Journal of Scientific Exploration, Vol. 7, No. 3, pp. 259–269, 1993 0892-3310/93

Selection Versus Influence in Remote REG Anomalies

YORK H. DOBYNS

Princeton Engineering Anomalies Research, Princeton University, Princeton, NJ 08544

Abstract — A large body of remote human-machine interaction data has been collected in a protocol structurally similar to that used for experiments in remote perception, with somewhat comparable anomalous results. This suggests that the effects seen in the former could be attributable to a selection or sorting process on a reservoir of unperturbed data, rather than to any remote influence on the machine behavior per se. Fortunately, the statistical consequences of these two modalities are clearly distinguishable within the available empirical data. When properly evaluated by Bayesian hypothesis-comparison methods, the experimental results overwhelmingly favor the direct influence hypothesis over any selection mechanism.

Background

A substantial body of experiments in remote human/machine interaction involving various types of random event generator (REG) devices has displayed significant correlations between operator intentions and the mean values of the output distributions of the machines, despite considerable spatial separations between operator and machine (Dunne and Jahn, 1992). These results are not only consistent with those produced under local conditions, but also bear some similarity to those of precognitive remote perception (PRP) experiments, where individuals have demonstrated a statistically significant ability to acquire information about locations remote in distance and time (Dunne, Dobyns, and Intner, 1989).

From one point of view, the remote REG experiments could be regarded as a task in remote influence on the machine's behavior that automatically incorporates a number of controls that are unavoidably lacking in the local procedure. For instance, the physical absence of the operator from the laboratory effectively eliminates any possibilities for operator deception, while the possibility of experimenter influence is precluded by the experimenters remaining blind to the operators' intentions until after the data are generated and recorded. From another perspective, the future outcome of a remote REG session, as recorded in a logbook and in a computer file, could be regarded as a form of simple PRP target in which the operator attempts to identify the highest and lowest of the three recorded mean values. This alternative hypothesis becomes more plausible in view of certain dissimilarities of results using different REG noise sources. While electronic diode sources display statistically significant anomalous yields in both local and remote studies, the various pseudo-random sources show significant yields only in the remote experiments. If the remote interaction is driven by a remote perception effect, to which the nature of the noise source would be irrelevant, the presence of anomalous yield from both sources would be a natural result. (The contrast between remote and local performance on the pseudo-random experiments would then suggest different mechanisms for those effects, but that issue is beyond the scope of this paper.)

Invoking such a selection model may seem simply to be handwaving to explain one anomaly with another. However, these two models make sufficiently different statistical predictions that they can be distinguished with some confidence even on a rather modest dataset. It should be noted that this analysis is directed only peripherally at the issue of the existence of a real effect. The central question addressed is not the credibility of the effect per se, but rather the consistency of the effect with a particular class of mechanisms, which must be the first step in the production of useful theoretical models.

Selection versus Influence

For purposes of this discussion, the *selection* model denotes the hypothesis that the operator's efforts do not in any way affect the operation of the device, but that the operator is able by unspecified means to assign the three intentions of each experimental session in accordance with the actual values that will subsequently be produced. The *influence* model proposes in contrast that the behavior of the device is actually different under the various experimental intentions. In both cases the assumed target of the experiment is the mean of the experimental data, and the anomalous result is the deviation of that mean from null-hypothesis expectation. As shown in Fig. 1, the salient difference in the two models lies in the relationship between the mean shifts in the data and the relative rankings of the intentional runs. Under the selection model, the behavior of the device is undisturbed and any alteration in the mean values of, for example, high runs must derive from their preferential origin as the highest of three randomly generated values. In the influence model, the differential ranking of the high intention relative to the others is driven by, rather than driving, the underlying difference in distributions.

Statistics of the Selection Model

The theoretical distribution of run-terminal scores is essentially normal, and all following discussion will assume the scores have been reduced to standard normal deviates. Since the selection model assumes the effect to be driven by the assignment of intentions to the three runs of each tripolar set, the crucial issue is thus the distribution of the highest, lowest, and midmost of a set of three normally distributed values. The probability density for x drawn from a standard normal distribution, conditional upon x being the highest of 3 such draws, is:

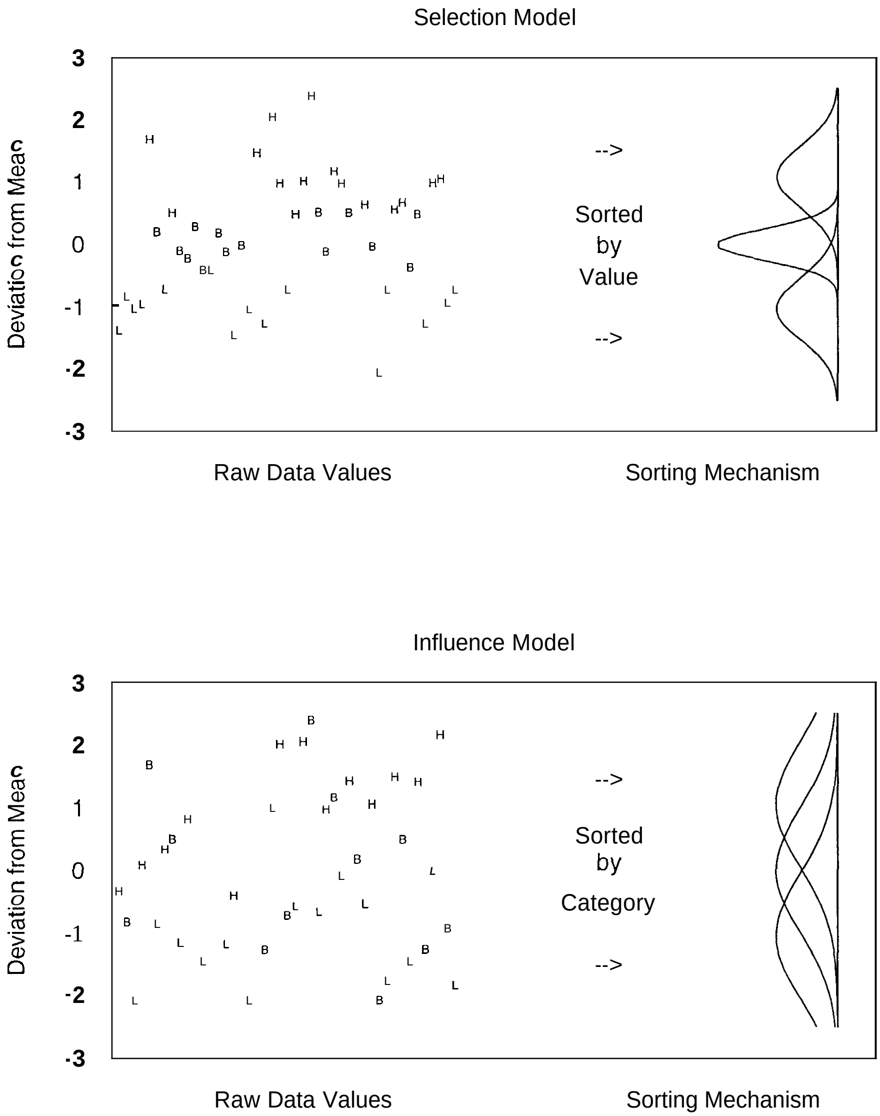

Fig. 1. Exaggerated Illustration of Differences in Models.

$$p(x) = f(x)F(x)^2 / \int_{-\infty}^{+\infty} f(x)F(x)^2\,dx, \tag{1}$$

where

$$f(x) = \left(1/\sqrt{2\pi}\right)e^{-x^2/2} \qquad F(x) = \int_{-\infty}^{x} f(t)dt$$

are the probability density and cumulative probability distribution of the standard normal function. The numerator of (1) is the joint probability of a value x appearing and of two lower values being drawn from the same distribution; the denominator simply normalizes the expression so that the total integral of p is 1. Integrating the moments of (1) numerically, one finds that the highest-of-three distribution has the mean μ_H = 0.8463 and the standard deviation σ_H = 0.7480. The lowest-of-three distribution by symmetry has the same standard deviation and opposite mean, and the middle-of-three has $\mu_M = 0$, $\sigma_M = 0.6698$. If these three distributions are recombined in equal proportions, one of course retrieves the parent standard normal distribution. (This is equivalent to saying that if you are required to label one of three numbers from a standard normal distribution as "high", but have equal probabilities of choosing the highest, lowest, or midmost of the three for the label, over many runs your choices of "high" will themselves follow a standard normal distribution.)

The essence of the selection model is the supposition that the operator is able to choose the highest of the three runs to be labeled as "high," and/or the lowest to be labeled as "low," with some probability greater than the chance 113. For the moment, consider only the operator's declared high intention, which actually turns out to be the lowest run of the three a fraction p_L of the time, the highest run a fraction p_H of the time, and on the remaining occasions, $p_M = 1 - (p_H + p_L)$, the middle run. Given the three distributions just described, it follows that the distribution of high runs has the mean and variance:

$$\mu = p_L\,\mu_L + p_M\,\mu_M + p_H\,\mu_H = \mu_H(p_H - p_L),$$

$$\sigma^2 = \sum_{x \in L, M, H} p_x\left(\mu_x^2 + \sigma_x^2\right) - \mu^2 = \sigma_M^2 + \left(\mu_H^2 + \sigma_H^2 - \sigma_M^2\right)\left(p_H + p_L\right) - \mu^2. \tag{2}$$

These equations may be inverted to give the selection frequencies needed to produce a given mean and variance:

$$p_L = \frac{1}{2}\left(\frac{\sigma^2 + \mu^2 - \sigma_M^2}{\sigma_H^2 + \mu_H^2 - \sigma_M^2} - \frac{\mu}{\mu_H}\right);$$

$$p_H = p_L + \frac{\mu}{\mu_H};$$

$$p_M = 1 - \left(p_H + p_L\right). \tag{3}$$

Obviously, similar relations can be derived for the low and baseline efforts. It should be noted, however, that all of these equations will producep's that are valid probabilities only for a limited range of μ and σ. No possible selection scheme can produce $\mu > \mu_H$, for example. Figure 2 illustrates the region in (μ, σ) space that can be potentially accomodated by a selection model.

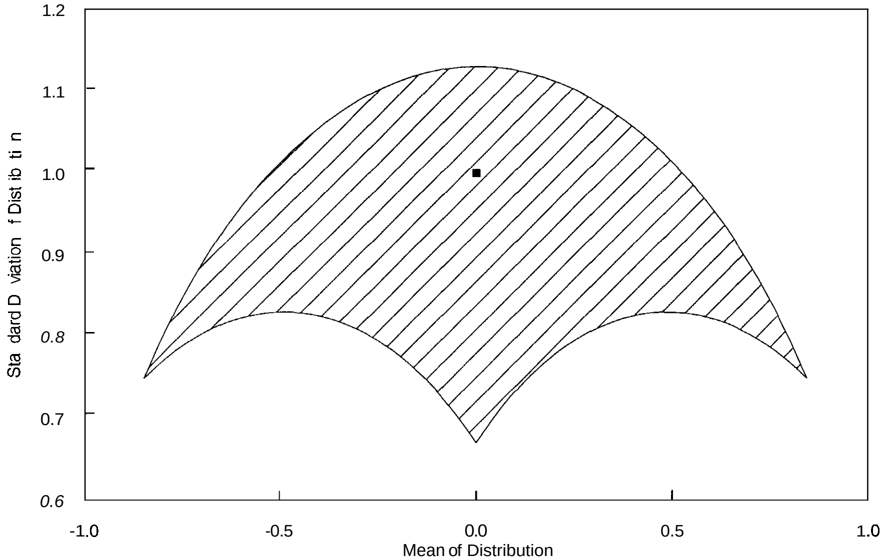

Fig. 2. Region of Possible Parameter Values for Selection Model.

Statistics of the Influence Model

The prior section showed how the rankfrequencies p_H, p„ p_M for any intention can be derived, under the selection model, from the mean and variance of the scores in that intention. Rank frequencies can likewise be derived for the influence model. If presented with three datasets whose empirical mean and variance estimates are m_1, m_2, m,, s_1^2, s_2^2, s_3^2 respectively, the frequencies with which a value drawn from distribution 1 will be highest or lowest are:

$$p_{1H} = \int_{-\infty}^{+\infty} F\left(\frac{x-m_2}{s_2}\right) F\left(\frac{x-m_3}{s_3}\right) \frac{1}{s_1} f\left(\frac{x-m_1}{s_1}\right) dx,$$

$$p_{1L} = \int_{-\infty}^{+\infty} \left[1-F\left(\frac{x-m_2}{s_2}\right)\right]\left[1-F\left(\frac{x-m_3}{s_3}\right)\right] \frac{1}{s_1} f\left(\frac{x-m_1}{s_1}\right) dx,$$

$$p_{1M} = 1 - \left(p_{1H} + p_{1L}\right). \tag{4}$$

Consistency Criteria

There are two constraints on the values of the rank frequencies. The first, that every run (and therefore every run with a particular intentional label) must be assigned some rank, is automatically satisfied by the formulae given above; it does, however, mean that only two of the three rank frequencies for any intention need be calculated. (This ignores the possibility of ties. Ties are present but infrequent in the actual data, and are most easily dealt with simply by discarding that relatively small number of tripolar sets.) The second is that every run, whatever its rank, must be assigned an intention, and therefore the proportions of (for example) highest-ranked runs assigned to high, low, and baseline intentions must also add up to one. This appears to present three more constraint equations on the rank frequencies, but in fact one of them is redundant. Since this places a total of five constraint conditions on the nine rank frequencies, determining four of the rank frequencies will suffice to determine the entire system. The calculations detailed below fix the matrix of rank frequencies by computing the four values R_{HH} (probability with which the high intention has the highest rank), R_{HL} (probability that the high intention has the lowest rank), R_{LH} and R_{LL} as the key values that determine the matrix.

A final consistency criterion applies to the normalization of the data. Since the selection model assumes the standard normal distribution for the source, the consistency constraints on the rank frequencies actually impose a condition on the means and standard deviations of the three intentions such that, when they are recombined, the composite distribution must have exactly $\mu = 0$, $\sigma = 1$. Therefore, in order to make valid rank frequency calculations in the selection model, the data must be normalized to their own, overall statistics across the three intentions (rather than to the theoretical performance of the machine). The influence model does not impose such a condition, since it will make exactly the same predictions for relative rankings regardless of what normalization is used for the raw data.

Statistics of the Actual Data

The remote REG database consists of 494 tripolar sets, of which 398 were generated using a microelectronic noise diode source and 96 using some form of pseudo-random device. All systems were extensively calibrated and failed to deviate significantly from the theoretical distribution in those calibration runs. Four of the experimental sets, all generated with the diode device, contain ties and are excluded; when the remaining 490 sets are normalized to their own collective mean and standard deviation, the results take the following form:

Int.	Mean	Std. Dev.	N Highest	N Middle	N Lowest
H	0.08825	1.03035	180	167	143
B	-0.03619	0.97074	159	156	175
L	-0.05207	0.99413	151	167	172

Interpreting the Observation

Using the relations developed earlier, we can construct predicted rank frequencies from the means and standard deviations of the observed data, for either the selection model or the influence model, for comparison with the observed rank frequencies. Since these predictions are quite different, the observation should provide better support for one hypothesis than the other.

The most straightforward way of assessing such evidential support is by an elementary application of Bayes' theorem of conditional probabilities. This asserts that the relative support an observation gives to two competing hypotheses is simply the ratio of the probabilities or "likelihoods" of that observation under each hypothesis. For each intention we are confronted with, in essence, a three-way choice experiment conducted n times. The three possible outcomes are observed n_1, n_2, n_3 times respectively ($n_1 + n_2 + n_3 = n$). The probability of this observation under a hypothesis p_1, p_2, p_3 concerning the probabilities of the individual outcomes is

$$P\left[n_1, n_2, n_3 \middle| p_1, p_2, p_3\right] = \frac{n!}{n_1!n_2!n_3!} p_1^{n_1}, p_2^{n_2}, p_3^{n_3} . \tag{5}$$

The combinatorial factor depends only on the observations, and therefore cancels out in the probability ratio between different hypotheses on the same observation. If the two hypotheses predict (p_1, p_2, p_3) and (q_1, q_2, q_3) respectively, the ratio of relative support is

$$B(p/q) = \frac{p_1^{n_1} p_2^{n_2} p_3^{n_3}}{q_1^{n_1} q_2^{n_2} q_3^{n_3}} = \left(\frac{p_1}{q_1}\right)^{n_1} \left(\frac{p_2}{q_2}\right)^{n_2} \left(\frac{p_3}{q_3}\right)^{n_3} . \tag{6}$$

The aggregate likelihood of the hypothesis over all three intentions may be calculated by repeating the individual likelihood calculation for each intention, and the total likelihood ratio will simply be the product of factors such as (6) above for each of the three intentions. The above discussion could as easily have been framed in terms of the assignment of intentions to ranks, rather than of ranks to intentions; this amounts to considering the columns rather than the rows of the 3 x 3 rank frequency matrix first, and leads to the same final result for the overall odds ratio.

A Complication

The foregoing discussion makes a tacit assumption that in calculating the predictions of each hypothesis, we know the exact distribution statistics of the three intentions' run scores, when in fact the observation only estimates these parameters. Because the selection model's predictions in particular are strongly dependent on these distribution parameters, our calculation of relative support will be erroneous if we pretend that the observed values are exact. This may be illustrated by considering an extreme example. Suppose that the selection model is in fact true and that, moreover, the highest run in a tripolar set is being assigned to the high intention with 100% efficiency. Suppose further that, due to statistical fluctuations, the actual mean of the high dataset is slightly higher than its theoretical population value $\mu_H = 0.85$, as will happen by chance about half the time. The selection model cannot accomodate so high a value at all, and thus has a likelihood of exactly zero regardless of the observed rank frequencies, if the observed value is taken as an exact population value. Since this model is true by hypothesis, the calculation must itself be invalid, and the error lies in taking the measured value as exact.

Therefore, to complete the application of Bayesian hypothesis comparison, we must consider that the observation on the mean and standard deviation of each intention defines a probability distribution on the population values giving rise to the measurement. Calculating the aggregate likelihood of each hypothesis thus becomes slightly more involved. The previous section showed how to calculate a likelihood from a set of rank frequency observations and theoretical predictions. Earlier sections showed how to predict rank frequencies from distribution statistics for both selection and influence models. Taken together, these results give us a functional mapping from distribution statistics onto likelihoods for each model; that is, either the selection model or the influence model can proceed from a set of means and standard deviations, and the associated rank frequency observations, to a final likelihood.

To implement this approach, let $L(\mu_h, \mu_l, \sigma_h, \sigma_l)$ be the function that describes the likelihood of a set of distribution parameters given the observed rank frequencies. (L is of course a different function for the two hypotheses.) The actual distribution parameters are known only probabilistically from the observations. The observed value m_h for the mean of the high distribution, for example, produces a normal likelihood distribution $p(\mu_h|m_h)$ for the actual value of the mean. Therefore, the overall likelihood of the hypothesis, given both the observed means and standard deviations and the observed rank frequencies, is

$$P \propto \int\int\int\int d\mu_h d\mu_l d\sigma_h d\sigma_l \, p(\mu_h|m_h) p(\mu_l|m_l) p(\sigma_h|s_h) p(\sigma_l|s_l) L(\mu_h, \mu_l, \sigma_h, \sigma_l). \tag{7}$$

Evaluating this four-dimensional integral is somewhat tedious, but amenable to standard numeric quadrature techniques. Formula (7) is expressed as a proportionality rather than an equation because we have not troubled with normalizing the assorted probabilities and likelihoods appropriately. Since we will be calculating this quantity twice, using the two different functions L required by the two models, and taking the ratio, overall normalization is unimportant so long as the same normalization is used for each hypothesis.

Analysis Results and Interpretations

In summary, the procedure for comparing the experimental support for the selection model versus the influence model involves three steps:

1. Compute the probability distributions for each of the model parameters μ_h, μ_l, σ_h, σ_l that follow from the corresponding observed values m_h, m_l, s_h, s_l.
2. Calculate for each model the weighted integrals (7) of the likelihoods of the observed rank frequencies over the range of possible model parameters.

3. Take the ratio of the likelihoods to determine the odds adjustment factor B between the two hypotheses that follows from the observation.

When this procedure is applied to the values given in the section "Statistics of the Actual Data," one finds that the odds ratio is 28.9 to 1 in favor of the influence model. In other words, given only the observation and no prior information concerning the relative plausibility of the two models, one would conclude that the influence model is about 29 times more likely as an explanation of the facts than the selection model. This factor is also the numerical adjustment that would be applied to a preexisting odds ratio between the two hypotheses, e.g. from prior experiments or theoretical considerations, in computing the effect of the current experiment on the overall relative credibility of the hypotheses.

It should be noted that while this sort of odds ratio is not directly comparable to a traditional p-value, it is clearly imposing. Consider a simple, one-parameter test against a null hypothesis that the parameter has a certain value, in which the possible measurement error is normally distributed. If a Bayesian hypothesis test returns odds of 28.9 in favor of the null, it is easy to show that the classical p-value must be no larger than $p = 0.0095$, and may well be considerably smaller. (This best case assumes that the alternate hypothesis is the "maximum likelihood hypothesis," namely that the observed value is exactly the actual value. Any other alternative hypothesis, to be favored over the null by a factor of 28.9, requires an observation such that the classical p-value is considerably less than 0.0095.)

As a check on the validity of the analysis, the procedure was repeated with two synthetic datasets, wherein the source of deviations between intentions was known by construction. In the synthetic "influence" data, a uniform mean shift was applied to the high intention, and an opposite one to the low intention, of magnitude similar to that seen in the actual data. The analysis on the synthetic data returned an odds ratio of 31.4 in favor of the influence model. The second dataset was constructed by selection, that is, by generating random tripolar datasets and preferentially choosing those that were in the "right" rank order between intentions, again in such a way as to approximate the effect size seen in the actual data. The odds calculation procedure on these data produced odds of 0.00018 favoring influence, or in other words of 5500 to 1 in favor of selection. The test calculation thus appears to discriminate effectively between the two idealized mechanisms. Equally important, the quantitative scale of the odds ratio for the actual data is almost as large as for the ideal "influence" case, indicating that the evidence in favor of the influence model is about as strong as one could reasonably expect, given the size of effect and the amount of available data.

Further Considerations

It was noted in the introduction that the statistical test here used does not address the evidence for the existence of any effect, versus the null hypothesis that all differences between intentional groupings are due to statistical fluctuation. That issue has been thoroughly addressed elsewhere (Dunne & Jahn 1992). In fact, applying the test to datasets that, by construction, contain no effect, yields strong odds (ranging, in a modest Monte Carlo database, from 8.5 to over 100) in favor of the influence model. It is easy enough to see why this should be so. In the completely randomized Monte Carlo calculation, the differences between datasets are due to random fluctuations and therefore display the same rank-frequency character as real population differences. Since selection is not taking place, the selection model is strongly refuted, and in the two-hypothesis test this manifests as evidence in favor of influence.

It might be suspected that the exclusion of the tied datasets is in some way prejudicial. This has been compensated for in two ways:

(1) In the construction of the artificial datasets above ties were discarded in exactly the same manner; and
(2) an earlier analysis, in which ties were less awkward because consistency criteria were not taken into account, found that including the tied runs changed the final odds ratio by only about 10%.

As noted above, the 490 (non-tied) tripolar runs were produced by two different types of noise source, one of which is only pseudo-random. Selection models are superficially more appealing than influence models for the pseudo-random data, since it is difficult to imagine how a deterministic string of pseudo-random values might be "changed." Thus, one might expect that in the pseudo-random data, at least, the observation would favor the selection model. However, this proves not to be the case. If the data are separated into subsets according to whether their source was a noise diode or a pseudo-random mechanism, we find:

(1) The diode source data produce odds of 25.6 to one in favor of influence. The slight reduction in strength of evidence is entirely attributable to the reduced amount of data.
(2) The pseudo-random data produce odds of 6.3 to one in favor of influence. Here the dataset is much smaller, so that strong statistical conclusions are not to be expected; nonetheless the evidence seems to show modest preference for the influence model.

Conclusions

The analysis described above thus allows some general conclusions:

1. Despite its conceptual appeal, the selection model is a considerably poorer predictor of the data structure than the influence model.
2. This remains true for either type of noise source, despite the fact that the selection model might seem more natural for the pseudo-random source than an influence model.
3. Since the selection model addresses phenomenology rather than mechanisms, evidence against it is evidence against any predictive, perceptual,

or data-sorting model of remote human-machine anomalies. It is also evidence against deception or failure of controls in the remote protocol; if the effect were due to accidental or deliberate cuing of operators, or to any other protocol failure that allowed them knowledge of the run results before their intentions were recorded, the output would necessarily show the statistical character of a selection model.

We thus conclude that whatever the operators are doing, they are not simply sorting the undisturbed output of the device into intentional "bins." Nor are they, by accident, chicanery, or anomalous means, finding out the values of existing runs and choosing the intentions to suit. Rather, in so far as the evidence allows us to judge, the mean output level of the device is genuinely, albeit slightly, different from one operator intention to another. Any explanatory model for the apparent anomaly must take this fact into account as its starting point.

Acknowledgements

The author wishes to thank Robert Jahn, Brenda Dunne, and Roger Nelson for advice regarding the presentation of this paper, which is a revised and expanded treatment of a presentation to the SSE's June 1992 Annual Meeting. The author is also indebted to Peter Sturrock and Helmut Schmidt for their comments at that meeting. The Engineering Anomalies Research Program is supported in part by major grants from the McDonnell Foundation, the Fetzer Institute, and Laurance S. Rockefeller.

Correspondence should be addressed to York Dobyns, Princeton Engineering Anomalies Research, C-131 Engineering Quadrangle, Princeton University, Princeton, New Jersey 08544-5263.

References

Dunne, B.J., Dobyns, Y.H., and Intner, S.M. (1989) ***Precognitive Remote Perception III: Complete Binary Data Base with Analytical Refinements.*** Princeton University: PEAR Technical Note.

Dunne, B.J., and Jahn, R.G. (1992) Experiments in Remote Human/Machine Interaction. ***Journal of Scientific Exploration.*** 6(4), pp. 311-332, 1992.