Count Population Profiles in Engineering Anomalies Experiments

ROBERT G. JAHN, YORK H. DOBYNS, and BRENDA J. DUNNE

Princeton Engineering Anomalies Research, Princeton University

Abstract—Four technically and conceptually distinct experiments—a random binary generator driven by a microelectronic noise diode; a deterministic pseudorandom generator; a large-scale random mechanical cascade; and a digitized remote perception protocol—display strikingly similar patterns of count deviations from their corresponding chance distributions. Specifically, each conforms to a statistical linear regression of the form $\Delta n/n = \delta(x - \mu)$, where $\Delta n/n$ is the deviation from chance expectation of the population frequency of the score value x divided by its chance frequency, μ is the mean of the chance distribution, and δ is the slope of the regression line, constant for a given data subset, but parametrically dependent on the experimental device, the particular operator or data concatenation, and the prevailing secondary conditions. In each case, the result is tantamount to a simple marginal transposition of the appropriate chance Gaussian distribution to a new mean value $\mu' = \mu + N\epsilon$, where N is the sample size, or equivalently to a change in the elemental probability of the basic binary process to $p' = p + \epsilon$, where p is the chance value and $\epsilon = \delta/4$. Proposition of a common psychophysical mechanism by which the consciousness of the operator may achieve these elemental probability shifts is thwarted by the complexity and disparity of the several technical and logical tasks that would be involved. More parsimonious, albeit more radical, explication may be posed via a holistic information-theoretic approach, wherein the consciousness adds some increment of information, in the technical sense, into the particular experimental system, which then deploys it in the most efficient fashion to achieve the experimental goal, i.e., the volition-correlated mean shift. The relationship of this technical information transfer to the subjective teleological processes of the consciousness remains to be understood.

Introduction

Over the past twelve years, the Princeton Engineering Anomalies Research (PEAR) program has accumulated very large data bases on a number of human/machine experiments, including a variety of microelectronic random and pseudorandom binary generators, a macroscopic random mechanical cascade, and an assortment of analogue experiments using optical, mechanical, and fluid dynamical devices. For the most part, attention has been focused on anomalous shifts of the mean values of the statistical output distributions of these machines, compared to their calibration behavior and/

or theoretical expectations, in correlation with prerecorded intentions of their human operators. Full descriptions of the specific equipment utilized, its qualification and calibration, the experimental protocols, safeguards against spurious artifacts, data collection and processing techniques, the detailed results, and their interpretation are available in many references (Jahn, Dunne, & Nelson, 1987; Dunne, Nelson, & Jahn, 1988; Jahn & Dunne, 1987; Nelson, Dunne, & Jahn, 1984, 1988a). In briefest summary, the following hierarchy of effects has been established:

1. Systematic anomalous deviations of the output distribution means of such devices can be replicably achieved by a large number of common human operators.

2. The scale of the anomalous effects are invariably quite small—tantamount to a few bits per thousand displacement from chance expectation in the binary machines—but over large data bases can compound to highly significant statistical deviations.

3. The primary correlate of these deviations is the pre-stated intention of the operator. Randomly interspersed accumulations of data taken under three states of intention—to get higher values; to get lower values; to maintain the chance value—show clear and systematic separations from one another and from the chance expectation.

4. The form and scale of these "tripolar" separations are identifiably operator-specific. Some operators achieve in both high and low intentions; some in just one; some in neither; some regularly obtain shifts inverse to their intentions; some even produce anomalous data under the null intention. But these results are sufficiently characteristic of the individuals that they may reasonably be termed operator "signatures."

5. The dependence of the results on the particular machines employed is less clear. Some operators seem to transfer their signatures across two or more devices, others show substantial differences from one device to another, but in most cases the scales of achievement are preserved.

6. Sensitivity of the effects to a variety of secondary technical parameters, such as length or pace of the experiment, form of feedback, whether the operator chooses the direction of effort or is instructed by a random selector, etc., is also found to be quite operator-specific. Some operators' performances are profoundly affected by such options; others seem oblivious to them.

7. Two operators attempting to influence the machine in a cooperative fashion do not simply compound their signatures; rather, their joint results tend to be distinctively characteristic of the particular pair.

8. These effects can be obtained by operators widely displaced from the machines—up to distances of several thousand miles. In a number of cases, the characteristics of a given operator's "remote" data closely concur with his "proximate" data.

9. In some cases, remote effects can also be obtained when the operator's

effort is substantially displaced in time from the actual operation of the machine.

10. Although little systematic correlation with psychological or physiological indicators has been attempted in this laboratory, there appears to be little dependence of individual operator achievement on personality, style, or strategy. All of the over 100 operators so far employed in these studies have been anonymous and uncompensated, and none has claimed extraordinary abilities before or after the experiments. When their individual effect sizes are arrayed in ascending order, the distribution covers a continuous range only slightly displaced from chance (Dunne, Nelson, Dobyns, & Jahn, 1988).

Although most of the data treatment and interpretation of these experiments has heretofore concentrated on the anomalous deviations of the output means, some incidental attention has also been paid to the behavior of higher moments of the experimental distributions—variance, skew, kurtosis, goodness-of-fit, and other more elaborate statistical criteria—in the hope of illuminating some aspects of the physical, psychological, and epistemological mechanisms of these phenomena. The ultimate assessment of this type, however, must entail systematic examination of the complete distribution profiles, at whatever incremental scale the particular experimental readout permits. For our ensemble of random event generators, for example, based as they are on binary combinatorials, the most natural incremental units are simply the integer count reports that distribute about the sample mean, e.g., for 200-sample binaries, the populations of counts . . . 98, 99, 100, 101, 102 . . . , etc. From such detailed distributions one may then determine whether the shift of the mean is driven by an excess or deficiency of nearby counts, counts in the tails, counts near the standard deviation, or by some regularly distributed pattern of differences or by a totally random array. One may also inquire whether individual operators invoke identifiably different count patterns for their achievements, and how replicably so, thereby supplementing their cumulative deviation mean signatures in characterizing their individual effects.

Unfortunately, given the highly stochastic nature of the process, the marginal scale of the anomalous effects, and the number of parameters that can influence operator performance, monumental amounts of data must be accumulated before any credible systematic trends in the individual count populations can be identified, and it is only relatively recently that our total data base has reached adequate dimensions to support such an effort. To date, the microelectronic random event generator (REG) ensemble of experiments has compounded to some six million trials, encompassing 108 operators, three machines, and three major protocol variations. From this reservoir, a number of sufficiently large and self-consistent data subsets may now be extracted, including some individual sets for a few of our more prolific operators, that can reasonably be submitted to this count distribution analysis.

Count Population Profiles in REG Experiments

The theoretically expected count distribution for any random binary generator trial follows from the Bernoulli binomial combinatorial

$$n(x) = \binom{N}{x} p^x (1-p)^{N-x} \tag{1}$$

where $n(x)$ is the frequency of count x, N is the number of binary samples per trial, and p is the elemental binary probability of each sample. For all of the experiments described here, $N = 200$, and all machines are designed to provide $p = 0.5$. The corresponding binomial distribution is very well approximated by the Gaussian function

$$n(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-1/2(x-\mu/\sigma)^2} \tag{2}$$

with mean $\mu = Np = 100$ and standard deviation $\sigma = [Np(1-p)]^{1/2} = 7.0707. \ldots$ All calibration data conform appropriately to these theoretical expectations, within an empirical uncertainty on p of 5×10^{-5} (Nelson, Bradish, & Dobyns, 1989).

For virtually all composite or individual operator data subsets for which statistically significant anomalous deviations of the mean are found, the corresponding count distributions appear, well within the inherent experimental noise, to retain essentially Gaussian profiles, albeit centered about a displaced mean $\mu' = 100 + \Delta\mu$, where $\Delta\mu$ is characteristic of the particular subset, the intended direction of effort, and an assortment of prevailing secondary parameters. The essential question we wish to examine here is whether the detailed forms of these translated distributions can provide any insight into the basic nature of the anomalous phenomenon.

Given the tiny scale of $\Delta\mu$, it is best to deal in terms of the *differential* count frequencies between the displaced distributions $n'(x)$ achieved by the operators, and the chance distribution, $n(x)$ i.e., $\Delta n(x) = n'(x) - n(x)$. Figure 1 shows sets of such $\Delta n(x)$ profiles for an overall REG data base of some 3.8 million trials performed by 92 operators over a period of eleven years under the tripolar intentions of high, low, and null, compared to an appropriate set of calibration data. In this case, the null intention and calibration deviations are found to conform very well to statistical chance expectation, as evidenced by the essentially random alternation of positive and negative values, and by the small number that exceed the two-tailed .05 confidence envelopes sketched on the graphs. The high and low intention data, in contrast, show clear distributions of imbalance consonant with the total mean shifts for these subsets. From these displays it is evident that the mean shifts are being broadly supported by a preponderance of the individual count deviations, rather than by more localized excesses or deficiencies in the tails or elsewhere.
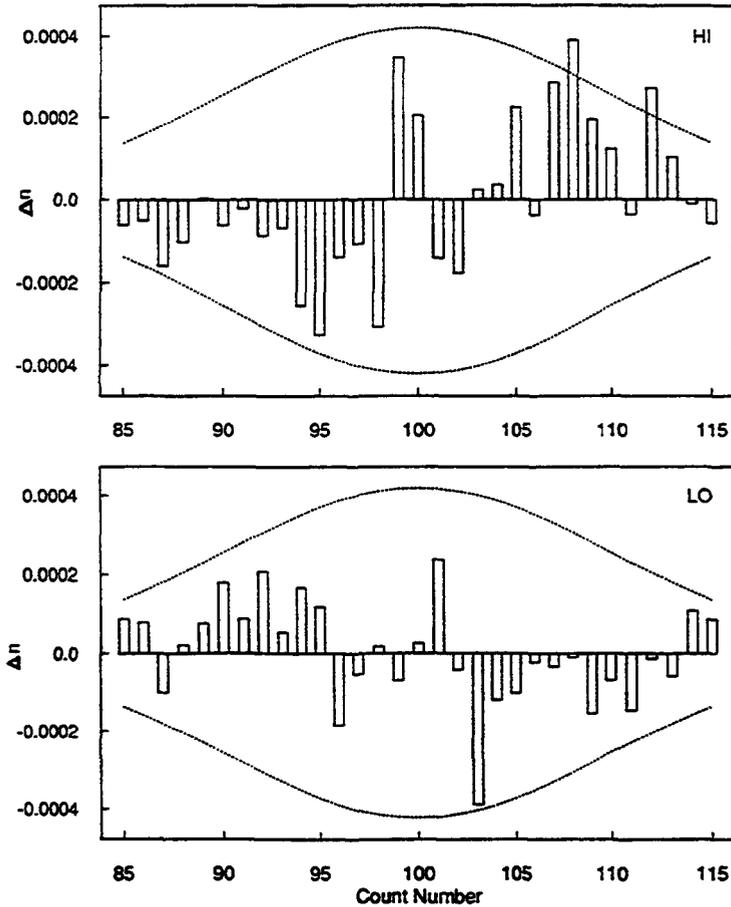
Fig. 1a. Count frequency deviations. All diode REG: high, low.

Figure 2 shows the patterns of count frequency deviations for a single operator's contribution to this same data set. Although these distributions are considerably noisier because of the smaller data base size, similar first conclusions may be drawn. From a full array of such graphs for all individual operator contributions to this and all other REG data subsets, it may be quite generally concluded that significant anomalous displacements of the mean are almost invariably broadly distributed across the corresponding count distributions, rather than appearing as more localized departures from chance expectations.

Temptations to search for yet more subtle features in these differential count patterns are largely deterred by the inherent statistical noise which even for these very large data sets tends to obscure any secondary regulari-
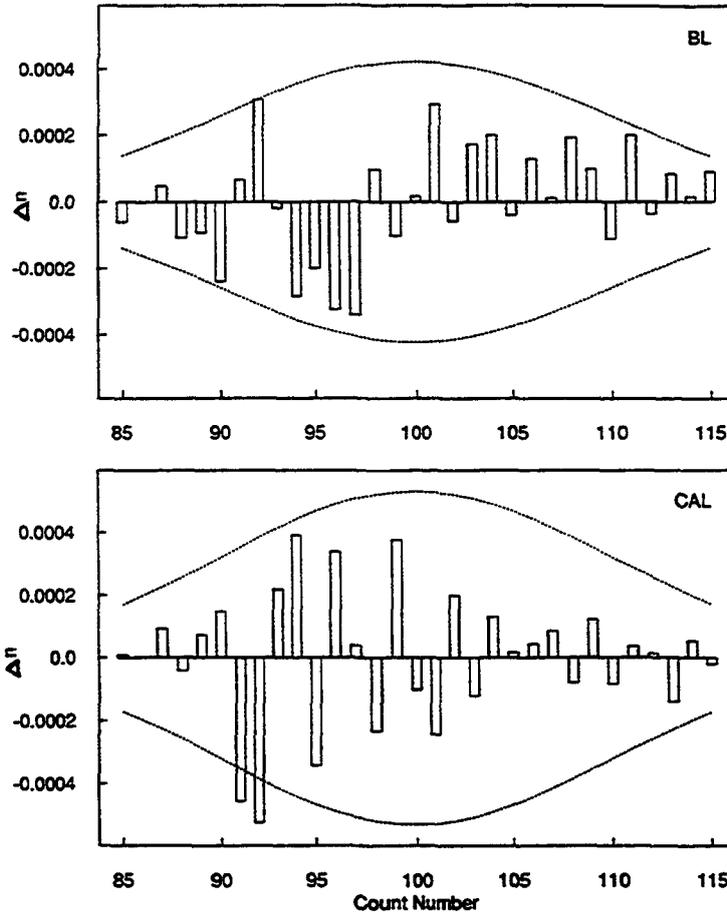
Fig. 1b. Count frequency deviations. All diode REG: baseline, calibration.

ties. It is possible, however, to take another analytical step that may help to illuminate the basic nature of the process underlying these anomalies. Namely, if the count data are re-plotted as proportional deviations, i.e., as the differences between the operator-generated distributions and the chance distribution, normalized by the latter, ($\Delta n/n$ values in x,) a particularly simple functional relationship tends to emerge in most cases. The data of Figure 1 are regraphed in this form in Figure 3. (Counts below 85 and above 115 have now been excluded because their populations are too small to provide stable estimates.) Despite their large stochastic swings, each of these patterns can be statistically well fit by a linear regression of the form

$$\frac{\Delta n}{n} = \delta(x - \mu) \tag{3}$$

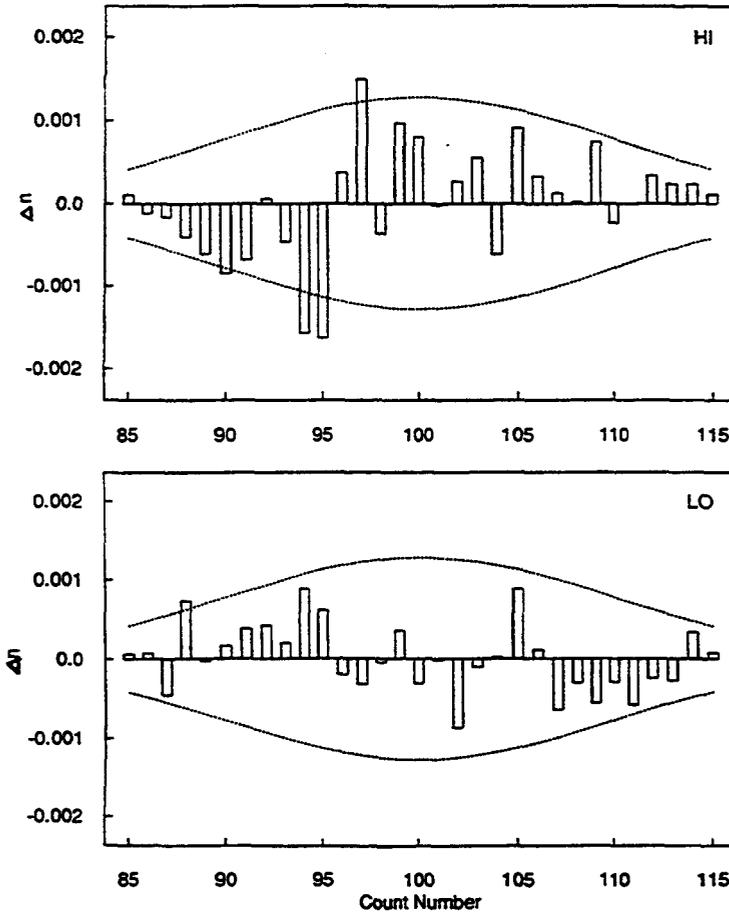Fig. 2. Count frequency deviations. Diode REG: Operator 010.

i.e., by a straight line through the mean count, with characteristic slope δ. (In these and subsequent figures, Z_1 is the z-score denoting the significance of the linear term in the regression; Z_2 denotes the improvement that would be made by retaining a quadratic term. Other details of the regression calculation are presented in the Appendix.) As shown in Figure 4, this form also obtains at the level of the individual operator data of Figure 2, and has been confirmed over many other composite and individual data sets not shown here. In virtually all cases, the linear fit is statistically sufficient; the quadratic, cubic, and higher terms are usually not needed.

This striking result is most parsimoniously consistent with the hypothesis that the displaced distribution is still Gaussian, but with its elemental binary probability, $p = 0.5$, somehow altered to a new value, $p' = 0.5 + \epsilon$, where the size of ϵ is characteristic of the subset or individual operator. This can be seen
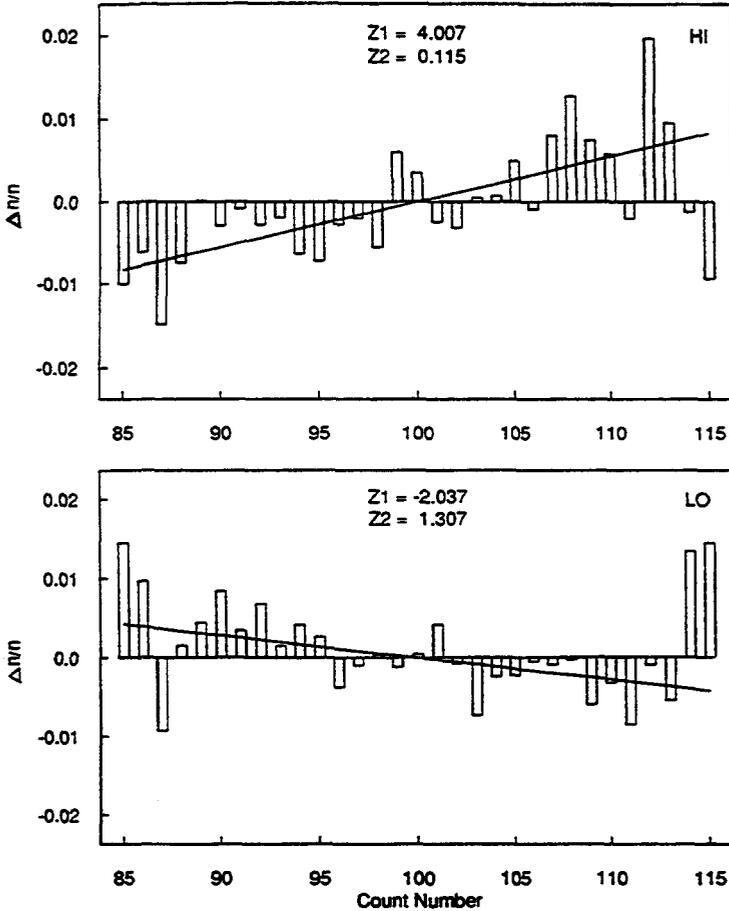
Fig. 3a. Proportional frequency deviations. All diode REG: high, low.

most directly by partial differentiation of the Gaussian function with respect to the mean value, μ, holding x constant as parameter:

$$\frac{\Delta n}{n} = \frac{1}{n}\frac{\partial n}{\partial \mu}\Delta\mu = \left(\frac{x - \mu}{\sigma^2}\right)\Delta\mu \tag{4}$$

Inserting

$$\Delta\mu = \mu' - \mu = N(p + \epsilon) - Np = 200\epsilon \tag{5}$$

and

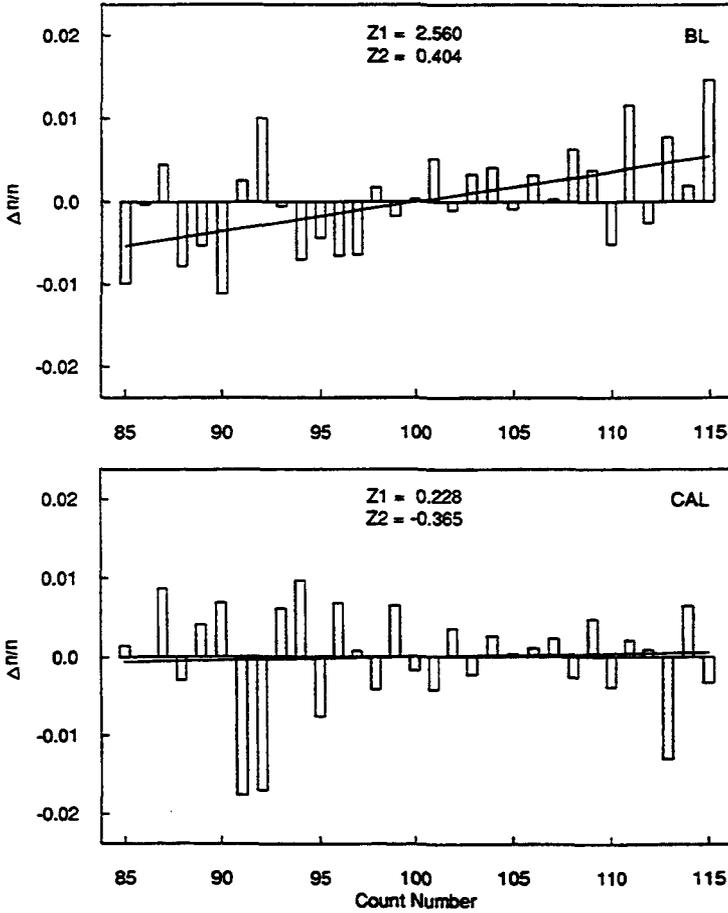$$\sigma^2 = Np(1 - p) = 50 \tag{6}$$

Fig. 3b. Proportional frequency deviations. All diode REG: baseline, calibration.

yields

$$\frac{\Delta n}{n} = \delta(x - \mu) = 4\epsilon(x - 100) \tag{7}$$

In other words, a single incremental distortion of the elemental binary probability, commensurate with the overall mean shift of the Gaussian, manifests as the observed linear displacements of $\Delta n/n$.

This does not, of course, imply that each count deviation contributes equally to the overall shift of the distribution mean. For this assessment one requires the first moment of the deviations about the mean, which we might term the leverage distribution, $L(x) = \Delta n(x - \mu)$. Empirically, this distribu-
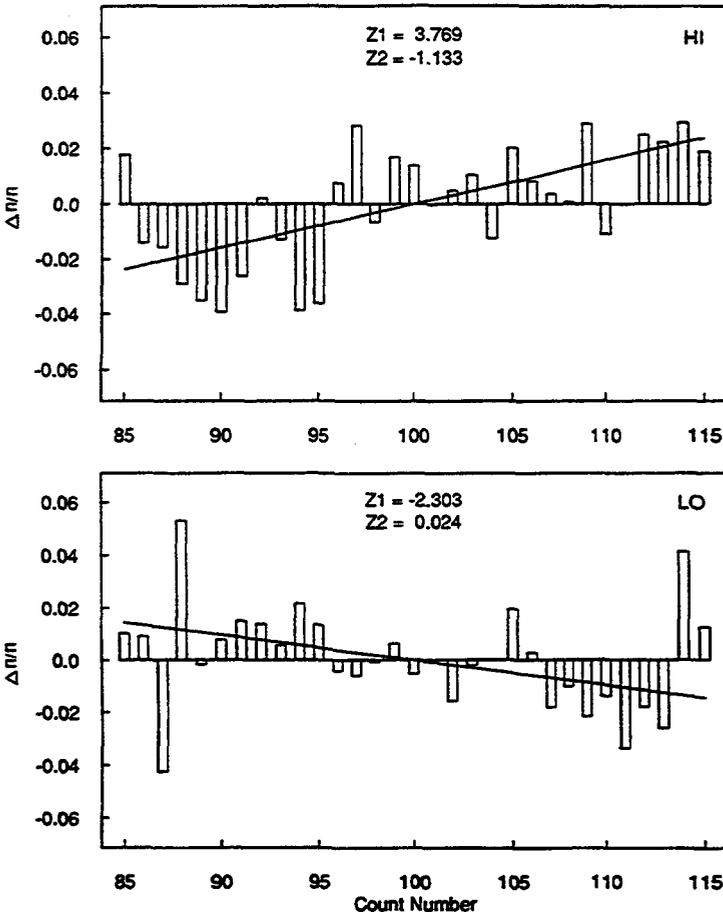
Fig. 4. Proportional frequency deviations. Diode REG: Operator 010.

tion is found to reach a maximum at a count somewhat outside of the first standard deviation. A theoretical leverage distribution can be computed from the linear fit of Eq. (7) and the Gaussian $n(x)$ approximation (2).

$$L(x) = 4\epsilon n(x - \mu)^2 = Ce^{-1/2(x-\mu/\sigma)^2} \cdot (x - \mu)^2 \qquad (8)$$

where C subsumes both the constant coefficients and the bias parameter, ϵ. For our values of μ and σ, this function has maxima at $(x - \mu) = \pm\sqrt{2}\sigma$, i.e., at the counts 110 and 90, consistent with the experimental distributions sketched in Figures 5 and 6 for the overall diode data base and the single operator, respectively.

It is tempting to leap from the generic result of relations (3) and (7) to speculate on possible modes of interaction of the consciousness of the human operator with the REG noise source, whereby the elemental probability
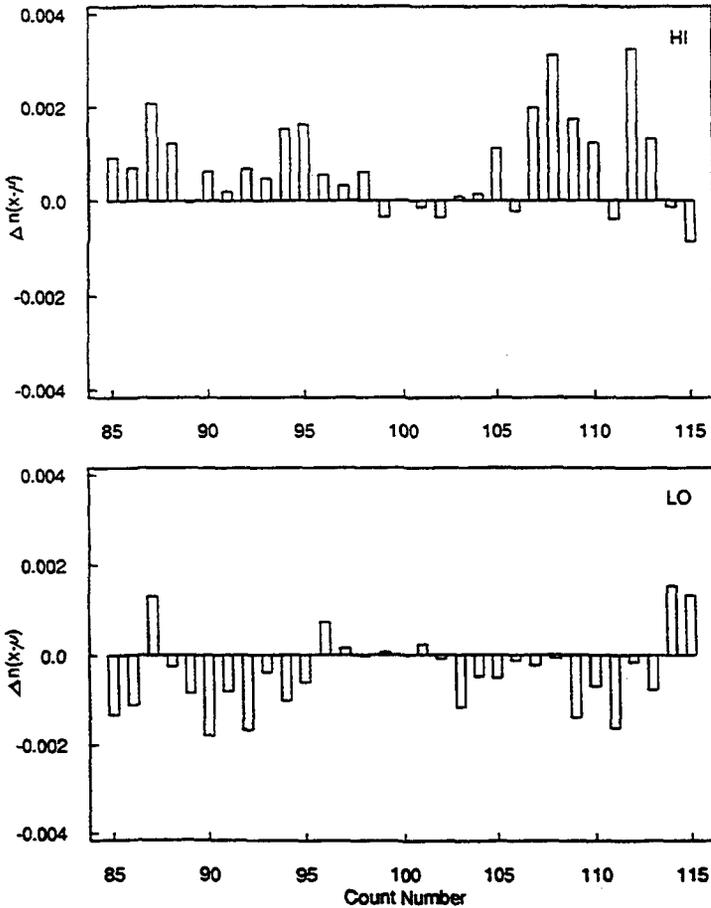
Fig. 5. Count leverage on mean shift. All diode REG: high, low.

of the microelectronic processes occurring therein might be altered by the indicated amounts. It should first be noted, however, that considerable electronic processing is applied to the output of the primary noise diode prior to the bit-counting stage, including the imposition of an alternating template that compares the raw random bit string with a regularly alternating criterion $(+, -, +, -, +, -)$ and counts the coincidences of sign. It is these coincident count distributions that constitute the experimental data and are used as feedback for the operators. While this strategy has the powerful technical advantage of eliminating any possible influence of remanent DC bias that might escape the extensive voltage regulation failsafe system, it considerably complicates and delimits the modalities by which the operators' influence may be imposed. In particular, introduction of a simple systematic bias in the raw noise pattern or the primary bit string will not avail. Rather, the
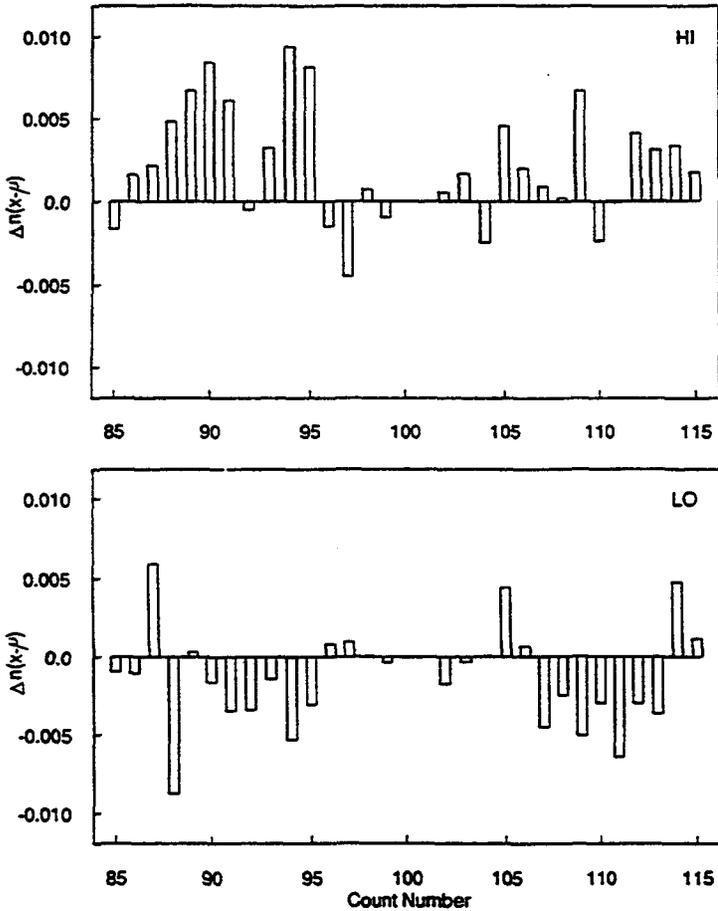
Fig. 6. Count leverage on mean shift. Diode REG, Operator 010.

effect must be exercised much further along in the preparation of the output count sequence.

This complexity of interaction is underscored by similar treatment of results from "pseudorandom" versions of the REG experiments. Early in the REG program, in an effort to assess the sensitivity of the observed effects to the specific nature of the noise source, a number of other microelectronic sources were developed, among them a hard-wired array of shift registers that produces a deterministic string of pseudorandom digits (Jahn, Dunne, & Nelson, 1987; Nelson, Dunne, & Jahn, 1984). This continuously running string, which repeats itself every thirty hours, can replace the diode random source in the composite REG circuit box by a flip of an external switch, leaving all other functions of the experiment, including its feedback and data processing, identical. Overall results using this pseudorandom version, in

terms of intention-correlated, operator-specific shifts of the mean, are essentially the same as with the electronic diode noise sources.

Since the only evident mode of operator influence on the output distributions of such a deterministic device is via the points of incursion into the bit string, examination of the count population patterns should be particularly interesting. Figure 7 shows $\Delta n/n$ for the entire pseudorandom data base of some 700,000 trials. (No individual operator has yet generated sufficient data on this device to support this analytical treatment.) Note once again the linear fits of magnitude consistent with the overall mean shifts. Clearly, if we are to retain the image of an induced bias in the underlying elemental probability, its mode of attainment must be extremely subtle. This point will be further illustrated in the context of the more substantially different experiments discussed in the next section.
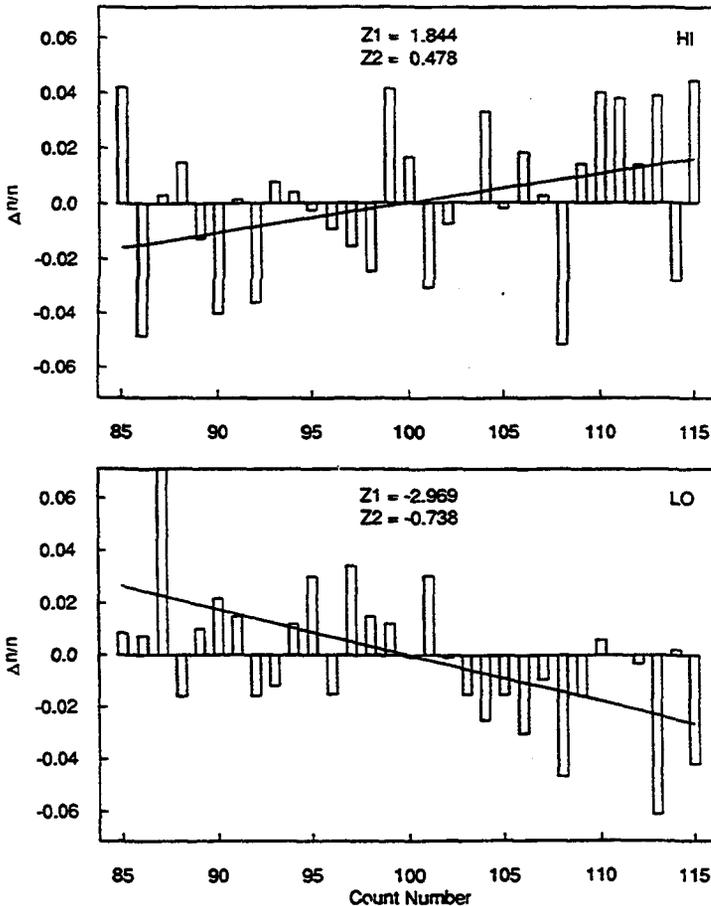

Fig. 7. Proportional frequency deviations. All pseudo-REG.

Bin Populations in a Random Mechanical Cascade Experiment

REG experiments like those described above have served as the bench-mark studies of the PEAR program since its inception. From these have radiated a number of ancillary experiments designed to explore systemati-cally the dependence of the anomalous phenomena on various technical, aesthetic, and procedural parameters, such as the scale of the machine, its physical domain, the character of its feedback, the quality of its randomicity, etc. Of these, the most extensively developed and deployed counterpart to the microelectronic REGs has been a room-size "Random Mechanical Cas-cade" (RMC) apparatus, shown in Figure 8a, wherein 9000 polystyrene spheres are allowed to trickle downward through an array of 330 nylon pegs to distribute themselves among 19 collecting bins lined across the bottom of the device. Photodetectors track the entrance of each ball into each bin and appropriate computational programs record and display on line the growing bin populations, followed by the terminal distribution properties.
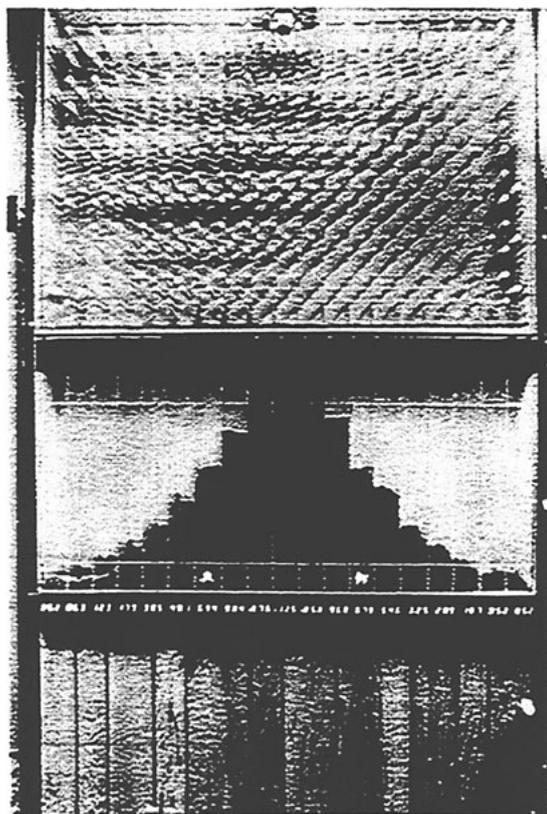

Fig. 8a. Random Mechanical Cascade Apparatus.

The detailed mechanics of the ball trajectories through the peg maze are much too complex to permit calculation of a theoretical bin population distribution, but when properly aligned, the machine yields calibration distributions sufficiently close to Gaussian to allow analysis by conventional parametric statistics (Figure 8b). Even so, various environmental vagaries, most notably hydroscopic changes in ball and peg resilience and long term ball and peg wear, preclude comparison of active experimental distributions with any fixed reference. Rather, data are invariably taken and processed on a tripolar differential basis wherein operator efforts to shift the distribution to the right or to the left, or to take an undisturbed null, are compared locally with one another.

Further details on the design, operation, and calibration of this machine, and full presentations of the large bodies of operator-specific and composite data, are available in the references (Nelson, Dunne, & Jahn, 1988a). Most briefly, anomalous effects quite comparable in scale and character to those obtained on the REG devices are found in these RMC experiments. In fact, the entire hierarchical list of salient effects sketched in the Introduction would apply equally well to this macroscopic mechanical facility. Of primary interest here, however, are the interior details of the terminal output distributions.

By its design and by its inherent statistical leverage, the RMC is even better disposed than the REGs to display of incremental count distributions, namely via the individual bin populations themselves. Indeed, although the absolute number of experimental runs of this machine are substantially smaller than for the REGs, there are fewer bins than REG counts, and the bin population data are actually somewhat less stochastic and easier to process. Figures 9 and 10, for example, show the mean differences in bin popula-
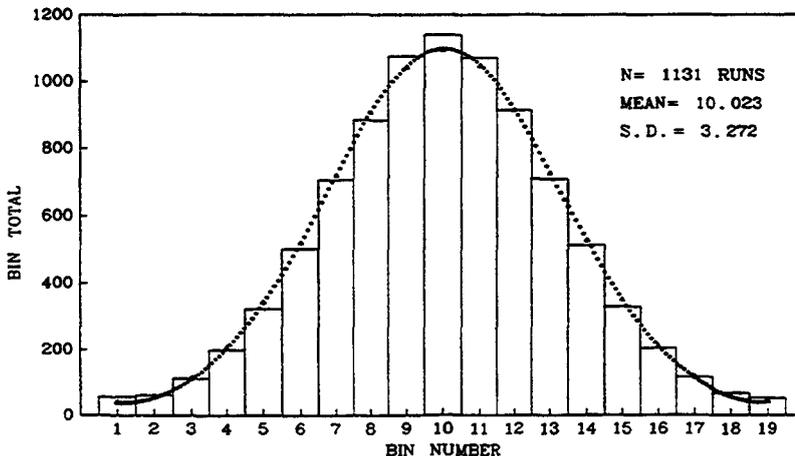

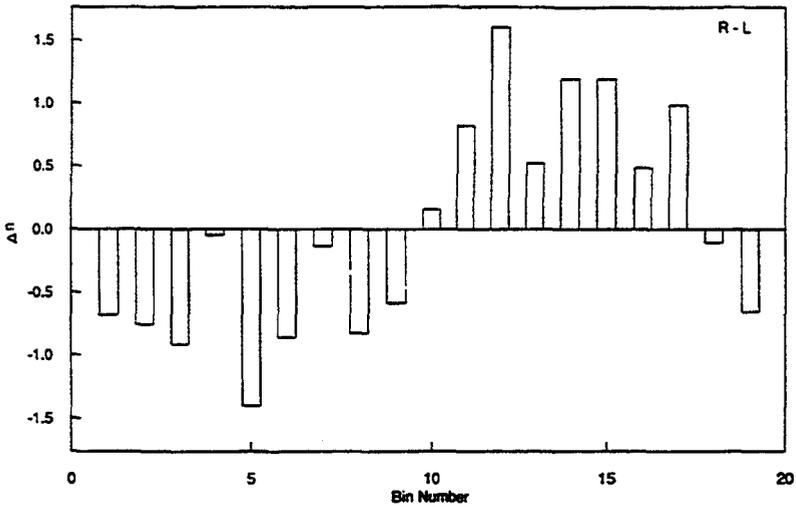Fig. 8b. RMC baseline mean bin totals on theoretical Gaussian.

Fig. 9. Mean bin population differences. All RMC.

tions for right versus left efforts for the entire RMC data base presented in Dunne, Nelson, and Jahn (1988) and Nelson, Dunne, and Jahn (1988a), and for the same individual operator, respectively. Note once again that a large number of the bins contribute substantially to the overall mean shift, rather than leaving the burden to any particular few.

As before, it is instructive to display these population decrements in $\Delta n/n$ form, where the reference n here is the average of the right and left effort populations for the corresponding bins. The results, like those shown in
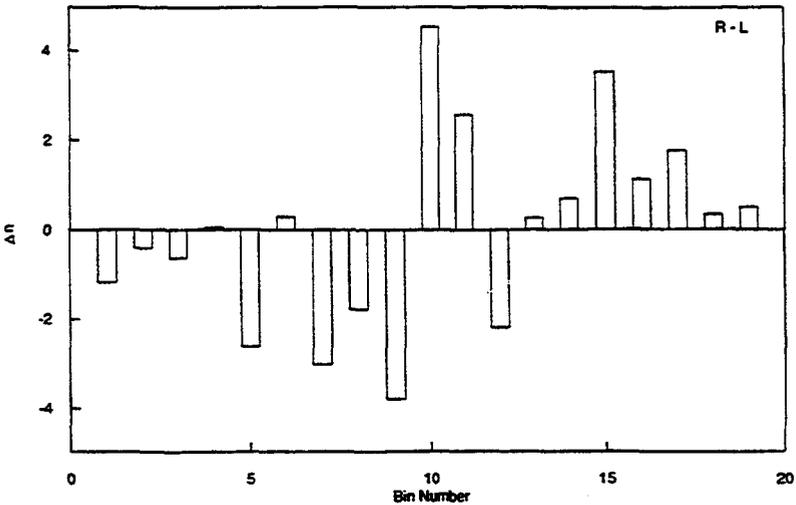

Fig. 10. Mean bin population differences. RMC, Operator 010.
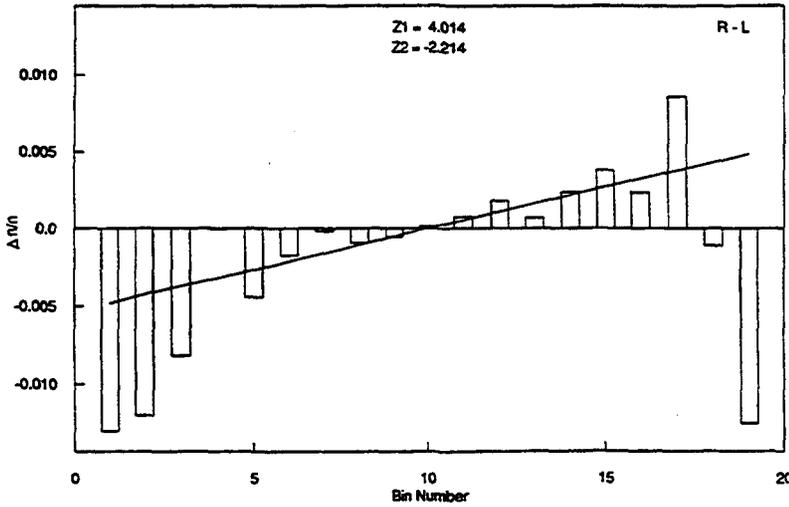
Fig. 11. Mean proportional bin population differences. All RMC.

Figures 11 and 12 and for many other subsets not shown, again lend themselves to a simple linear regression of the form

$$\frac{\Delta n}{n} = \delta(B - 10) \tag{9}$$

where B is the bin number from extreme left to extreme right (10 is the center bin) and δ is the slope of the fit. (The significant quadratic correction,

Fig. 12. Mean proportional bin population differences. RMC, Operator 010.

$Z_2 = -2.214$, in Figure 11, is attributable to the spurious effect of the RMC sidewalls, which tend to distort the end bin populations somewhat. If the calculations are repeated with bins number 1 and 19 excluded, $Z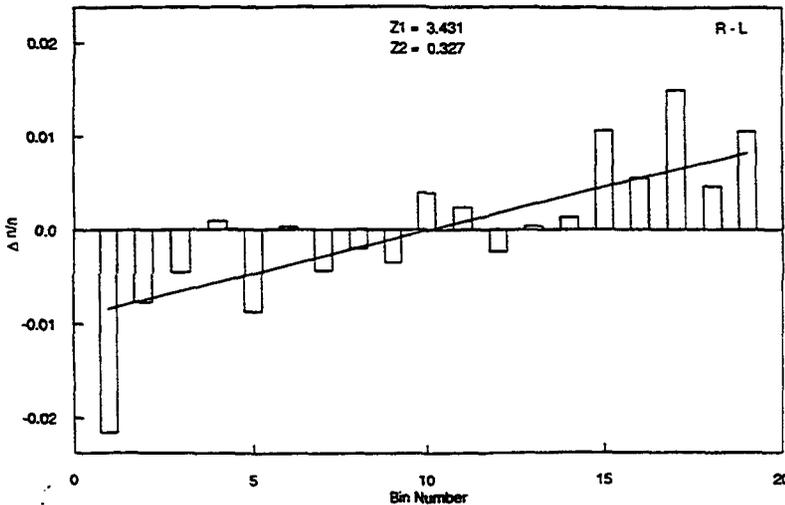_2$ drops to -0.754; if 2 and 18 are also excluded, Z_2 becomes 0.132. In both cases the value of Z_1 and δ remain nearly the same.)

The conformance of a large body of data on this macroscopical mechanical RMC device to the same form of linear $\Delta n/n$ relationship found for the microelectronic and pseudo REG results must have important implications for comprehension of the generic anomaly at work in these vastly disparate systems. In the previous section we commented on the technical complexity of the task facing the operator in attempting to impose his characteristic increment, ϵ, on the basic binary probability, p, to achieve the observed output results in the microelectronic and pseudorandom REGs. To this we now must add the yet more imposing complexity of the RMC system, where the elemental probability is deeply buried in a host of technical features, such as the ball inlet conditions, the configuration and resilience of the scattering pegs, and the high unpredictability of the ball/ball collisions that supplement the ball/peg interactions. Indeed, although we have on occasion made some attempts to interpret the observed bin population distributions of this machine in terms of a "quasi-binary" combinatorial, it is quite clear that the basic scattering events span wide ranges of elemental probability, and compound in a highly hierarchical and non-linear fashion, so that the resemblance of the output distributions to the Gaussian must be far more fortuitous than fundamental. In this light, the similarity of the observed bin population deviation patterns to those of the REGs becomes even more remarkable than simply the scale and genre differences of these classes of machine would suggest. In fact, it makes any common mechanistic interpretation so implausible as to require a more paradigmatic resolution.

Application to Remote Perception Data

The ubiquitous appearance of the count population patterns uncovered in the various human/machine experiments outlined above and the implications thereof for the basic mechanisms involved can be projected even further by one final example drawn from a substantially different sector of our research program. For over twelve years, this laboratory has also carried forth a coordinated experimental and theoretical component concerned with the acquisition, evaluation, and interpretation of data in a variety of protocols subsumed under the nomen of "Precognitive Remote Perception" (PRP)—the acquisition of information about physical target scenes, remote in both space and time, by common human percipients, using other than normal sensory processes. The primary focus of this effort has been the development of analytical judging techniques to quantify the anomalous

information contained in the several hundred formal target perceptions acquired in these experiments. These analytical strategies, fully described in a sequence of publications (Jahn, Dunne, & Jahn, 1980; Dunne, Jahn, & Nelson, 1983; Jahn et al., 1982; Dunne, Dobyns, & Intner, 1989), ultimately yield distributions of perception scores that can be compared with empirical chance distributions for the same scoring methods. Such pairs of score distributions are found to correspond closely enough to Gaussian forms to allow parametric statistical evaluation of the mean shifts and higher moments, much like that applied to the REG and RMC data. Despite the much smaller size of the PRP data base, the statistical significance of the mean shifts of the score distributions are considerably greater than for the human/machine experiments, with probabilities against chance ranging from 10^{-6} to 10^{-12}, depending on the particular data subset and scoring method.

If the PRP experimental and chance score distributions are binned into discrete increments, they may be subjected to the same Δn and $\Delta n/n$ analyses used for the REG and RMC data. Figures 13 and 14 show the results of such treatment of 277 formal PRP trials, ab initio encoded, employing a global target pool and both instructed and volitional protocols (Dunne, Dobyns, & Intner, 1989). The similarities of form to the human/machine data are unmistakable; the interpretation, however, is even more obscure.

All of the PRP scoring methods are based on an array of thirty binary descriptors ranging from very factual to very impressionistic details, such as whether the scene is outdoors, whether it is noisy, whether people are present, etc., that are answered by the percipients after dictating their free response narratives. Each of these descriptors has a particular frequency of occurrence across the target pool of all scenes in a given data set. The percip-
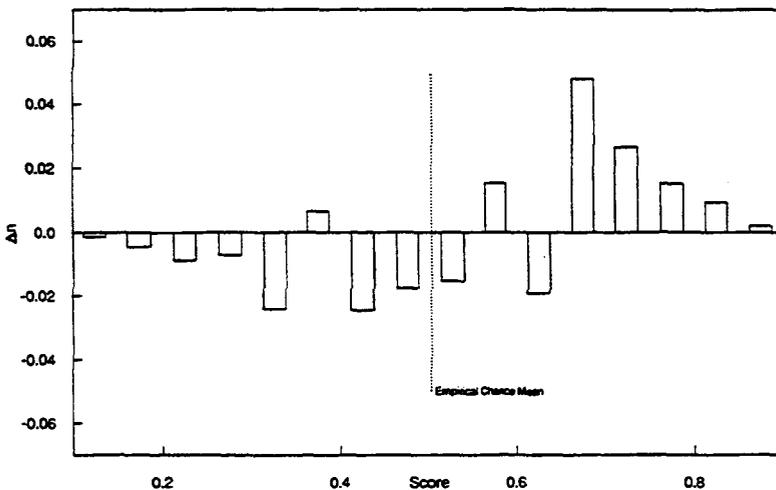

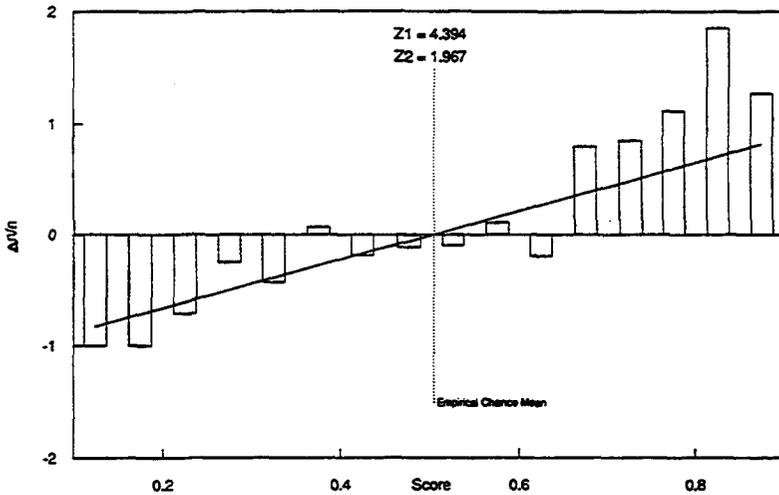Fig. 13. PRP score frequency differences. All ab initio data.

Fig. 14. PRP proportional score frequency differences. All ab initio data.

ient's response to each descriptor is compared to the proper statement for the given target, and that result is weighted in terms of the descriptor frequency. Thus, the linear $\Delta n/n$ pattern of Figure 14 is equivalent to a uniform slight improvement in the statistical likelihood of the percipients' proper identification of the target descriptors, beyond their normal chance occurrence across the utilized target pool. Mechanistically, this seems far removed from the alteration of bit probabilities in an REG, or from the systematic marginal migration of balls in an RMC, yet the $\Delta n/n$ pattern similarities among them are unmistakable.

Summary

The tantalizingly similar simple patterns of the count deviations in these four disparate experiments, tantamount in each case to marginal transposition of the chance Gaussian distributions consistent with incremental changes in their elemental binary probabilities, begs for some correspondingly simple common hypothesis for the attainment of these several empirical anomalies. Yet, the individual complexity and collective differences in the interior technical processes involved quickly render any such hypothetical mechanisms extremely convoluted at best, suggesting that a more generic and holistic approach, even if more radical in its paradigmatic implications, may ultimately be more productive. With no pretense of empirical verification or theoretical uniqueness, the remaining paragraphs offer one such possible representation.

In a strictly technical sense, the only difference between the chance expectations of the various experimental outputs and their demonstrated anomalous results is a matter of *information*. In each case, the anticipated random array of output digits, bin populations, or target descriptor scores has been slightly ordered, thereby decreasing its overall entropy, and raising its overall information content. Since the only empirically demonstrated primary correlate of this achievement is the pre-stated intention of the human operator, it is reasonable to assume that the source of this information increment is the consciousness of that operator. Whether the process is regarded as a direct transfer of information from the operator's consciousness to the machine's "consciousness," or as an internal rearrangement of the total information content of the bonded operator/machine system—although a philosophically intriguing distinction in its own right—is not of primary relevance here. Rather, the essential feature is that the particular output pattern now finds itself obliged to assimilate this increment of information.

In simplistic terms, the pattern has two options: it may retain a Gaussian distribution, displaced by the requisite amount to accommodate the full information increment solely within its mean shift; or, it may distribute some or all of the information into an internal rearrangement of the count distribution, e.g., by changing its variance, developing a skew or kurtosis, or forming a more pathological pattern. In opting for the former reaction, as our empirical results indicate, the added information is utilized in the most efficient fashion to fulfill the stated teleological task of the operator, in this case to shift the mean score, squandering none of it in unproductive internal distortions of the distribution. This option, which seems to be commonly elected in all of the various experiments described, thus takes the form of a "minimum information" principle, with the consciousness of the operator somehow specifying the nature of the experimental goal, and the output pattern deploying the minimum information necessary to achieve it.

A critical test of this hypothesis would be for the operators to address tasks other than shifts of the mean in otherwise similar experimental protocols, e.g., to attempt to change the distribution variance, to develop asymmetries, or to overpopulate particular counts, to see whether the systems respond in similarly efficient modes to fulfill these volitions. Unfortunately, given the huge data bases that would be needed to substantiate such statistical patterns, we are a long way from being able to validate the model in this more general form. The only relevant evidence in hand is from two operators who, in relatively small data sets, succeeded in significantly altering REG distribution variances, without substantially affecting the mean or other moments, and from a third who achieved similar effects on the RMC. One other operator has had minor success in over-populating particular preselected RMC bins. Many more data of this sort are clearly required.

Any information transfer model for the observed phenomena inevitably entails energy transfer considerations as well, on purely physical grounds. Of

the three recognized currencies of the physical world—substantial matter (mass), energy, and information—the relationship between the first two has been unequivocally established via Einstein's monumental equation and has dominated much of twentieth century physical science. The relationship between energy and information has been less incisively formulated and less extensively exploited, although models and empirical examples exist in many physical sectors, notably the Second Law of Thermodynamics; the quantum mechanical exchange energy of covalent molecular bonds; various electromagnetic resonance and coherence situations such as lasers and masers; and, of course, fundamental information theory à la Shannon and its many derivatives. The classical separation of these three physical currencies over most of scientific history is attributable to the huge size of the transmutation coefficients that relate mass to energy, and energy to information, respectively.

In our experimental situation, the inversion of a small fraction of the information bits from their chance configurations or, equivalently, the shift in the apparent elemental probabilities, also has energetic as well as informative implications, although the former are of miniscule scale, again given the magnitude of the transmutation coefficient. Nonetheless, what is of overarching interest here is the possibility that the consciousness of the operator, using that capacity for which it is most extraordinarily equipped—the processing of information—has in these interactions entered proactively into the affairs of the physical world, rearranging not only a portion of its information array, but thereby accessing its energy, and thence, by inference, its very substance. Extrapolating to more general implications, this model would thus suggest that the third side of the mass-energy-information triangle of physical currencies can provide a natural entry for human consciousness to participate in the construction of tangible reality, if we can but comprehend the dynamics of transfer of the subjective information of the mind to the technical information of the cosmos.

References

Dunne, B. J., Dobyns, Y. H., & Intner, S. M. (1989). Precognitive Remote Perception III: Complete Binary Data Base with Analytical Refinements (Technical Note PEAR 89002). Princeton Engineering Anomalies Research, Princeton University, School of Engineering/ Applied Science.

Dunne, B. J., Jahn, R. G., & Nelson, R. D. (1983). Precognitive Remote Perception (Technical Note PEAR 83003). Princeton Engineering Anomalies Research, Princeton University, School of Engineering/Applied Science.

Dunne, B. J., Nelson, R. D., & Jahn, R. G. (1988). Operator-related anomalies in a random mechanical cascade. *Journal of Scientific Exploration, 2,* 155–179.

Dunne, B. J., Nelson, R. D., Dobyns, Y. H., & Jahn, R. G. (1988). Individual Operator Contributions in Large Data Base Anomalies Experiments (Technical Note PEAR 88002). Princeton Engineering Anomalies Research, Princeton University, School of Engineering/Applied Science.

Jahn, R. G., & Dunne, B. J. (1987). *Margins of reality*. San Diego, New York, London: Harcourt Brace Jovanovich.

Jahn, R. G., Dunne, B. J., & Jahn, E. G. (1980). Analytical judging procedure for remote perception experiments. *Journal of Parapsychology, 44*, 207–231.

Jahn, R. G., Dunne, B. J., & Nelson, R. D. (1987). Engineering anomalies research. *Journal of Scientific Exploration, 1*, 21–50.

Jahn, R. G., Dunne, B. J., Nelson, R. D., Jahn, E. G., Curtis, T. A., & Cook, I. A. (1982). Analytical Judging Procedure for Remote Perception Experiments II: Ternary Coding and Generalized Descriptors. (Technical Note PEAR 82002). Princeton Engineering Anomalies Research, Princeton University, School of Engineering/Applied Science.

Nelson, R. D., Bradish, G. J., & Dobyns, Y. H. (1989). Random Event Generator: Qualification, Calibration and Analysis (Technical Note PEAR 89001). Princeton Engineering Anomalies Research, Princeton University, School of Engineering/Applied Science.

Nelson, R. D., Dunne, B. J., & Jahn, R. G. (1984). An REG Experiment with Large Data Base Capability III: Operator Related Anomalies (Technical Note PEAR 84003). Princeton Engineering Anomalies Research, Princeton University, School of Engineering/Applied Science.

Nelson, R. D., Dunne, B. J., & Jahn, R. G. (1988a). Operator Related Anomalies in a Random Mechanical Cascade Experiment (Technical Note PEAR 88001), and (1988b). Operator Related Anomalies in a Mechanical Cascade Experiment, Supplement: Individual Operator Series and Concatenations (Technical Note PEAR 88001.S). Princeton Engineering Anomalies Research, Princeton University, School of Engineering/Applied Science.

Appendix

Error-Weighted Linear Regression

All of the data consist of many observations of a dependent variable y_i, each associated with an independent variable value x_i and involving a measurement uncertainty σ_i. A linear model assumes that the data are of the form

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \tag{1}$$

where ϵ_i is an error term with mean zero. In an unweighted regression it is usually assumed that the error terms are all drawn from a common distribution, whereas for the weighted regression employed here the ϵ_i are presumed to have variance σ_i^2. It is further assumed that the various ϵ_i are independent. The goal of a regression analysis is to construct sample estimates b_0 of β_0 and b_1 of β_1.

A least-squares approach that minimizes the total, normalized squared error,

$$E \equiv \sum_i \left(\frac{b_0 + b_1 x_i - y_i}{\sigma_i} \right)^2, \tag{2}$$

seems a natural choice in that it gives the error term a χ^2 functional form when the ϵ_i are normal, and includes the unweighted regression form as a

special case. This total error may be minimized by finding zeroes of its partial derivatives,

$$\frac{\partial E}{\partial b_0} = \sum_i \frac{2(b_0 + b_1 x_i - y_i)}{\sigma_i^2} = 0$$

$$\frac{\partial E}{\partial b_1} = \sum_i \frac{2(b_0 + b_1 x_i - y_i)x_i}{\sigma_i^2} = 0. \tag{3}$$

On simplification, Eqs. (3) become a set of linear equations in two unknowns,

$$\left(\sum_i \frac{1}{\sigma_i^2}\right)b_0 + \left(\sum_i \frac{x_i}{\sigma_i^2}\right)b_1 = \left(\sum_i \frac{y_i}{\sigma_i^2}\right)$$

$$\left(\sum_i \frac{x_i}{\sigma_i^2}\right)b_0 + \left(\sum_i \frac{x_i^2}{\sigma_i^2}\right)b_1 = \left(\sum_i \frac{x_i y_i}{\sigma_i^2}\right), \tag{4}$$

which may readily be solved for b_0 and b_1, yielding

$$b_0 = \frac{\left(\sum_i \frac{x_i^2}{\sigma_i^2}\right)\left(\sum_i \frac{y_i}{\sigma_i^2}\right) - \left(\sum_i \frac{x_i}{\sigma_i^2}\right)\left(\sum_i \frac{x_i y_i}{\sigma_i^2}\right)}{\left(\sum_i \frac{1}{\sigma_i^2}\right)\left(\sum_i \frac{x_i^2}{\sigma_i^2}\right) - \left(\sum_i \frac{x_i}{\sigma_i^2}\right)^2}$$

$$b_1 = \frac{\left(\sum_i \frac{1}{\sigma_i^2}\right)\left(\sum_i \frac{x_i y_i}{\sigma_i^2}\right) - \left(\sum_i \frac{x_i}{\sigma_i^2}\right)\left(\sum_i \frac{y_i}{\sigma_i^2}\right)}{\left(\sum_i \frac{1}{\sigma_i^2}\right)\left(\sum_i \frac{x_i^2}{\sigma_i^2}\right) - \left(\sum_i \frac{x_i}{\sigma_i^2}\right)^2} \tag{5}$$

Note that both b_0 and b_1 are *linear* combinations of the y_i, so that

$$b_0 = \sum_i k_{0i} y_i \quad \text{and} \quad b_1 = \sum_i k_{1i} y_i, \tag{6}$$

where

$$k_{0i} \equiv \frac{\left(\sum_i \frac{x_i^2}{\sigma_i^2}\right)1/\sigma_i^2 - \left(\sum_i \frac{x_i}{\sigma_i^2}\right)x_i/\sigma_i^2}{\left(\sum_i \frac{1}{\sigma_i^2}\right)\left(\sum_i \frac{x_i^2}{\sigma_i^2}\right) - \left(\sum_i \frac{x_i}{\sigma_i^2}\right)^2} \tag{7}$$

and

$$k_{1i} \equiv \frac{\left(\sum_i \frac{1}{\sigma_i^2}\right)x_i/\sigma_i^2 - \left(\sum_i \frac{x_i}{\sigma_i^2}\right)1/\sigma_i^2}{\left(\sum_i \frac{1}{\sigma_i^2}\right)\left(\sum_i \frac{x_i^2}{\sigma_i^2}\right) - \left(\sum_i \frac{x_i}{\sigma_i^2}\right)^2}. \tag{8}$$

These k coefficients have the addition properties: $\sum k_{0i} = 1$, $\sum k_{0i} x_i = 0$, $\sum k_{1i} = 0$, and $\sum k_{1i} x_i = 1$, from which it follows that the expectation values of b_0, b_1 are indeed the model parameters β_0, β_1:

$$E[b_0] = E[\sum k_{0i} y_i] = E[\sum k_{0i} (\beta_0 + \beta_1 x_i + \epsilon_i)]$$

$$= \beta_0 \sum k_{0i} + \beta_1 \sum k_{0i} x_i + \sum k_{0i} E[\epsilon_i] \qquad (9)$$

$$= \beta_0,$$

where the last equality follows from the addition properties of k_{0i} and the fact that $E[\epsilon_i] = 0$ for all i, by definition. A parallel derivation with the k_{1i} shows that $E[b_1] = \beta_1$.

In order to establish confidence intervals for the model parameters, it is also necessary to know their variances. Using $\sigma^2[x]$ to denote the variance of a formula x, we note from the rules for variances of linear combinations that

$$\sigma^2[b_0] = \sigma^2[\sum k_{0i} y_i] = \sum k_{0i}^2 \sigma^2[y_i]. \qquad (10)$$

But, by the assumptions of the model, $\sigma^2[y_i] = \sigma^2[\epsilon_i] = \sigma_i^2$. Therefore $\sigma^2[b_0] = \sum (k_{0i} \sigma_i)^2$ and, by a similar derivation, $\sigma^2[b_1] = \sum (k_{1i} \sigma_i)^2$. This allows confidence intervals on b_0 and b_1 to be established.

In addition to inferences regarding the individual parameters, it is also often desirable to form joint inferences about the regression line that results from combining them. The hyperbolic confidence band that results from such calculations is an envelope about the regression line such that, with the stated likelihood, the actual model line $Y = \beta_0 + \beta_1 x$ lies entirely within the envelope. Establishing the confidence band requires calculation of the model prediction variance $\sigma^2[\hat{Y}]$ where $\hat{Y} = b_0 + b_1 x$ is the predicted value of the regression line at a given point x. The may be computed as

$$\sigma^2[\hat{Y}] = \sigma^2[b_0 + b_1 x]$$

$$= \sigma^2[\sum k_{0i} y_i + \sum k_{1i} y_i x]$$

$$= \sigma^2[\sum (k_{0i} + k_{1i} x) y_i] \qquad (11)$$

$$= \sum (k_{0i} + k_{1i} x)^2 \sigma^2[y_i]$$

$$= \sum (k_{0i}^2 + 2 k_{0i} k_{1i} x + k_{1i}^2 x^2) \sigma_i^2$$

where the intermediate expansion is needed because, while the variances of b_0 and b_1 are known, they are not independent. Given a formula for $\sigma^2[\hat{Y}]$, the locations of the confidence band limits are $\hat{Y} \pm \sqrt{2F[1 - \alpha; 2, n - 2] \sigma^2[\hat{Y}]}$ where n is the number of data points and $F[1 - \alpha; 2, n - 2]$ is the $1 - \alpha$ quantile of the F distribution for 2 and $n - 2$ degrees of freedom. In other words, it is the number such that an F distribution with 2 d.f. in the numerator and $(n - 2)$ d.f. in the denominator has a probability α of producing a larger value (Neter & Wasserman, 1974).

An important concern of any linear regression model is whether it leaves

some non-random component of the data unaccounted for. The simplest way to test this is by examination of the residuals after the regression line prediction is subtracted from the data. A rescaling of the x axis values to x^2, x^3 or any other suspected contributing term can then be performed, and the resulting data fitted with a new regression line to be examined for significant slope. The absence of such a slope does not totally guarantee that the contributing term is not present, but only that the data are insufficient to resolve it if it is.

Regression Line and Mean Shift

As demonstrated in the body of the text, if the theoretical frequency for the ith count value is denoted by n_i; the empirical frequency is n_i'; the deviation of the ith count from the theoretical mean is x_i; and the elementary success probability is $0.5 + \delta$, then

$$\frac{n_i' - n_i}{n_i} = \frac{\Delta n_i}{n_i} = 4\delta x_i \tag{12}$$

is the expected proportional shift in the population of the ith count value. Two other quantities that are useful for this discussion are P_i, the actual population of the ith count in a given dataset, and t_i, the theoretical population expected in a dataset of N trials. Note that $P_i = Nn_i'$, $t_i = Nn_i$, and $\sum t_i = \sum P_i = N$ all hold as identities from their respective definitions. Alternatively $\sum P_i = N$ may be regarded as the definition of N.

To consider this in the context of the regression analysis above, we may identify $y_i = (P_i - t_i)/t_i$ with the proportional shift in population for each count value. We then note that under the null hypothesis P_i is random with expectation t_i and standard deviation $\sqrt{t_i}$. Since σ_i are the standard deviations of y_i for the regression analysis, $\sigma_i = 1/\sqrt{t_i}$ follows.

The mean shift is clearly coupled to the individual count totals. The mean shift $m - \mu$ of a given population is the total deviation from the expected value, divided by the total population, that is,

$$m - \mu = \frac{\sum P_i x_i}{\sum P_i} = \frac{\sum P_i x_i}{N} . \tag{13}$$

From the definitions above, however, $P_i = y_i t_i + t_i$. Since the x_i are defined as deviations from the mean value, and the t_i are always symmetrically distributed about the mean value, the term $\sum x_i t_i$ vanishes identically, so that

$$m - \mu = \frac{\sum y_i t_i x_i}{N} = \sum n_i x_i y_i . \tag{14}$$

The δ value that corresponds to this mean shift is calculated as follows: for $p = 0.5$, μ is the expected number of hits in however many binaries are gath-

ered to form one trial; therefore the number of binaries in question is 2μ. The probability of a hit in one binary is $m/2\mu$, where m is the empirical trial mean, so that

$$\delta \equiv p - 0.5 = (m/2\mu) - 0.5 = \frac{m - \mu}{2\mu}. \qquad (15)$$

Combining (14) and (15), we obtain $\delta = (1/2\mu) \sum n_i x_i y_i$, which may be used in the right hand side of (12) to compute the predicted proportional shift $\hat{Y}_i$ for each count value:

$$\hat{Y}_i = \frac{2}{\mu} (\sum_j n_j x_j y_j) x_i. \qquad (16)$$

Now let us consider the linear regression fit that is computed for x_i, y_i, and σ_i as defined above. First, since the bin probabilities are symmetric about the theoretical mean, and x_i is by definition the deviation from that mean, $\sum x_i/\sigma_i^2 = \sum x_i t_i = 0$. This simplifies the formulas (5) for the regression coefficients to

$$b_0 = \frac{\sum t_i x_i^2 \sum t_i y_i}{\sum t_i \sum t_i x_i^2}$$

$$\qquad (17)$$

$$b_1 = \frac{\sum t_i \sum t_i x_i y_i}{\sum t_i \sum t_i x_i^2}.$$

Noting that $\sum P_i = \sum t_i = N$ and $P_i = y_i t_i + t_i$ are both required, we conclude that $\sum t_i y_i = 0$; this is in essence a constraint equation on y_i that follows from the fact that the t_i are normalized to the actual population. Applying this relation in (17) we see that b_0 now vanishes identically. The formula for b_1 may be simplified by noting that $\sum t_i x_i^2$ is N times the trial variance. Since the number of binaries per trial is 2μ, from the binomial formula the trial standard deviation is $\sqrt{2\mu pq} = \sqrt{\mu/2}$ for $p = q = 0.5$. Therefore the trial variance is just $\mu/2$ and $b_1 = \sum t_i x_i y_i / N(\mu/2) = (2/\mu) \sum n_i x_i y_i$. The regression line prediction for the ith bin is thus $\hat{Y}_i = (2/\mu)(\sum n_j x_j y_j) x_i$ which is identical to Eq. 16. In other words, the regression line and the mean-shift prediction line are identical, unless the set of counts is incomplete, so that count-conservation conditions such as $\sum P_i = \sum t_i$ are not perfectly obeyed. This latter situation actually arises in many cases, including those covered by this paper, whenever extreme count values, i.e., those in the tails of the distributions, are insufficiently populated to allow stable estimates, and must be excluded.

One can also show that the regression line preserves the mean shift, or, to paraphrase, that a hypothetical data set, whose count population all fall exactly on the regression line obtained from an actual data set, has exactly the same mean shift as that data set. From (14) the mean shift produced by the actual data is $\Delta M = \sum n_i x_i y_i$. The various symmetries and conservation

conditions coerce the regression parameters to $b_0 = 0$, $b_1 = (\sum n_i x_i y_i)/$ $(\sum n_i x_i^2)$. The predicted Y values for the regression line are then $\hat{Y}_i$ $= x_i (\sum n_j x_j y_j)/(\sum n_j x_j^2)$. But formula (14) for the mean shift applies to any set of y_i, so the mean shift for the hypothetical set $\hat{Y}_i$ can be calculated:

$$\Delta M = \sum n_i x_i \hat{Y}_i$$

$$= \sum n_i x_i \left(x_i \frac{\sum n_j x_j y_j}{\sum n_j x_j^2} \right)$$

$$= \frac{(\sum n_i x_i^2)(\sum n_i x_i y_i)}{\sum n_i x_i^2} \qquad (18)$$

$$= \sum n_i x_i y_i.$$

Reference

Neter, J., and Wasserman, W. (1974). *Applied linear statistical models.* Homewood, Illinois: Richard D. Irwin Inc.