

## Selection Versus Influence Revisited: New Method and Conclusions

YORK H. DOBYNS

Princeton Engineering Anomalies Research, C-131 Engineering Quadrangle, Princeton University,  
Princeton, NJ 08544-5263

**Abstract** — A previous paper by the author (Dobyns, 1993) on this topic mistakenly fails to account for nonindependence in certain test measures, leading to exaggerated conclusions. An analysis that avoids this problem produces weaker statistical evidence, although the qualitative conclusions of the earlier analysis are sustained.<sup>1</sup>

### I. Background

A private communication from Jessica Utts called my attention to a difficulty in my previous article on the selection model for remote REG experiments (Dobyns, 1993). The problem appears on p. 265, immediately following formula (6): “The aggregate likelihood of the hypothesis over all three intentions may be calculated by repeating the individual likelihood calculation for each intention...” Unfortunately, while the Bernoulli formula used in eqs. (5) and (6) correctly accounts for the constraint equations governing populations and probabilities within an intention, it fails to account for the nonindependence induced by the further constraint conditions (see Section III below) operating between intentions. The alternative formulation discussed later in the same paragraph fails for the same reason; while the roles have been switched, the formula is still correcting for one set of constraints and ignoring the other. Since the component likelihoods do not derive from independent events, the aggregate likelihood formed by multiplying them is in error.

The conclusion might be salvaged by deriving a correction factor for the effects of nonindependence, but with the raw data readily available, it seems more productive to reformulate the analysis in such a way as to avoid the non-independence problem entirely.

### II. A Brief Reprise: Selection and Influence Models

For the current article to stand alone, a brief discussion of terms and experimental background seems necessary. The experimental database considered comes from remote experiments using a *Random Event Generator* (REG), a

---

<sup>1</sup>I am indebted to Jessica Utts for her detection and communication of the error in the earlier analysis. The Engineering Anomalies Research program is supported in part by grants from the Fetzer Institute, Laurance S. Rockefeller, and Helix Investments.

device which records digitized random output from a noise diode or other source. In these remote experiments, operators distant<sup>2</sup> from the device attempt to alter the machine's mean output level. These experiments are *tripolar*; each remote session comprises three consecutive runs started at twenty-minute intervals, with the operator attempting to increase the mean output level in one of the runs (H intention), decrease it in another (L intention), and leave it unaltered in a third (B intention). Each of these runs consists of  $2 \times 10^5$  binary random samples collected as 1000 sums of 200 bits each. Because of the lack of contact between operator and device, it can plausibly be proposed that the observed experimental effect derives not from an actual change in the output, but from judicious choice of intentional labeling to correspond with the random outputs of an undisturbed device generating tripolar sets. This type of effect could constitute a genuine anomaly, in which operators by some as-yet-unknown and probably unconscious means acquire information about the run outcomes and choose their intentions to suit, or it could represent a breakdown of the experimental controls in which operators somehow learned of the run outcomes before reporting their intentions. Regardless of the details, any model in which the effect is achieved by selecting the intention ordering to fit otherwise unmodified output can be considered a *selection model*. In contrast, an *influence model* assumes that any observed effects are due to actual differences in the machine's performance under different intentional conditions. It should be noted that this analysis does not, strictly speaking, address the question of whether there is in fact an effect; it aims at distinguishing the origin and nature of an effect *if* one is present. The reality or otherwise of the effect has been discussed elsewhere (Dunne and Jahn, 1992).

### III. Redefined Rank Frequency

The 1993 paper uses the term *rank frequency* to refer to the frequency with which a given intention possesses a given ordinal rank within its tripolar set. Since each run within a tripolar set has a definite rank and must also be assigned a definite intention, there are nine rank frequencies, which can most intelligibly be arrayed in a  $3 \times 3$  matrix of intentions versus ordinal ranks. The dual constraints of one of each rank, and one of each intention, in each tripolar set, manifest as a set of five independent constraint equations on these nine matrix elements. (Actually, there are six constraining conditions, one on each row and column of the matrix; however, any one of the six equations may be expressed as a linear combination of the other five, and thus eliminated.)

The nonindependence problem can be avoided by formulating the problem in terms of tripolar sets, rather than individual runs. There are just six distinct ways in which three nonidentical run outcomes can be assigned to the three intentions under the protocol constraint that each tripolar set contains exactly

---

<sup>2</sup>Distant is here taken to mean that the operator-device separation is at least on the order of a mile and frequently ranges up to hundreds or thousands of miles. Further details of the remote experiments can be found in Dunne and Jahn, 1992.

one instance of each intention. If we consider the frequency with which tripolar sets appear in each of these possible configurations, we are clearly examining a single set of six exhaustive and mutually exclusive alternatives, rather than three correlated sets of three such alternatives; the nonindependence problem that corrupted the previous analysis then becomes irrelevant. The remainder of this analysis shall be cast in terms of these frequencies of the six possible tripolar rankings. (Note that these “rank frequencies” are thus defined differently than in the 1993 paper.)

Where individual identification is necessary, rank frequencies will be labeled by the intentional subscripts assigned to the highest, middle, and lowest run of the set, respectively. Thus  $\rho_{HBL}$  refers to the frequency of appearance of the “correct” tripolar labeling in which the highest run is assigned to the high intention and the lowest to the low intention;  $\rho_{BLH}$  refers to the frequency of tripolar sets in which the highest run is labeled a baseline, the middle run a low, and the lowest run a high; and so forth. The key to the analysis is that influence and selection models predict different functional relationships between the rank frequencies and the distribution statistics of the observed data.

#### IV. Observational Database

As noted in the 1993 paper, the database comprises 494 tripolar sets. Four of these sets contain ties between intentions, a consequence of the discrete nature of the experiment but not one that can readily be dealt with in this continuous formalism. They may be discarded without appreciably altering the statistics. The overall bit-level deviations from expectation in the remaining 490 sets show the following means and standard deviations:

TABLE 1  
Bit Deviations

Intention	Mean	Std Dev
H	32.81	225.5
B	5.578	212.5
L	2.102	217.6

The theoretically expected distribution for these bit deviations is normal with a mean  $\mu = 0$  and standard deviation  $\sigma = 223.6$ .

However, to conduct an analysis against a selection model one must normalize the data, not by their theoretical distribution, but by the empirical mean and standard deviation of the pooled data themselves. This is necessary to avoid improper prejudice against the selection model. This model assumes the process is applied to  $\mu = 0$ ,  $\sigma = 1$  normal data, and its predictions necessarily have the property that if the three intentions are combined into a single collective pool, the aggregate will have  $\mu = 0$ ,  $\sigma = 1$ . The influence model, on the other hand, is indifferent to a uniform linear transformation applied to all of

the data. Another way of understanding this consideration is to say that the selection model, at least as regards means and variances, predicts the *relative* status of the intentions within the aggregate database, rather than their absolute status under the machine's theoretical output.

So, when the raw data are normalized to have an overall mean of 0 and standard deviation of 1 across all three intentions, the results are:

TABLE 2  
Normalized Statistics

Intention	Mean	Std Dev	Skew	Krt.
H	0.0883	1.0304	-0.1339	-0.0697
B	-0.0362	0.9707	0.1005	0.1481
L	-0.0521	0.9941	-0.1803	-0.4903

Table 2 has included the higher moments that will be used in statistical evaluations. These have the same values for the non-normalized data, since they are unaffected by linear transformations. The observed rank frequencies are:

TABLE 3  
Rank Frequencies

Rank Order	N (out of 490)	Observed $\rho$
HBL	92	0.188
HLB	88	0.180
BHL	80	0.163
BLH	79	0.161
LHB	87	0.178
LBH	64	0.131

The quoted values for  $\rho$  have a one- $\sigma$  statistical uncertainty of  $\pm 0.017$ , due to the number of observations.

### V. Inferences from Models

The influence model treats the distribution data in Table 2 as primary; the expected rank frequencies can be calculated from these distribution statistics through a straightforward if tedious process of numerical integration. The selection model, on the other hand, treats the rank frequencies of Table 3 as primary, and allows distribution statistics to be calculated from them. It is not, however, immediately obvious how one may interpret the results of such calculations. In the one case the prediction generates a set of distribution statistics to be compared with the observation; in the other, a set of rank frequencies is predicted. It is not clear how one may construct a single goodness-of-fit parameter that can be applied in both cases to compare the relative merits of the two hypotheses. The previous work avoided this problem by inverting the functional dependence of the selection model calculations (the integrals involved in the influence model are not readily invertible), allowing a calculation

from distribution statistics to rank frequencies for both models. This allowed a direct comparison of rank frequency predictions between the two models. Unfortunately, the redefined rank frequencies in Table 3 are not amenable to such functional inversion for the selection model; the relevant equations are nonlinear in the rank frequencies, and admit of multiple solutions for a given set of distribution statistics.

In the absence of a clear theoretical model for a goodness-of-fit comparison, it is nonetheless possible to determine the goodness of fit for each model empirically via a Monte Carlo procedure. This not only establishes each model's ideal predictions for the input data, but also directly determines the distribution of variations in the predictions, making it possible to test each model's fit to the observed data without the need for a functional inversion to create comparable predictions. As an added bonus, using the actual data in the Monte Carlo process assures that the real characteristics of the data are being accounted for to exactly the degree that they are statistically established, without any risk that simplifying assumptions in a theoretical model<sup>3</sup> are distorting the predictions.

## VI. Monte Carlo Algorithms

### *Selection by Monte Carlo*

The selection model assumes that the data are the result of a selection procedure applied to the extant tripolar sets. A certain proportion of them are correctly identified as to their ordinal rank, with the highest run labeled H, the lowest labeled L, and the middle run labeled B. Likewise varying proportions of the tripolar sets, as detailed in Table 3, are "misabeled" to various degrees of inaccuracy.

The question we ask of the selection model may be expressed thus: Given the 490 tripolar sets made available for the selection process, and given also the rank frequencies of Table 3 as the definition of the efficiency of the selection process, how likely are the observed statistics of the three intentional distributions? The question, thus phrased, is in itself nearly a specification of the desired Monte Carlo algorithm. First, we internally sort the tripolar sets so that, for each set, we can identify its highest, lowest, and midmost element. We then randomly choose 92 of the 490 to receive the "correct" HBL labeling; we assign the HLB labeling (highest run labeled H and lowest labeled B) to 88 randomly selected sets out of the remaining 398; and so forth. Once the sets have been distributed among the six possible intentional labelings according to the population figures in Table 3, we find which runs have been assigned to each of the three intentions and calculate intentional distribution statistics accordingly. Finally, we repeat the whole process many times, and see how the actual twelve-element matrix (mean, standard deviation, skew, and kurtosis for each

---

<sup>3</sup>As, for example, the assumption of normality in an influence model integration.

of three intentions) of intentional statistics compares to the distribution of many such matrices calculated in the Monte Carlo process.

### *Influence by Monte Carlo*

For the influence model, the question is: How likely is the observed set of rank frequencies, given the observed statistics of high, low, and baseline runs? Here, the procedure is to preserve the intentional identity of each run and to scramble the tripolar sets. A new group of 490 tripolar sets is created by randomly drawing (without replacement) one each from the high, baseline, and low datasets, until the data are exhausted. The rank frequencies of this re-arranged dataset are then calculated and recorded. This process is then iterated to build up an empirical distribution of rank frequency predictions.

## VII. Goodness of Fit: The Empirical Distance Parameter

To compare the single set of observational values with the distributions generated by the Monte Carlo procedure, it is simplest to regard the set of numbers as defining a single point in a multidimensional space. For the selection model, which generates twelve statistical measures, the parameter space is twelve-dimensional;<sup>4</sup> for the influence model, which produces six rank frequency predictions, the space is six-dimensional. Once we start thinking in terms of a spatial representation of the data format, however multidimensional, it becomes quite natural to think of summarizing the many differences between (say) any two Monte Carlo outcomes by the *distance* between two points in these many-dimensioned spaces.

This parameter distance presents the ultimate key to quantifying the question of whether the observed values are “like” or “unlike” the predictions emerging from the Monte Carlo calculation. For each model, we calculate the centerpoint of the distribution of Monte Carlo outcomes by taking the mean value of each “coordinate.” We can then calculate the distribution of parameter distances from all of the individual Monte Carlo outcomes to this centerpoint, and compare this distribution to the distance between the Monte Carlo centerpoint and the observed data. Figure 1 demonstrates the application of this concept in a readily visualizable parameter space of 2 dimensions. The scatterplot shows 500 points generated with a Gaussian radial density

$$p(x, y) \propto e^{-r^2/2} = e^{-(x^2+y^2)/2}.$$

(Note that this is equivalent to independent variations on both the  $x$  and  $y$

---

<sup>4</sup>Actually, the two constraints on mean and variance imposed by the selection model confine the points to a ten-dimensional hypersurface in the twelve-dimensional space. This is automatically handled correctly by the empirical treatment, since the normalized observational data obey the same constraint. A similar dimensional reduction, caused by the constraint  $\Sigma p = 1$ , applies to the rank frequency calculation; again, the use of an empirical distribution obeying the same constraint instead of a theoretical calculation dependent on the number of dimensions used automatically compensates for this problem without any need for explicitly taking it into account.

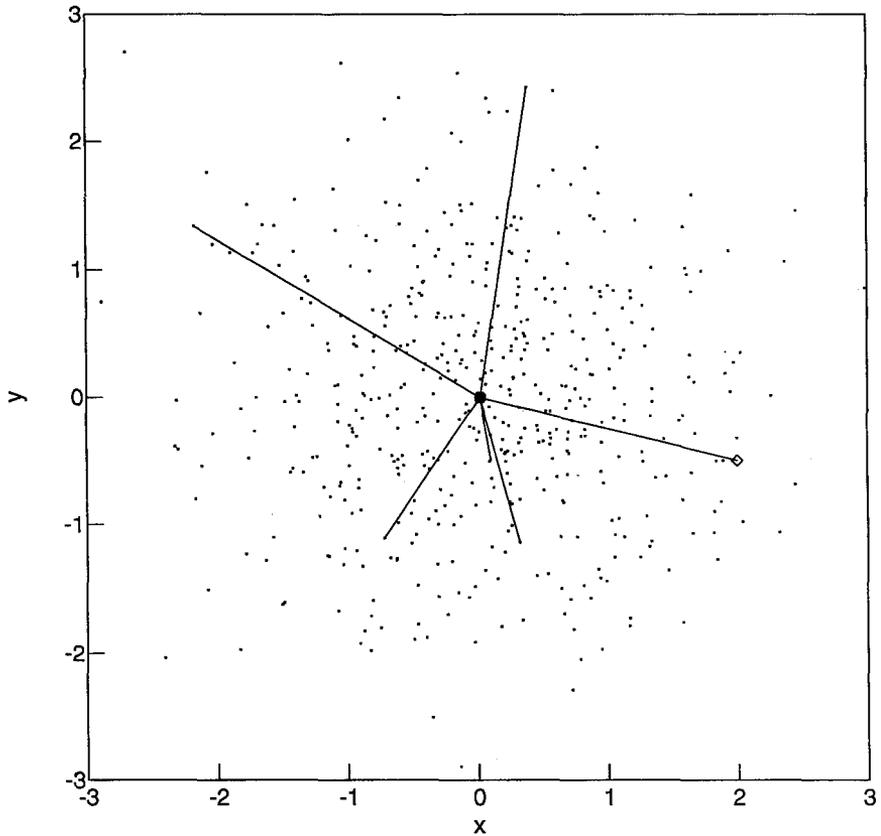


Fig. 1. The distance problem in two dimensions.

axes). The center of the distribution is marked with a filled circle; a few radial lines from the centerpoint to some of the individual points are shown. Also shown is the radial line to an arbitrarily chosen point  $x = 2$ ,  $y = -0.5$  marked by a diamond. If the statistics of the scattered points were not known *a priori* we could construct a statistical test for the likelihood that the diamond is an ordinary member of the distribution by comparing its radial distance from the collective centerpoint with the distribution of all other radial distances such as the examples shown.

#### *Axis Normalization*

An extra complication appears in the Monte Carlo distribution of selection-model statistics, since the various statistical measures calculated are not equally stable. This requires that the distance measure be normalized, as shown in Figure 2.

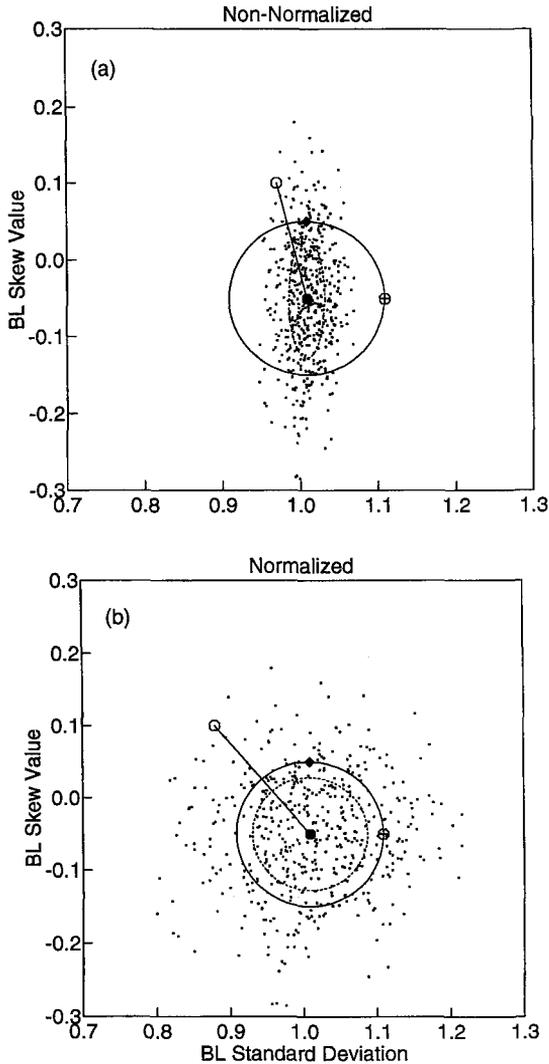
This figure shows two dimensions from an actual sequence of 500 Monte Carlo selection model runs, specifically plotting the baseline skew against the

baseline standard deviation. Again, the individual runs are plotted by points and the center of the distribution is shown by a filled circle. Figure 2a shows the distribution in absolute units; it is evident that the skew is intrinsically more variable than the standard deviation, as expected from first principles. The density contours are elliptical rather than circular in these two dimensions; the dotted ellipse is an approximate contour. The solid circle, with two points marked, shows why the distance calculation needs to be normalized for such cases. Although all points on the circle clearly share the same distance  $r$  from the center of the distribution, it is obvious that the point marked by a filled diamond is fairly typical of the distribution, while the point marked by a circle and crosshairs is extremely atypical. Since both typical and extreme points can share a common radius, the non-normalized radius is clearly not an adequate representation of how well a given point fits the distribution. For illustration, the open circle connected to the distribution centerpoint by a radial line shows the values of these two parameters in the observed data.

Figure 2b shows the effect of normalizing, in this case by amplifying all lateral distances by a suitably chosen scale factor. (The same scale is used on the  $x$ -axis for display purposes, and no longer reports the actual standard deviation value of the plotted runs.) We can see that the density contours are now approximately circular and that two points at the same distance  $r$  are in comparable regions of the distribution, regardless of their angular position. The rescaling has shifted the position of the observed data point as well as of the individual Monte Carlo outcomes; its renormalized radius is now suitable for comparison with the renormalized radii of the individual Monte Carlo outcomes as discussed with the example of Figure 1.

### VIII. Conclusions from Monte Carlo

To evaluate the two models,  $10^5$  iterations of Monte Carlo were run for each. The distances of individual Monte Carlo runs from the aggregate population centerpoints were calculated and binned to establish the probability density of the distance parameter for each model. The results are shown graphically in Figure 3. The raw bin populations are shown by the scatterplot of crosses; a smoothed version is illustrated by a continuous line. The position of the actual observed data on the distance scale is shown by a labeled vertical spike. It is quite evident that the actual data fall moderately far out on the tail of the selection model's distribution of predictions (top graph in Figure 3), while they are quite close to the peak of the distribution of predictions from the influence model (lower graph in Figure 3). The tail-area  $p$ -value can be calculated quite directly simply by counting the number of runs in the upper tail for each model, that is, the fraction of the Monte Carlo runs that are more unlike the average prediction than the observation. For the selection model, this  $p$ -value is  $0.0296 \pm 0.0005$  (the uncertainty quoted is the statistical one- $\sigma$  error in establishing a binomial probability from  $10^5$  observations). For the influence model, on the other hand,  $p = 0.347 \pm 0.002$ . Thus, by a standard  $p = 0.05$  sig-



(2a)

(2b)

Fig. 2. Why normalization is needed.

nificance criterion, the predictions of the selection model *can* be distinguished from the structure of the observed data, whereas the predictions of the influence model *cannot*. Or, to express the consequences of the Monte Carlo analysis more directly: When influence is assumed, and the existing data distributions are used to construct rank frequencies, the result is statistically indistinguishable from the actual data. In contrast, when selection is assumed, and the existing rank frequencies and tripolar sets are used to construct data distributions, the result is statistically distinct from the actual data structure.

## IX. Theoretical Concerns

### *The Nelson Problem*

R. D. Nelson, in private communication, outlined a possibility for the selection process that would increase the difficulty of data interpretation by at least an order of magnitude.

The current analysis presumes that the hypothetical selection by the operator is purely qualitative. It is assumed, for that model, that the operator has some erratic ability to discern which of the three run means is highest (or lowest, etc.). What if the efficacy of such an ability is conditioned by the distinctness of the runs? To draw a visual analogy, it is clear that humans can distinguish two different primary colors under much more adverse conditions than two lightly contrasting shades of beige. It does not seem unreasonable *a priori* that if the human participants have some ability to distinguish and sort among the experimental runs, they could more readily distinguish a (normalized) split of  $\pm 3$  between two intentions than one of  $\pm 0.001$ .

If the rank frequencies are data-dependent, the problem of predicting the expected selection distributions from them becomes very much harder. For one thing, the observed rank frequencies are already somewhat uncertain, simply due to the limited number of observations available to characterize them. If the additional dimension of variation with regard to run mean is added, we have no hope of being able to characterize their variation on the basis of the data, and would have to assume a model for such variation. Furthermore, even given a model, the calculation of expected statistics from such variable rank frequencies becomes quite intractable.

Fortunately, the proposition of data-dependent rank frequencies is amenable to a direct test, or rather, to several. It is necessary first to quantify the degree of accuracy an operator displays in making a particular assignment of intentions to a tripolar set. Clearly, there is some sense in which the "HBL" assignment is "completely right" and the backwards "LBH" assignment is "completely wrong," but how should intermediate assignment orders be ranked? There are three binary decisions that can be made in evaluating the relative rankings of a tripolar set: Is the H run higher than the B? Is the B run higher than the L? Is the H run higher than the L? (It should be noted that these three decisions are not independent, but this is irrelevant to the analysis.) If we define an accuracy index by the number of these conditions that are satisfied by a given set, we find that a set in the HBL order has an accuracy index of 3, while a set in the LBH order has an index of 0. For the other four orderings, both HLB and BHL have an index of 2, while both BLH and LHB have an index of 1. Since, in each case, there is no obvious qualitative way in which one of the two rankings with an equal index is "better" or "more accurate" than the other, this index seems a satisfactory quantitative measure of the somewhat vague notion of accuracy in judgement.

The other measure of interest is the *span* of the tripolar set, the interval be-

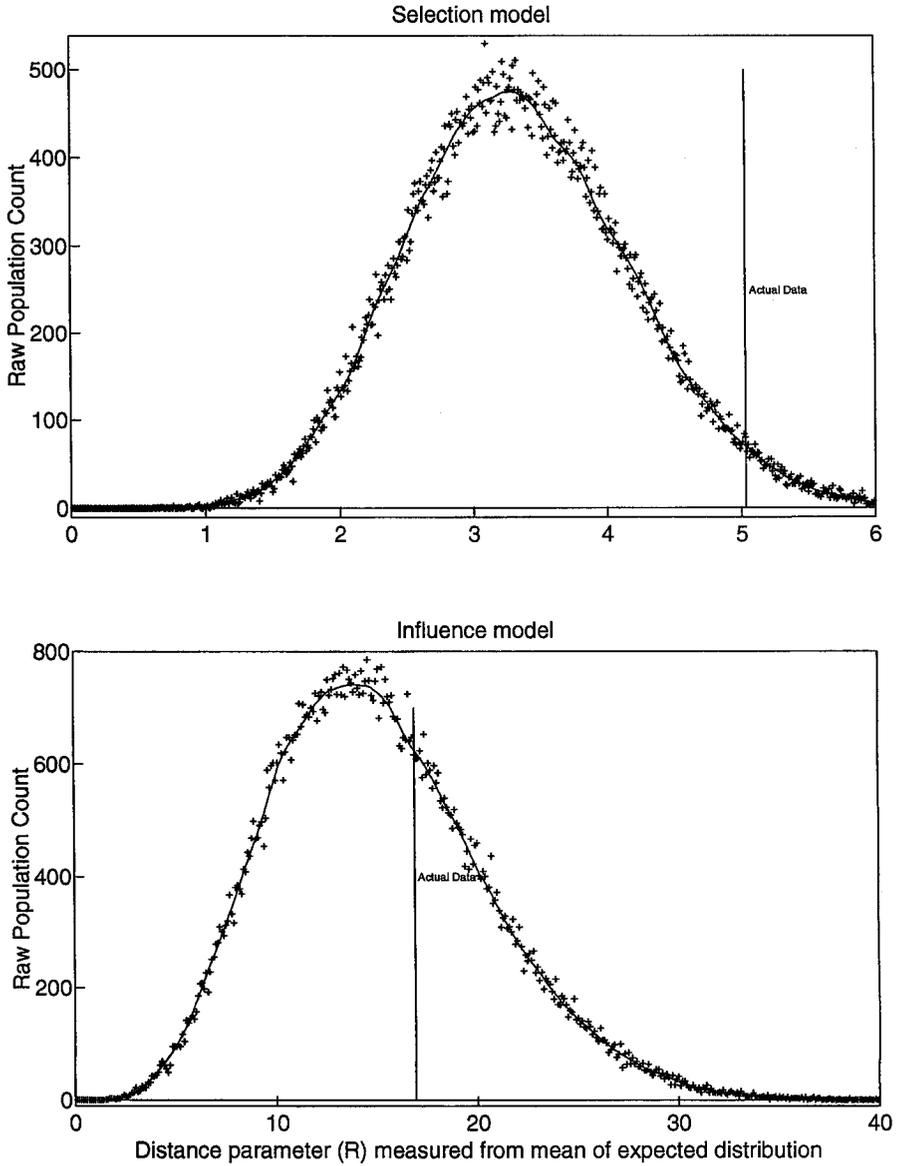


Fig. 3. Fit of theory to observation, both models.

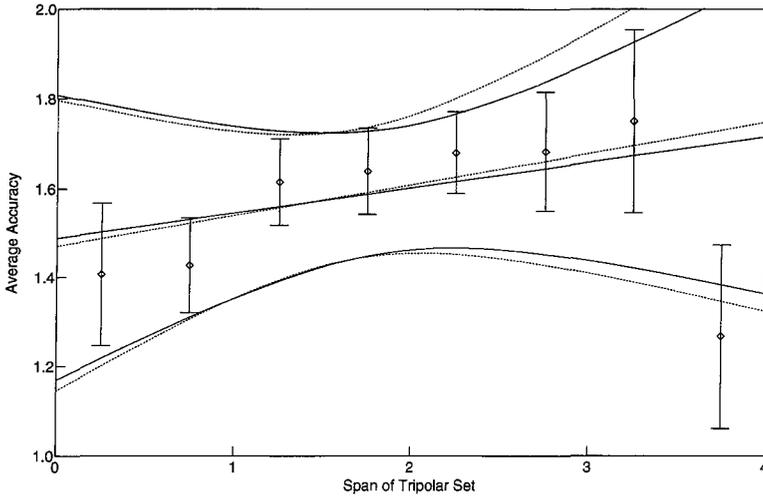


Fig. 4. Accuracy as a function of span.

tween its highest and lowest elements. The data-dependent selection hypothesis posits that accuracy should be greatest for those sets with the widest span. Figure 4 illustrates that this is not the case. The icons with error bars show the average accuracy index and associated standard error, as a function of the span of the tripolar set. These averages were calculated by a binning process; the first point represents the average accuracy for all sets with a span less than 0.5, and so forth. The rightmost bin, nominally [3.5,4.0], includes a contribution from two sets with span values greater than 4. This inclusion does not appreciably change its statistics. Two regression lines with their 95% confidence hyperbolas are shown: the solid line is a weighted regression to the averaged points, the dotted line is the regression to the actual 490 span and accuracy values. This latter may be considered more accurate, since some information is inevitably lost through the binning process; binning was conducted simply because the scatterplot of the 490 accuracy values (not shown) is difficult and uninformative to judge by eye.

The conclusion of the regressions is clear: while some slope is visible in the regression lines, the confidence hyperbolas include lines of the opposite slope, and in consequence the slope of the regression line is not statistically distinguishable from zero. Looking at the binwise averages we can note some modest suggestion, not of a trend, but of some kind of distinction: the data appear to consist of two groups, an extreme group of very large and small spans for which accuracy is poor, and a range of intermediate spans for which accuracy is somewhat better. Resolving the reality of this apparent structure must await the collection of more data, preferably in independent replications. For the purposes of the current analysis, it is sufficient to note that any span-dependence of the rank frequencies is too weak to be detectable in the experimental database. We may therefore resolve the Nelson problem by noting that the as-

sumption of constant selection efficiency used in the foregoing analysis is an adequate approximation for treating the given database.

### *Timing Selection*

The analysis above has addressed a selection model in which intentions are assigned to suit the respective outcomes of a tripolar set, or at least the probability of a match is increased by some anomalous knowledge on the part of the operator. The remote protocol used at PEAR does, however, allow one (and only one) other volitional choice by the operator, namely the time at which data collection is to start. This allows another potential venue in which a selection process could operate: rather than choosing the intentional order to fit the outcome, an operator aware of the machine's future behavior could choose to start collecting data at a moment when its variations would correspond to a chosen intentional order.

The operator has only a single choice of timing for each tripolar set; the second and third runs are started at twenty-minute intervals after the first. This means that, as with intentional selection, timing selection is a process that must be analyzed in terms of entire tripolar sets rather than individual runs.

Clearly, timing selection is potentially far more powerful than intentional selection. A perfectly efficient intentional selection process is limited in its abilities. The best it can do, by labeling each set optimally, is to put the high intention in the distribution generated by taking the highest of three independent standard normal deviates, and the low intention in the symmetric lowest-of-three distribution. In contrast, a perfectly efficient timing selector is limited only by the number of possible outcomes available for choice. Given a sufficiently broad "menu" of alternatives, a timing selector with perfect discrimination could create any output distribution desired for the three intentional categories.

However, with a single constraint, timing selection can be analyzed with the same tools used above — in fact, it makes exactly the same predictions as intentional selection, and therefore the same conclusions already reached will apply. The constraint is simply that an operator who is using timing selection to favor desirable results will tend only to choose a moment that produces results in line with the intentions, without further optimizing the outcome. In other words, if one assumes that an operator searching (perhaps subconsciously) for an auspicious time to begin the series is satisfied by finding some moment that gives mean shifts in the declared directions, rather than searching among a broad range of possibilities to find the very best, timing selection predicts the same relationships between rank frequencies and distribution parameters as intentional selection.

The reason for this should be clear. Consider a constrained timing selection process that is always successful at choosing the HBL order. By the constraint assumption, its outputs are unbiased selections from the distribution of tripolar sets that happen to be in the correct order. The output of the process, over

many samples, is equivalent to the algorithm: "Generate a tripolar set with attached intention labels. If it is in the correct order, keep it; if not, discard and try again." An intentional selection process that always succeeds is equivalent to the algorithm: "Generate a tripolar set. Label the lowest L, the highest H, and the third B." But both of these procedures will create exactly the same distributions. In each case the probability density of H runs is given (up to an overall normalization) by the joint probability that a normal deviate will take on a given value while two other independent normal deviates take on lower values. The probability distributions for L and B run values are likewise identical in both processes.

The constraint on timing selection may seem arbitrary, but in fact it is highly plausible for the dataset under consideration. It is obvious that the operators are not functioning anywhere near the regime of perfect efficiency. Table 3 shows that, if a selection process is operating, it achieves only a modest increase in the probability of correctly labeled sets. If this is the outcome of a timing selection process, it seems that operators are frequently wrong in their judgement that a given initiation time will produce results in the desired direction. Is it credible that they are managing to optimize their choices within the distribution of correctly-labeled sets, given that they are only marginally successful at identifying such sets at all? The solution of the Nelson problem discussed in the preceding section also supports the notion that operators do not seem to be optimizing their successful choices.

In short, the limited efficiency of rank frequency selection seen in Table 3, and the negative outcome of the Nelson test, strongly indicate that any timing selection process present must be operating in a regime where its effects are indistinguishable from intention-based selection. If this is so, then timing selection under these circumstances makes the same predictions as an intention selection model, and the  $p \approx 0.03$  rejection seen above applies to it as well. Timing selection together with intention selection spans the possible range of selection models for this experiment, so the result can with considerable confidence be applied to all selection models for this particular database. A more generalized timing selection model is not refuted, but must at a minimum include some explanation for the oddity that operators who are only modestly successful at choosing starting times that make the H run high and the L run low nonetheless are choosing their moments so cleverly as to spuriously mimic the statistical features of an influence model.

## X. Conclusions

We arrive, finally, at the conclusion that the selection hypothesis gives a poor fit to the data structure, while the influence hypothesis gives about as good a fit as can be expected. The result is statistically weaker than that reported in 1993:  $p \approx 0.03$  instead of  $p \approx 0.0095$  as in the previous analysis. Since it is known that the confound in the previous analysis was an inflation of its significance, it should not be surprising that this is the case.

We therefore may safely continue to conclude, albeit with less force, that the observed data do not support a run-based selection process for the apparent remote REG anomaly. Aside from its implications for theoretical modeling, this also reinforces the validity of the experiment, since failure of the experimental controls would manifest as just such a run-based selection effect.

### References

- Dobyns, Y. H. (1993). Selection versus influence in remote REG anomalies. *Journal of Scientific Exploration*, 7, 259.
- Dunne, B. J., and Jahn, R. G. (1992). Experiments in remote human/machine interaction. *Journal of Scientific Exploration*, 6, 311.