her $D_3'$ figures are invalid. In computing any statistic, we need a certain minimum sample size in order to assume it is reasonably representative of the population it is drawn from. For Chi-square tests, this is usually expressed as the rule that the expected value in each cell of the matrix must be five or greater.

In doing her triplet level tests, Dr. Gatlin is spreading a mere 500 data points over 1,000 cells, for an expected value of only one-half in each, violating the minimum expectation rule by a factor of 10. Simple calculation will show that violation of this rule leads to grossly inflated Chi-square values. For example, if every triplet in this sample of 500 were exactly equiprobable, that is that they were distributed over 500 separate cells in the triplet table, calculation would give a super-significant Chi-square, corresponding to a CR of -22! Dr. Gatlin errs in even presenting such invalid measures, much less interpreting what she thinks they show about the percipients' response patterns, or interpreting them as evidence for the existence of higher order bias patterns in the target sequences. Indeed, Dr. Gatlin must not have bothered to compare her conclusions about these invalid $D_3'$ abstractions with the actual distribution of patterns in the data available to her, or she could not have made the interpretive statements she made about them. Much of her claim about the alleged superior sensitivity of her

D-measures then, rests on invalid analysis procedures. I shall drop all further reference to Dr. Gatlin's triplet data.

Dr. Gatlin reports that the scoring rate of the percipients is significantly correlated with the degree of departure from an equiprobability and serial independence model. This is true. Unfortunately, she goes on to make the classical student error of equating correlation with causation when she states "The scoring rate is significantly positively correlated with $D_1'$ (T) and $D_2'$ (T) . . . . . Clearly the patterning in the target <u>is being used by</u> (my italics) the subjects to inflate their scores." (Gatlin, 1978c, p ).

I again suspect occasional PK by the percipients as the cause of this correlation, with the more successful percipients occasionally (unconsciously) trying a PK strategy in addition to their ESP strategies. Whatever the cause of this correlation, the considerations I shall now discuss establish that the existence of significant (and consistent) bias patterns in a target sequence establishes only a <u>potential</u> for using a mathematical estimator strategy, a potential that may not be practically useful. It is important to note that this <u>potential</u> existed in only two of the ten target sequences (for P3 and P5) of the first Training Study, and a simple and traditional way to handle it would be to just delete the data from those two percipients. This would still leave the overall results enormously significant (495 hits in 4,000 trials, CR = 5.01, $p < 6 \times 10^{-7}$, 2-tailed), but it is more useful and revealing to examine the nature of this potential, and then consider the more pertinent questions of just how much scoring can be gotten from efficient application of such potential, and whether or not there is evidence that such a potential was <u>actually utilized</u> by the percipients.

<u>Chi-Square Bias Tests Are Poor Predictor Measures:</u>

If we think that a percipient might take advantage of biases he discovers

through feedback in a target source, it is probably common to assume that the magnitude of the Chi-square measures (or Dr. Gatlin's D-measures) of departure from an equiprobability and serial independence model is a measure of how predictable the sequence is. This is incorrect, as the following examples will show.

Suppose a percipient works in an experiment involving guessing the numbers one to ten, in an experiment fixed at 200 trials. We shall deliberately use a biased target source, such that outputs 1, 2, 3, 4, and 5 all have a probability of .15, while the outputs 6, 7, 8, 9, and 10 have a probability of .05, instead of all targets being equiprobable. The observed distribution of targets in the first half of our experiment (100 trials) would look like this, deliberately giving it a perfect reflection of the target bias pattern for simplicity of illustration

| TARGETS | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| FREQUENCY | 15 | 15 | 15 | 15 | 15 | 5 | 5 | 5 | 5 | 5 |

SLIDE #5-2

The standard Chi-square test for equal frequency of observed targets would tell us that this sample of 100 is not from a random source, as we get a Chi-square of 25.00 with 10 degrees of freedom, $p < .01$, one-tailed.

Suppose our percipient takes the first 100 trials to catch on to this singlet bias pattern, so that while he has only scored the 10 hits expected under our assumed equiprobability model in the first 100 trials, he will use a mathematical inference strategy, based on his new knowledge, for the remaining 100 trials of the experiment, His best strategy is to guess a 1, 2, 3, 4, or 5 on every trial. It does not matter whether he picks one of the high five and always guesses it or randomly alternates among the high five: we would expect him to score about 15 hits in the second 100 trials. If we assess the significance of this second-half score under our assumption of equiprobability, we compute a CR of 1.67, $p < .05$, one-tailed. For the whole experiment of 200 trials, we now have (10+15) =25 hits with an associated CR

of 1.18, which, while not reaching statistical significance, might suggest to an experimenter that something was happening.

If the experiment was longer than 200 trials total and the bias pattern and mathematical inference strategy were consistent, the percipient could obviously attain conventional levels of significance as he went further. For 300 trials, for example, we would have (10+15+15)=40 hits, for a CR of 1.92, $p < .05$, one-tailed. We shall stay with a 200 trial experiment for now, however, to illustrate certain points.

Now consider a target source with a quite different sort of bias, where we observe the following distribution of targets in the first 100 trials:

| TARGETS | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| FREQUENCY | 24 | 10 | ~~8~~ 10 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |

SLIDE # S-3

This generator is highly biased toward producing ones, with no other large biases. The Chi-square test for equiprobability of singlets for this distribution gives a value of 22.40, $p < .02$, one-tailed. If we mistakenly assumed that the magnitude of the Chi-square values reflected the degree of predictability of this and the previous target sources for a mathematical inference strategy, we would think this second target source was equally or slightly less predictable than the source in the previous example. We should be quite wrong.

Again assume that it takes our percipient the first 100 trials to catch on to the bias, so he scores only 10 hits in the first 100. Now he follows the optimal strategy of calling a one for every one of the remaining 100 trials, and scores about 24 hits. For the second 100 trials alone, this gives a CR of 4.67, $p < 10^{-4}$, one-tailed. For the whole experiment of 200 trials, we have (10+24)=34 hits, with a CR of 3.30, $p < .0005$, one-tailed.

For equal Chi-square values in tests of bias, two sequences may differ enormously in usefulness for a mathematical inference strategy. Further,

there will be far, far more possible bias patterns of less usefulness for a mathematical inference strategy than there will be highly useful ones for a given Chi-square value. There are many, many ways to rearrange the bias pattern in our first example without giving it the single number peak bias pattern of our second example that is so useful in a mathematical inference strategy.

Consider a third example where the following frequences of targets are observed in the first 100 trials:

| TARGETS | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| FREQUENCY | 18 | 10 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 |

SLIDE # 5-4

The standard Chi-square test of equiprobability tells us that this is not a biased sequence, for Chi-square is only equal to 7.20, which would occur more than half the time by chance alone. Yet if our target source is really biased toward 18% ones in this way, and our percipient decides to call all ones in the second 100 trials of the experiment, he could make 18 hits there, for a second-half CR of 2.67, $p < .01$, one-tailed, and a total of (10+18)=28 hits for the whole experiment, CR = 1.89, $p < .05$, one-tailed. It is especially interesting to note that the entire sequence of 200 targets for this percipient, with the 18% bias continuing through the second 100 trials, still does not show any significant bias: Chi-square is 14.40, $p > .10$, one-tailed.

We may conclude the following for standard Chi-square tests of bias.

For long to infinite length experiments,

(1) Lack of significant Chi-square values in bias tests probably indicates lack of significant predictability by a mathematical inference strategy; and

(2) The presence of significant Chi-squares in bias tests indicates some degree of predictability by a mathematical inference strategy, but the magnitude of Chi-square does not indicate the degree of predictability.

For short to moderate length experiments of the type frequently carried out,

however,

(3) A significant Chi-square indication of bias does not necessarily indicate that a significant overall score can be obtained through a mathematical inference strategy;

(4) The magniture of the obtained Chi-square is a very poor measure of the magnitude of results that can be obtained with a mathematical inference strategy; and

(5) A mathematical inference strategy may produce significant results from a biased source which does not appear to be significantly biased by Chi-square evaluation.

The shortcomings of standard Chi-square measures of bias in realistic length experiments illustrate why more direct measures of predictability by mathematical inference strategies need to be developed.

The findings of significant singlet and doublet bias in a few of the target sequences used in the first Training Study, then, indicate a potential for some kind of mathematical estimation strategy being employed, but do not tell us either (a) how much it could usefully contribute to artifactually creating significant results; or (b) whether it was actually employed; or (c) how much such a strategy could do compared with the actual scores obtained by the percipients.

Let us now consider the important difference between prediction and postdiction.

Randomicity, Prediction, and Postdiction:

Probably the most disheartening aspect of several years of exchanges with Dr. Gatlin is her persistent failure to comprehend the enormous difference between prediction and postdiction. I shall quote three paragraphs of a letter of mine published earlier this year in the January issue of the Journal of the American Society for Psychical Research, pointing out this problem.

"Two meanings are generally associated with the concept of randomicity. The first is that no patternings or dependencies of any sort can be found in a sequence of random numerical data. The second is that randomicity means a lack of predictability of a numerical sequence: that is, given a sample of the sequence, one cannot predict subsequent numbers in the sequence with greater than chance success.

While the second meaning associated with the concept of randomicity is important for both psychological and parapsychological research, the first is false. Mathematically, one can take any sequence of numbers of any finite length, even if they have been generated by a truly random process, and find an algorithm which would deterministically generate that exact sequence of numbers. This seems to imply that the sequence of numbers was not random, but resulted deterministrically from that algorithm, and thus had a pattern to it that could be detected and made use of. However, the algorithm so determined will not successfully predict further numbers gathered from the same random source at a level beyond chance expectancy. To put it another way, we can always find some kind of pattern in retrospect, a process akin to the psychological process of rationalization or projection, but that does not mean that the sequence was actually generated in that fashion or that it is predictable."

More technical discussions of these points can be found in Chaitin, 1975, and Gardner, 19 . My letter continued:

"Thus the question of whether Dr. Gatlin's post-hoc analyses can find any kind of pattern (in the sense of departures from p equaling exactly one-tenth) in my target data is not really the relevant question: such patterns can be found, to varying degrees, in the data of any and every psychological and parapsychological experiment. The relevant question is whether such patternings, sequential dependencies, or biases exist in the target data

to a degree strong enough to have allowed percipients in the Training Study to figure out these biases as they went along (not post hoc), and make use of them to boost their scores to a level high enough to make unnecessary the occurrence of ESP as an explanation." (Tart, 1978, p 82).

Dr. Gatlin claimed in her JASPR letter (Gatlin, 1978a) to which I was responding that her "monotone" (singlet) strategy scored significantly with eight of the ten target sequences in my study. As I pointed out in my reply, this strategy apparently consisted of post hoc counting of the frequency of observed singlets in the entire sequence and then pretending you had called that highest singlet for your every response! In the real world percipients do not have all this data until after their calls are made, so Dr. Gatlin's postdictive procedure is quite spurious. I gave an example of performing a Gatlin monotone postdictive strategy on 25 random numbers taken from a random number table: I scored six spurious hits, for a binomial probability of .03. Dr. Gatlin is strong on higher order biases: using a doublet monotone postdictive strategy of the same type, I scored 10 hits in 25 trials, etc. The higher the level of this postdictive strategy, the higher your score is for a sequence with any bias in it, or even on a random sequence.

Given the total fallaciousness of any kind of postdictive strategy, I found it hard to believe that Dr. Gatlin would continue to use it after it was pointed out, but she has. In her latest publication (Gatlin, 1978b) it is now given the impressive sounding title of a "Maximal Markov-3 strategy." To quote Dr. Gatlin's current paper, "It is instructive to calculate how high the subjects could score if their estimates were 100% accurate. If we count the triplet frequencies in each individual target sequence and use these as a basis for a simple guessing strategy, which we will call a maximal Markov-3 strategy symbolized as MM3, wherein the subject guesses the symbol

most likely to occur, given the two preceeding symbols in the target, the Z-scores range from about 17 to 19 which is substantially higher than any observed in the experiment." (Gatlin, 1978c, p.13).

In her recent letter to the Journal of the American Society for Psychical Research (Gatlin, 1978b), Dr. Gatlin insists that she does know the difference between prediction and postdiction, yet she again gives a postdictive strategy as an example of her knowledge! Now perhaps I'm old-fashioned and conservative, but the dictionary definition of the verb "predict" is "To tell or declare beforehand . . .", being derived from Latin roots meaning to speak about something before it happens. Dr. Gatlin just gives another example of what percipients might have done, given her later knowledge, but this is hardly predicting. Her examples remind me of the newspaper columns of stock market analysts who always brilliantly explain why the market behaved the way it did last week. These analysts seldom make any money on the market.

Dr. Gatlin further argues in this letter of response that I misunderstook her monotone guessing strategy, and declares that ". . . in eight out of the 10 sequences the probability (Dr. Gatlin's italics) of scoring significantly is 10% to 40% . . ." (Gatlin, 1978b, p. 296). What this means in terms of actual data is that if a percipient had happened to guess the one correct out of ten possible monotone strategies right at the start of his or her responses they could have scored at the $p < .05$ level, 2-tailed. Aside from the fact that there are many more ways of guessing wrong with this strategy than guessing right with it, a simple inspection of the data would have revealed that no percipient used such a monotone strategy! Further inspection of the data would have shown that even if they had used it to maximal advantage, their total hits scores would have been enormously less significant than they actually were! If I had bought many shares of

a certain stock last week I would indeed have been rich this week, but . . . .

As long as we are on the unpleasant subject of meaningless statistical procedures, I should comment on the so-called "control" analyses reported by Dr. Gatlin that presumably support her main thesis. For these analyses she matched a computer pseudo-random number generator output against the target sequences and then artificially created a number of hits equal to those the actual percipients made on each target sequence by looking at the target list at random intervals and simply changing the pseudo-random output of the computer generator to match the target and thus make a hit. She reports that these response sequences showed none of the significant D-measures that the actual percipient response sequences did.

Her procedure amounts to taking a very small sample from a slightly biased sequence. A small sample, of course, is unlikely to have a <u>detectable</u> bias in it simply by the large reduction of N. Then mixing this small sample with several times as many pseudo-random numbers dilutes any bias even further. It is no wonder no biases were found. Indeed, the pseudo-random generator used was probably of the same type that Dr. Gatlin standardized her D-measures on in the first place. I cannot understand what meaning this so-called control analysis has.

<u>Magnitude of Bias Versus Pattern of Bias</u>:

Ignoring for the moment the triplet and other fallacies in Dr. Gatlin's analyses, supposing we assumed that her analyses at least demonstrated the <u>possibility</u> that a mathematical inference strategy might have been used by at least some percipients. If we look at the data available to Dr. Gatlin to see if they actually support this possibility, we shall see evidence that the percipients did <u>not</u> use such a strategy.

To use a mathematical inference strategy, a percipient should pattern his or her response strategies as close to his estimate of target bias patterns

as possible. We would then expect to see a correspondence between the most frequent bias patterns in the target sequence, the things that would be most useful for an inference strategy, and the percipients' response patterns. Dr. Gatlin claims that such bias pattern matchings exist, as in the opening statement of her discussion section when she states, "The matching patterns (my italics) demonstrated in the target and guess sequences of these two independent sets of ESP data indicate that extremely subtle bias at high n-tuple levels in finite sequences can be utilized by the human mind." (Gatlin, 1978c, pp. 15-16). This claim, however, has no empirical support at all presented for it in Dr. Gatlin's paper: she has examined the magnitude of biases, but presented no data at all on whether the specific patterns of bias in target and percipient data actually match. If a target sequence is highly biased toward threes following fives, for example, and a percipient is highly biased toward responding with sixes after a target has been five, these high magnitudes of bias will not be at all useful for scoring, as the patterns do not match.

One would have expected that Dr. Gatlin would have inspected the actual bias patterns in the target and percipient data to see if they did match. Since she has either not done so or chosen not to present such data, I carried out this analysis.

If a target sequence had a high singlet level bias for eights, e.g., we would expect to see the percipient showing most of his above-chance hits on eights, rather than other targets. If a doublet level mathematical estimation strategy was also useful, then we should see many of the above-chance hits on the second term of the doublet: if nines followed fives very frequently, for example, we should have many hits on nines, as well as on the eights (in this example) that we have already defined as useful for a singlet estimation strategy. We would not expect more than a chance number of hits on targets

that were not the high ones in an estimator strategy. How do the data actually look?

Table 2 shows the target numbers which had the high singlet and high doublet occurrences for the ten percipients and their respective target sequences. The vast majority of these do not, of course, represent statistically significant biases. As can be seen, only one percipient had his highest singlet bias toward the same target that was high in the target sequence, and no percipient had his highest doublet bias identical to the highest doublet bias in his target sequence. This result is not supportive of Dr. Gatlin's claims. Indeed, inspection of the variety of high singlets and doublets might lead us to wonder how the same electronic RNG could produce such different patterns for different percipients if it were as biased as Dr. Gatlin implies.

- - - - - - - - - - - - - - - - - -

Insert Table 2 about here SLIDE #5-6

- - - - - - - - - - - - - - - - - -

The data in Table 2 do not completely disprove that the percipients might have used a mathematical inference strategy, however, for their incorrect high response biases might have represented personal idiosyncracies, yet they might have still picked up enough bias toward the target sequence highs to artifactually produce the extra-chance hits which made their scores significant. A more direct test of the hypothesis is shown in Table 3. For simplicity, only the five percipients who scored significantly above chance are shown. The second column is the CR of their observed hits, given the equiprobability model. To compute the CRs in the third column, all the trials (and, of course, hits) on the most frequently occurring singlet target were deleted from the total trials: if the percipients were only using a singlet estimator strategy, this should reduce their scores to chance. Obviously it

does not. They were scoring quite significantly on many other targets than the one a singlet estimator strategy would give them an advantage on. To test an inference hypothesis even further, all hits on the relevant target predicted by the most frequent target doublet were also deleted, yet the CRs in the fourth column show the percipients still continue to score very significantly.

- - - - - - - - - - - - - - - - - - -

Insert Table 3 about here SLIDE #S-7

- - - - - - - - - - - - - - - - - - -

Thus straightforward inspection of data on the relevant bias estimates strongly suggests that the percipients were not using a useful estimator strategy, contrary to Dr. Gatlin's claims.

Gatlin's Optimal Estimation Windows:

I find Dr. Gatlin's section on optimal estimation windows rather difficult to follow. Presumably she is claiming that if significant and consistent bias patterns exist in the target sequence, a percipient must detect what these patterns are early enough in the experiment to be able to effectively use them to inflate his or her scoring. At least this is my reading of what would make sense. The procedure Dr. Gatlin follows to presumably demonstrate that this was possible is unclear to me, however. At one point it seems to involve the calculation of 300 correlation coefficients as she gets her $r_{net}$ values for $D_1'$, $D_2'$, and $D_3'$ for ten incremental sample lengths over ten percipients, and such a large number of correlations seem bound to yield some significant values by chance alone. The final outcome of all this is even more puzzling, however.

If a percipient actually figures out a significant bias pattern in a target sequence and exploits it, what we basically expect to see is scoring near chance level for a while, but increasing markedly and steadily once the bias pattern is grasped. That is, we would expect a "learning" curve, a

steady increase in scoring with further trials. This expectation of learning is especially applicable for higher order biases, where it might take some time and a few partially successful starts for the percipient to get it right. As you know, my primary interest in this data was in looking for learning, and while I reported some evidence for it (Tart, 1975, 1976a), it was hardly as pervasive as we would expect from the application of successful mathematical inference strategies by the percipients!

Consider the data presented by Dr. Gatlin in her Table 2 on optimal estimation windows. For convenience, I have presented them graphically in Figure . Her procedure seems to be a matter of correlating samples of the total target distribution for a percipient with the whole target distribution of that percipient. Since the sample sizes start as only a small (5%) sample but get increasingly bigger as they increment, we would expect these correlations to get increasingly bigger: after all, we are correlating increasingly more adequate samples of a distribution with itself. Yet Dr. Gatlin's own analyses show that the singlet samples get better for a while and then get worse: aside from the mathematical oddness of this, which makes me suspect some error, how is this supposed to be helpful to a percipient? The doublet level correlations stay relatively constant as soon as the sample size increases above its initial very small size, and this is probably mainly a reflection of the consistency of the experimenter error that led to the systematic depletion of XX doublets, an error I have already shown (Tart, 1977f) to require only a trivial correction of the results. Perhaps someone else can figure out exactly what Dr. Gatlin did to obtain these results and what significance, if any, they might have.

- - - - - - - - - - - - - - - - - - - -

Insert Figure about here

SLIDE # 5-8

- - - - - - - - - - - - - - - - - - - -

Applying a Powerful Inference Strategy:

From the analyses I have presented, I have shown that while some potential may exist in a small minority of the target sequences by which some sort of mathematical inference strategy might have contributed to at least some part of the observed results, Dr. Gatlin's claims about this are invalid because of rather basic procedural and statistical flaws. Indeed, if she had inspected the raw data she based her analyses on, she would have seen that they contradicted her assertions. As I mentioned at the beginning of this paper, however, the general question of the extent to which mathematical inference strategies could affect results in feedback studies is an important question, so let me briefly address it directly by describing the results of a powerful inference strategy applied to the data of the first Training Study. This will be familiar to some of you as I briefly touched on it in my Presidential address (Tart, 1977b) last year.

The hypothesis that percipients can inflate their scores by a mathematical inference strategy as a result of figuring out target biases needs to be cast in a specific and testable form to be scientifically useful. Fortunately, mathematical inference lends itself to precise definition. Eugene Dronek, a colleague in the Computer Sciences Department of the University of California at Berkeley, and I are now submitting for publication the results of a very powerful mathematical inference strategy that we call the Probablistic Predictor Program (PPP). I am primarily responsible for the basic strategy, and Dronek is primarily responsible for its practical implementation on the computer.

We set ourselves the task of devising a computer-assisted inferential calling strategy that would have enormously more power than we could reasonably attribute to human percipients. We gave our program powers such as an absolutely perfect memory for all previous targets to date, all previous target doublets, etc., up to all previous target sextuplets, as well as perfectly accurate and

well nigh instantaneous (in terms of human time) computing capacity to assess possible biases.

To get an overview of what the PPP does, assume that the 101st trial is coming up. To make its call, our PPP inference program looks at all hundred previous targets which have come up on previous trials. It has already sorted them into a singlet file, a doublet file, and so on through a sextuplet file. It looks at the singlet file, asks what has been the most frequent singlet to date, and, given 100 trials, what is the exact binomial probability that a singlet should have appeared with such an observed frequency compared to the null hypothesis that all singlets have an equal probability of one-tenth? This exact binomial probability is computed and stored. The program then asks if there is relevant information in its doublet file: that is, say the 100th target was a 7. Does the doublet file have any information on what 7s have been followed by in the previous 100 trials? If not, it will guess on the basis of the most improbable (compared to the null hypothesis) target to date in the singlet file, but if the doublet file does have relevant information, it will again compute the exact binomial probability of that many or more doublets having occurred in the 100 trials to date, compared to the null hypothesis of equal probability for all possible doublets. This binomial probability will then be compared to the binomial probability of the highest singlet to date: if the highest doublet to date is less probable, i.e., represents more of a departure from the model of sequential independence than the highest singlet to date represents as a departure from the equiprobability model, the program will use that doublet information as the basis of its guessing strategy. Similarly, if there is a relevant triplet, quadruplet, quintuplet, or sextuplet, the most radical departure from the model of equal probability and sequential independence will be used as a basis for the guessing strategy. On the 102nd trial, all computations will be re-done

because there is now a data base of 101 trials instead of 100, etc., so the program constantly updates itself in order to get the maximum information from all the material to date. Because of this updating, it is quite sensitive to locally shifting biases, as well as general biases.

Note that our PPP program is not based on any assumptions about representative samples or the like: it deals with what has actually been observed to date and always tries to capitalize on these observed frequencies.

Figure  is a comparison of what our inferential strategy program, with all its advantages, can do on the target sequences, compared to the scores of the actual percipients of the first Training Study. As you can see, the PPP manages to reach statistical significance on only two of the ten target sequences, and it is generally scoring well below the actual percipients' scores. In two cases of percipients who did not show individually significant ESP scores, the inferential strategy program did better, although it did not reach statistical significance. Indeed, the PPP scored at chance

- - - - - - - - - - - - - - - - - - -

Insert Figure  about here  SLIDE # T-6

- - - - - - - - - - - - - - - - - - -

(CR = .15) on the target sequence of the most successful percipient (P3), for the biases in the target distribution were not <u>useful</u> for prediction, even if they were <u>statistically</u> significant. In general, the PPP could get only about 30% as many hits as the actual percipients got over the whole study, and even if we adjusted the percipients' scores downward accordingly on an hypothesis that they arose from both ESP <u>and</u> a mathematical inference strategy, the amount of ESP in the experiment was still enormous.

Dr. Gatlin puts much emphasis on higher biases in speculating that a mathematical inference strategy was used by the percipients. I must note that we compared various levels of sensitivity of the PPP to higher level

biases, ranging from allowing it to operate only at the singlet level all the way up to the sextuplet level. Adding levels above the singlet was of very little help to the PPP, as practically all of its significant performances came from its use of singlet level biases: there simply were not higher level biases,up to the sextuplet level, that were of practical use for a mathematical estimator strategy. This is perfectly in accord with the kind of operation we would expect from an electronic roulette wheel type of RNG, of course.

Evidence that Percipients Did Not Use an Inference Strategy:

I could conclude at this point that less than one-third maximum of the scoring could be attributed to the best mathematical inference strategy we have been able to devise, and allow that claim to stand until someone empirically demonstrates that some other predictor program applied to this data does better. I stress empirically demonstrates, for I think concern with Dr. Gatlin's repeated claim of fantastic computing and pattern recognizing abilities of the human mind, supported by invalid analyses, is a waste of our time as it stands. Let the next claimant for a predictor strategy demonstrate that it predicts.

I think I can go much further in my conclusion, however, and claim that the data strongly suggest that either no mathematical inference strategy was used to any significant extent by the percipients, or, if one was used, it was a considerably less powerful one than Dronek and I have devised, and so would leave even more than 70% of the hitting in the first Training Study attributable to ESP. This conclusion has already been suggested by the analyses of the patterns of target and response bias I presented earlier, which showed that the percipients' strong response biases were almost always different from the slight biases in the target sequences. Further studies of the internal patterning of the data give even stronger support to this conclusion.

In my Presidential Address last year (Tart, 1977b) and elsewhere (Tart, 1977c), I reported on the discovery of a strong, negative relationship between the magnitude of ESP hitting on the real time target and the magnitude of hitting on the immediate future (+1) target, a finding independently significant in two separate studies. I proposed a theory of transtemporal inhibition, an information processing mechanism for ESP, to account for this relationship, and I also pointed out that the finding was quite robust: if all the target sequences showing significant singlet bias (three, in two studies) were deleted from the computations, the relationship was still strongly and significantly present. Figure shows a typical example of the temporal displacement pattern of hitting and missing with a talented percipient, with strong real time hitting, strong missing on the +1 future target and on the -1 and -2 past targets (due at least partly to response biases), and general positive and negative, largely chance, fluctuations for other time displacement registers.

- - - - - - - - - - - - - - - - - - -

Insert Figure about here

SLIDE # T-4

- - - - - - - - - - - - - - - - - - -

As a control on a possible mathematical estimation strategy artifactually creating the negative relationship between real time and +1 hitting, I ran the same sorts of analyses for temporal displacements on the responses of the PPP to the target sequences. Figure is a temporal displacement analysis on the same target sequence as that used in Figure , and it is typical of the PPP results. I could tabulate various parameters of these analyses statistically and show enormous differences, but the two figures convey the essential point: a mathematical estimation strategy like the PPP gives internal patterning to the data that is obviously nothing like that shown by actual percipients. The differences are discussed in more detail

elsewhere (Tart, 1977b).

- - - - - - - - - - - - - - - - - - -

Insert Figure about here SLIDE # T-5

- - - - - - - - - - - - - - - - - - -

Any proponent of a mathematical estimation theory, then, has a challenge here: what kind of predictive strategy can they devise that can be empirically demonstrated to both produce the enormous number of hits the percipients showed and the same internal patternings?

Conclusions:

Time does not allow me to adequately review the various points touched on in this paper, nor do I wish to dwell further on the inadequacies of Dr. Gatlin's arguments and analyses. I shall close by just mentioning the main positive contributions and conclusions that have come out of this discussion.

First, the potential importance of inflating scores through some kind of mathematical inference strategy in ESP studies employing immediate feedback of results has been underscored.

Second, this discussion has emphasized that the predictability of a target sequence is our primary concern in this matter, not departures from randomness per se.

Third, the inadequacy of the standard measure for bias, the Chi-square test (and Dr. Gatlin's D-measures) for telling us how predictable a target sequence is by mathematical inference strategies has been demonstrated empirically: I am sure some of you who are more mathematically inclined than I could demonstrate it in elegant mathematical form.

Fourth, the need to deal directly and empirically with the question of predictability of target sequences, rather than inferentially, has been underscored. The contribution of Dronek and myself that has been submitted for publication is a powerful start in directly dealing with this question.

Fifth, as to the specific question of whether the percipients in my first Training Study actually inflated their outstanding hitting scores by a mathematical inference strategy, in addition to or instead of using ESP, several analyses sketched here have demonstrated that they could not have done anywhere nearly as well as they did with a known estimator strategy; further, the evidence strongly suggests they did not use a known estimator strategy to any significant extent.

The results of this first Training Study have now been thoroughly questioned on a wide variety of grounds in the literature (Gatlin, 1978a; 1978b; Gardner, 1977; O'Brien, 1976; Stanford, 1977), and I believe the results indicating the presence of very high levels of ESP in the data have withstood this questioning extremely well (see Tart, 1976b; 1977d; 1977f; 1978a) making them some of the best data in contemporary parapsychological experimentation. They suggest, among other things, that serial selection procedures can find very high scoring percipients, that immediate feedback can at least sustain if not increase ESP functioning, and that we now have a glimpse of a basic information processing strategy for ESP, transtemporal inhibition, which in turn leads to a more sensitive test for the presence of ESP (Tart, 1977b). I would suggest that there is more profit in trying to replicate and expand these findings in new experimentation than in ~~taking further retrospective pot shots at them.~~ asking further questions about the first Training Study.

References

Chaitin, G., Randomness and mathematical proof. Scientific American, 1975, 232, 47-52.

Davis, J., & Akers, C., Randomization and tests for randomness. Journal of Parapsychology, 1974, 38, 393-407.

Gatlin, L., Comments on the critical exchange betweeen Drs. Stanford and Tart. Journal of the American Society for Psychical Research, 1978, 72, 77-81. (a)

Gatlin, L., Dr. Gatlin's reply to Dr. Tart. Journal of the American Society for Psychical Research, 1978, 72, 294-296. (b)

Gatlin, L., A new measure of bias in finite sequences with applications to ESP data. Paper, Parapsychological Association, St. Louis, 1978. (c)

Gardner, M.,

Gardner, M., ESP at random. New York Review of Books, 1977, August 14.

O'Brien, D., Review of Tart's "Application of Learning Theory to ESP Performance", Journal of Parapsychology, 1976, 40, 76-81.

Stanford, R., The application of learning theory to ESP performance: a review of Dr. C. T. Tart's monograph. Journal of the American Society for Psychical Researbh, 1977, 71, 55-80.

Tart, C., Card guessing tests: learning paradigm or extinction paradigm? Journal of the American Society for Psychical Research, 1966, 60, 46-55.

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

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

Tart, C., Reply to O'Brien. Journal of Parapsychology, 1976, 40, 240-246. (b)

Tart, C., Improving x real-time ESP by suppressing the future: trans-temporal inhibition. Paper, Insittute of Electrical and Electronic Engineers, New York, 1977. (a)

Tart, C., Space, time, and mind. Presidential Adress, Parapsychoclogical Association, Washington, D.C., 1977. (b)

Tart, C., Psi: Scientific Studies of the Psychic Realm. New York: Dutton, 1977. (c)

Tart, C., Toward humanistic experimentation in parapsychology: A reply to to Dr. Stanford. Journal of the American Society for Psychical Research, 1977, 71, 81-102. (d)

Tart, C., Toward conscious control of psi through immediate feedback training: some considerations of internal processes. Journal of the American Society for Psychical Research, 1977, 71, 375-408. (e)

Tart, C., Psi and science. New York Review of Books, 1977, October 13. (f)

Tart, C., Dr. Tart's reply to Dr. Gatlin. Journal of the American Society for Psychical Research, 1978, 72, 81-87.

Table 1

Bias Measures of Target Sequences

in the First Training Study

| Percipient | Singlet $\chi^2$ | Gatlin $D_1'(T)$ | Doublet $\chi^2$ | Gatlin $D_2'(T)$ |
|---|---|---|---|---|
| P5 | 38.81* | 6.47* | 99.08 | 2.73* |
| P3 | 35.53* | 7.66* | 137.70* | 6.48* |
| P4 | 17.04 | | 97.37 | 2.20* |
| P2 | 16.34 | | 107.84* | 2.51* |
| P14 | 15.08 | | 104.70* | 2.49* |
| P1 | 13.47 | | 87.81 | |
| P32 | 12.12 | | 76.23 | |
| P17 | 11.83 | | 101.43 | 2.04* |
| P11 | 5.13 | | 75.66 | |
| P7 | 2.64 | | 93.32 | 2.01* |

* indicates $P < .05$, 1-tailed.

Table 2

High Frequencies of Target Generator

Versus Percipients

| | Singlets+ | | Doublets+ | |
|---|---|---|---|---|
| Percipient | Target | Response | Target | Response |
| P5 | 7* | 7* | 8,5 | 7,5* |
| P3 | 5* | 2* | 9,5* | 9,2* |
| P4 | 9 | 7* | 5,9 | 7,9* |
| P2 | 9 | 5 | 5,9* | 2,9 |
| P14 | 9 | 5* | 9,0* | 5,3* |
| P1 | 7 | 9* | 4,7 | 8,9 |
| P32 | 1 | 7* | 1,0 | 2,8* |
| P17 | 3 | 8* | 9,5 | 4,8 |
| P11 | 1 | 4* | 4,8 | 4,6* |
| P7 | 7 | 9* | 7,5 | 5,6* |

* indicates that overall distribution departed from the equiprobable or serial independence model with $P < .05$, 1-tailed.

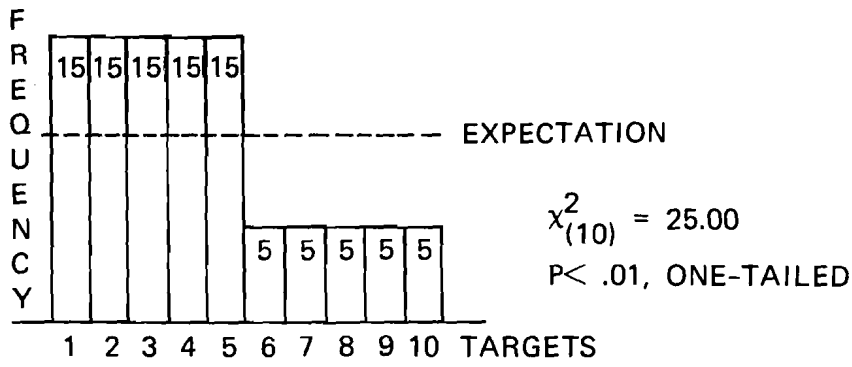
+ ties for highest rank were broken randomly

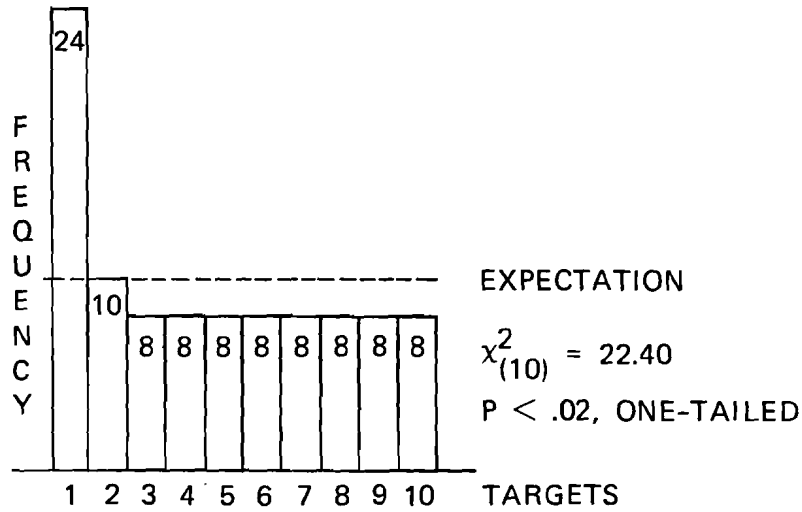
Table 3

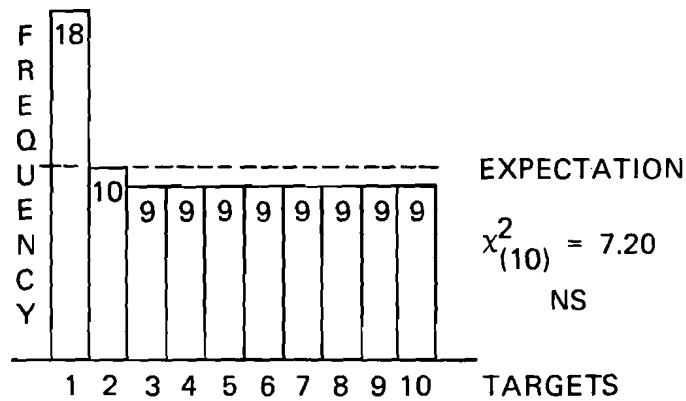
Extreme Test of Mathematical Predictive Strategy: Deleting Hits from Best Predictors as Postdicted, Singlet and Doublet Levels

| Percipient | Original Hits CR* | Deleting Hits on Highest Singlet: CR | Deleting Hits on Highest Singlet&Doublet CR |
|---|---|---|---|
| P3 | 11.03 | 9.80 | 8.06 |
| P5 | 7.60 | 4.54 | 2.28 |
| P4 | 4.47 | 3.14 | 3.14 |
| P2 | 4.32 | 3.80 | 3.68 |
| P1 | 4.17 | 3.82 | 3.82 |

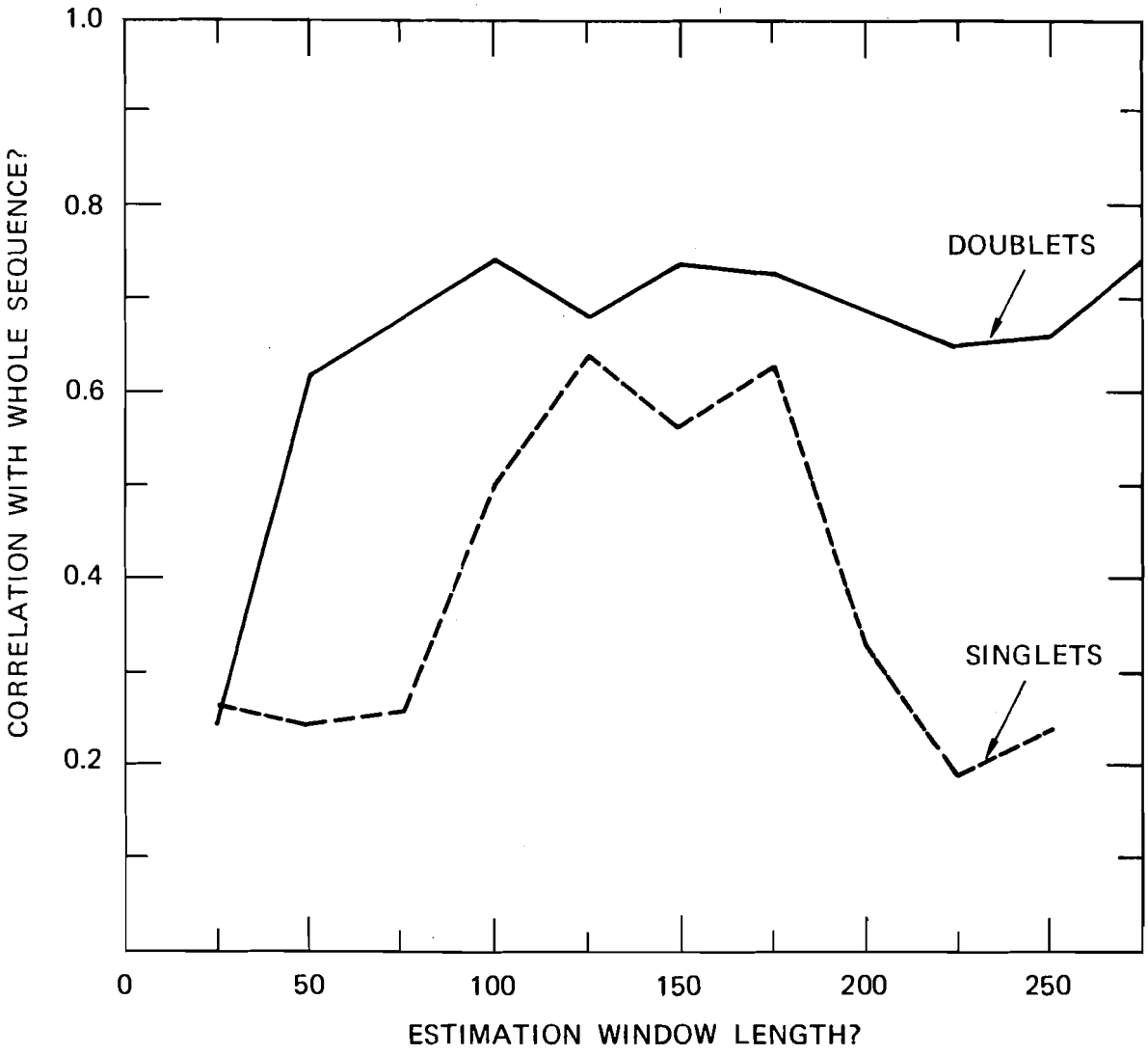
* These CRs may differ slightly from published data due to Gatlin's practice of substituting random responses for Pass data.

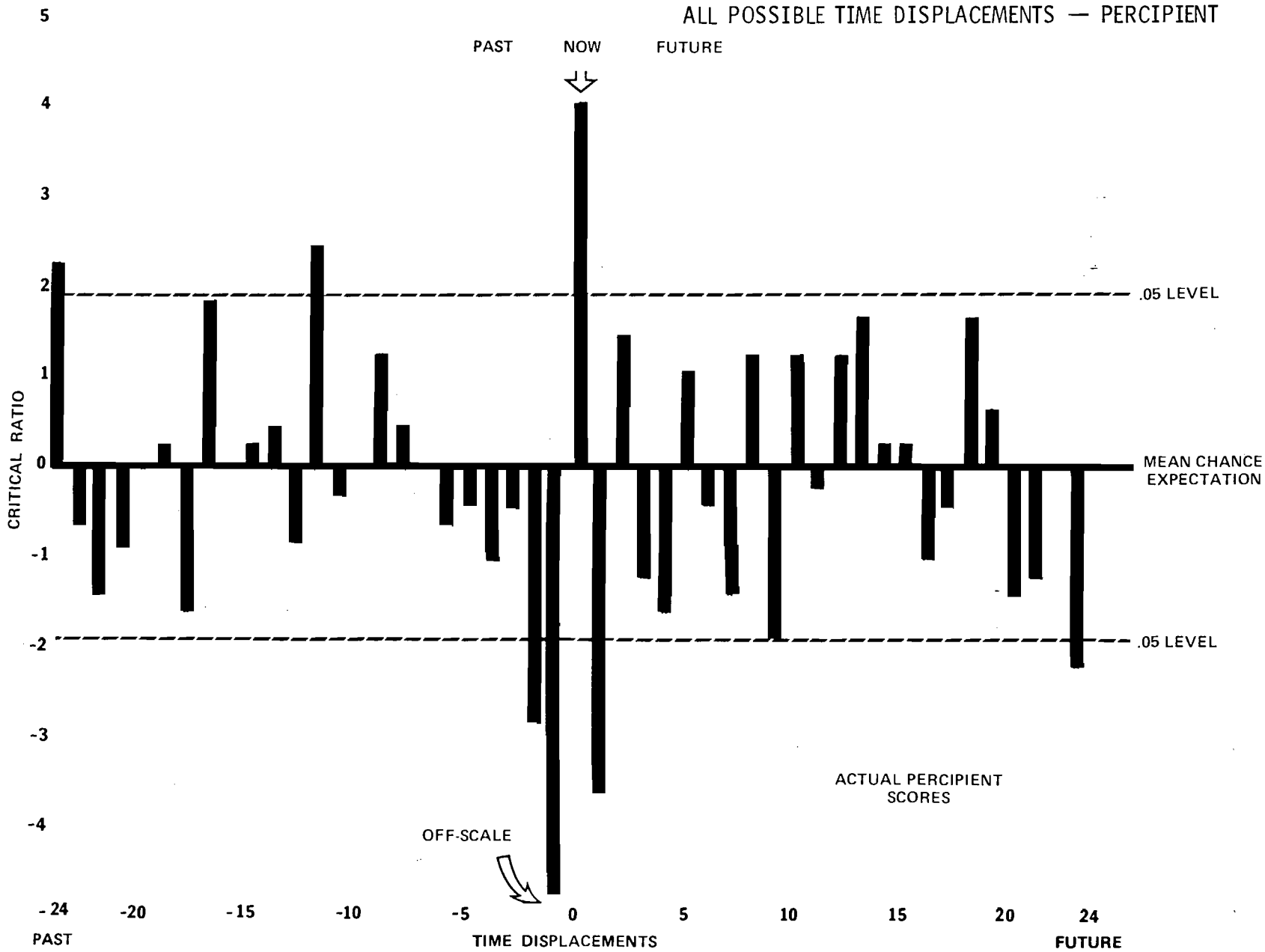

FREQUENCY
15 15 15 15 15
5 5 5 5 5
EXPECTATION
$\chi^2_{(10)} = 25.00$
P< .01, ONE-TAILED
1 2 3 4 5 6 7 8 9 10 TARGETS

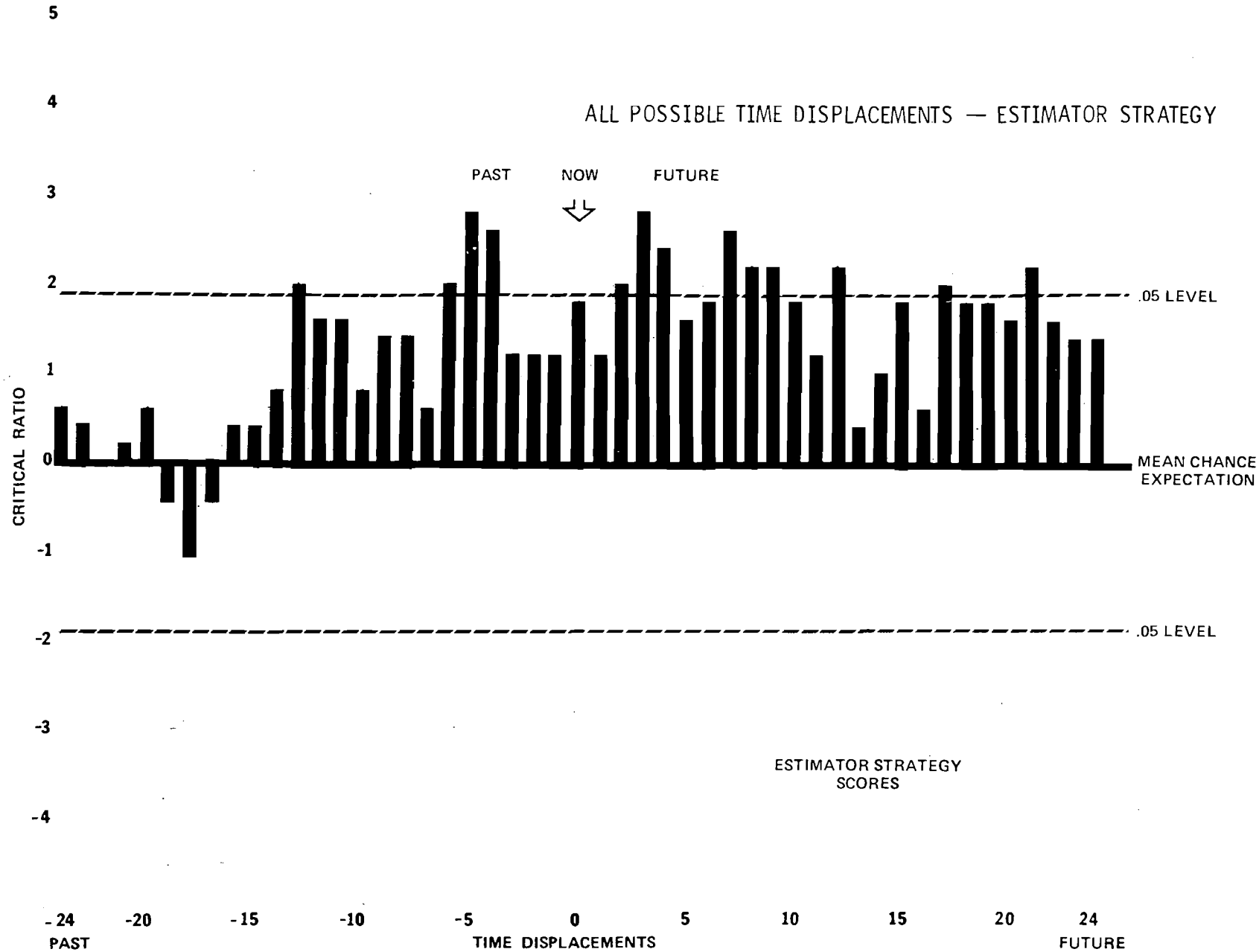

FREQUENCY
24
10
8 8 8 8 8 8 8 8
EXPECTATION
$\chi^2_{(10)} = 22.40$
P < .02, ONE-TAILED
1 2 3 4 5 6 7 8 9 10 TARGETS


FREQUENCY
18
10
9 9 9 9 9 9 9 9
EXPECTATION
$\chi^2_{(10)} = 7.20$
NS
1 2 3 4 5 6 7 8 9 10 TARGETS

# GATLIN'S OPTIMAL ESTIMATION WINDOWS




ALL POSSIBLE TIME DISPLACEMENTS — PERCIPIENT
PAST
NOW
FUTURE
5
4
3
2
1
0
-1
-2
-3
-4
CRITICAL RATIO
.05 LEVEL
MEAN CHANCE EXPECTATION
.05 LEVEL
ACTUAL PERCIPIENT SCORES
OFF-SCALE
-24
-20
-15
-10
-5
0
5
10
15
20
24
PAST
TIME DISPLACEMENTS
FUTURE


ALL POSSIBLE TIME DISPLACEMENTS — ESTIMATOR STRATEGY
PAST
NOW
FUTURE
5
4
3
2
1
0
-1
-2
-3
-4
CRITICAL RATIO
.05 LEVEL
MEAN CHANCE EXPECTATION
.05 LEVEL
ESTIMATOR STRATEGY SCORES
-24
-20
-15
-10
-5
0
5
10
15
20
24
PAST
TIME DISPLACEMENTS
FUTURE

Bias Measures of Target Sequences

in the First Training Study

| Percipient | Singlet $\chi^2$ | Gatlin $D_1'(T)$ | Doublet $\chi^2$ | Gatlin $D_2'(T)$ |
|---|---|---|---|---|
| P5 | 38.81* | 6.47* | 99.08 | 2.73* |
| P3 | 35.53* | 7.66* | 137.70* | 6.48* |
| P4 | 17.04 | | 97.37 | 2.20* |
| P2 | 16.34 | | 107.84* | 2.51* |
| P14 | 15.08 | | 104.70* | 2.49* |
| P1 | 13.47 | | 87.81 | |
| P32 | 12.12 | | 76.23 | |
| P17 | 11.83 | | 101.43 | 2.04* |
| P11 | 5.13 | | 75.66 | |
| P7 | 2.64 | | 93.32 | 2.01* |

* indicates $P < .05$, 1-tailed.

High Biases of Target Generator

versus Percipients

| | Singlets+ | | Doublets+ | |
|---|---|---|---|---|
| Percipient | Target | Response | Target | Response |
| P5 | 7* | 7* | 8,5 | 7,5* |
| P3 | 5* | 2* | 9,5* | 9,2* |
| P4 | 9 | 7* | 5,9 | 7,9* |
| P2 | 9 | 5 | 5,9* | 2,9 |
| P14 | 9 | 5* | 9,0* | 5,3* |
| P1 | 7 | 9* | 4,7 | 8,9 |
| P32 | 1 | 7* | 1,0 | 2,8* |
| P17 | 3 | 8* | 9,5 | 4,8 |
| P11 | 1 | 4* | 4,8 | 4,6* |
| P7 | 7 | 9* | 7,5 | 5,6* |

* indicates that overall distribution departed from the equiprobable or serial independence model with $P < .05$, 1-tailed.

+ ties for highest rank were broken randomly

Extreme Test of Mathematical Predictive Strategy: Deleting Hits from Best Predictors as Postdicted, Singlet and Doublet Levels

| Percipient | Original Hits CR* | Deleting Hits on Highest Singlet: CR | Deleting Hits on Highest Singlet&Doublet CR |
|---|---|---|---|
| P3 | 11.03 | 9.80 | 8.06 |
| P5 | 7.60 | 4.54 | 2.28 |
| P4 | 4.47 | 3.14 | 3.14 |
| P2 | 4.32 | 3.80 | 3.68 |
| P1 | 4.17 | 3.82 | 3.82 |

* These CRs may differ slightly from published data due to Gatlin's practice of substituting random responses for Pass data.