Applications of Decision Augmentation Theory

by

Edwin C. May, Ph.D
Science Applications International Corporation
Menlo Park, CA

Jessica M. Utts, Ph.D.
University of California, Davis
Division of Statistics
Davis, CA

and

S. James Spottiswoode (Consultant), Christine L. James
Science Applications International Corporation
Menlo Park, CA

Abstract

Decision Augmentation Theory (*DAT*) provides an informational mechanism for a class of anomalous mental phenomena which have hitherto been viewed as a causal, force-like, mechanism. Under the proper conditions, DAT's predictions for micro-anomalous perturbation databases are different from those of all force-like mechanisms, except for one degenerate case. For large random number generator (RNG) databases, *DAT* predicts a zero slope for a least squares fit to the (Z^2,n) scatter diagram, where n is the number of bits resulting from a single run and Z is the resulting Z-score. We find a slope of $(1.73 \pm 3.19) \times 10^{-6}$ $(t = 0.543, df = 126, p \leq 0.295)$ for the historical binary random number generator database. In a 2-sequence length analysis of a limited set of RNG data from the Princeton Engineering Anomalies Research laboratory, we find that a force-like explanation misses the observed data by 8.6-σ; however, the observed data is within 1.1-σ of the DAT prediction. We also apply *DAT* to one PRNG study and find that its predicted slope is not significantly different from the expected value. We review and comment on six published articles that discussed *DAT's* earlier formalism (i.e., Intuitive Data Sorting—IDS). Our *DAT* analysis of Braud's hemolysis study confirms his finding in favor of a causal model over a selection one (i.e., $p \leq 0.023$); so far, this is the only study we have found that supports anomalous perturbation (*AP*). We provide six circumstantial arguments, which are based upon experimental outcomes against the perturbation hypothesis. Our anomalous cognition (*AC*) research suggests that the quality of *AC* data is proportional to the total change of Shannon entropy. We demonstrate that the change of Shannon entropy of a binary sequence from chance is independent of sequence length; thus, we have provided a fundamental argument in favor of *DAT* over causal models. In our conclusion, we suggest that, except for one degenerate case, the physical RNG database cannot be explained by any causal model, and that Braud's contradicting evidence should inspire more *AP* investigations of biological systems in a way that would allow a valid *DAT* analysis.

Introduction

May, Utts, and Spottiswoode (1994) proposed Decision Augmentation Theory (*DAT*) as a general model of anomalous mental phenomena (*AMP*).* *DAT* holds that anomalous cognition (*AC*) information is included along with the usual inputs that result in a final human decision that favours a "desired" outcome. In statistical parlance, *DAT* says that a slight, systematic bias is introduced into the decision process by *AC*.

This concept has the advantage of being quite general. We know of no experiment that is devoid of at least one human decision; thus, *DAT* might be the underlying basis for *AMP*. May et al. (1994) mathematically developed this concept and constructed a retrospective test algorithm than can be applied to existing databases. In this paper, we summarize the theoretical predictions of *DAT*, review the criteria for valid retrospective tests, and analyze the historical random number generator (RNG) database. In addition, we summarize the findings from one prospective test of *DAT* (Radin and May, 1985) and comment on the published criticisms of an earlier formulation, which was then called Intuitive Data Sorting. We conclude with a discussion of RNG results that provide a strong circumstantial argument against a causal explanation. As part of this review, we show that one biological-AP experiment is better described by an influence model (Braud, 1990).

Review of Decision Augmentation Theory

Since the formal discussion of *DAT* is statistical, we will describe the overall context for the development of the model from that perspective. Consider a random variable, X, that can take on continuous values (e.g., the normal distribution). Examples of X might be the hit rate in an RNG experiment when the number of binary bits in the sequence is large, the swimming velocity of cells, or the mutation rate of bacteria. Let Y be the average computed over n values of X, where n is the number of items that are *collectively* subjected to an *AMP* influence as the result of a single decision—one trial, and let Z be the appropriate Z-score corresponding to Y. Often this may be equivalent to a single effort period, but it also may include repeated efforts. The key point is that, regardless of the effort style, the average value of the dependent variable is computed over the n values resulting from one decision point. In the examples above, n is the sequence length of a single run in an RNG experiment, the number of swimming cells measured during the trial, or the number of bacteria-containing test tubes present during the trial.

Under *DAT*, we assume that the underlying parent distribution of a physical system remains *unperturbed*; however, the measurements of the physical system are systematically biased by an *AC*-mediated informational process. Since the deviations seen in actual experiments tend to be small in magnitude, it is safe to assume that the measurement biases are small and that the sampling distribution will remain normal; therefore, we assume the bias appears as small shifts of the mean and variance of the sampling distribution as:

$$Z \sim N(\mu_z, \sigma_z^2).$$

* The Cognitive Sciences Laboratory has adopted the term *anomalous mental phenomena* instead of the more widely known *psi*. Likewise, we use the terms *anomalous cognition* and *anomalous perturbation* for *ESP* and *PK*, respectively. We have done so because we believe that these terms are more naturally descriptive of the observables and are neutral in that they do not imply mechanisms. These new terms will be used throughout this paper.

This notation means that Z is distributed as a normal distribution with a mean of μ_z and a standard deviation of σ_z. Under the null hypothesis, $\mu_z = 0.0$ and $\sigma_z = 1.0$.

Review of a Causal Model

For comparison sake, we summarize a model for *AP* interactions. We begin with the assumption that a putative *AP* force would give rise to a perturbational interaction. What we mean is that given an ensemble of entities (e.g., random binary bits), a small force perturbs, on the average, each member of the ensemble. We call this type of interaction perturbational *AP* (*PAP*).

In the simplest *PAP* model, the perturbation induces a change in the mean of the parent distribution but does not effect its variance. We parameterize the mean shift in terms of a multiplier of the initial standard deviation. Thus:

$$\mu_1 = \mu_0 + \varepsilon_{AP}\, \sigma_0,$$

where μ_1 and μ_0 are the means of the perturbed and unperturbed distributions, respectively, and where σ_0 is the standard deviation of the unperturbed distribution. ε_{AP} can be considered the *AP* effect size. Under the null hypothesis for binary RNG experiments, $\mu_1 = \mu_0 = 0.5$, $\sigma_0 = 0.5$, and $\varepsilon_{AP} = 0$.

The expected value and variance of Z^2 under the mean-chance-expectation (*MCE*), *PAP*, and *DAT* assumptions for the normal distribution are shown in Table 1. The details of the calculations can be found in May, Utts, and Spottiswoode (1994).

Table 1.

Normal Parent Distribution

Quantity	Mechanism		
	MCE	PAP	DAT
$E(Z^2)$	1	$1 + \varepsilon_{AP}^2 n$	$\mu_z^2 + \sigma_z^2$
$Var(Z^2)$	2	$2(1 + 2\varepsilon_{AP}^2 n)$	$2(\sigma_z^4 + 2\mu_z^2\sigma_z^2)$

Figure 1 displays these theoretical calculations for the three mechanisms graphically.

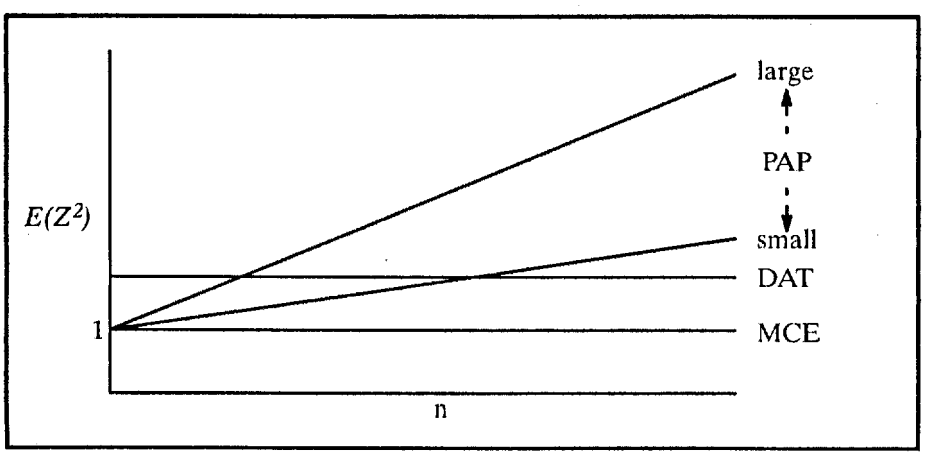

Figure 1. Predictions of *MCE*, *PAP*, and *DAT*.

This formulation predicts grossly different outcomes for these models and, therefore, is ultimately capable of separating them, even for very small perturbations.

Retrospective Tests

It is possible to apply *DAT* retrospectively to any body of data that meets certain constraints. It is critical to keep in mind the meaning of n—the number of measures of the dependent variable over which to compute an average during a single trial following a single decision. In terms of their predictions for experimental results, the crucial distinction between *DAT* and the *PAP* model is the dependence of the results upon n; therefore, experiments which are used to test these theories ideally should be those in which experiment participants are blind to n, and where the distribution of n does not contain extreme outliers.

Aside from these considerations, the application of *DAT* is straight forward. Having identified the unit of analysis and n, simply create a scatter diagram of points (Z^2, n) and compute a weighted least square fit to a straight line. Table 1 shows that for the *PAP* model, the slope of the resulting fit is the square of the *AP*-effect size. A student's t-test may be used to test the hypothesis that the *AP*-effect size is zero, and thus test for the validity of the *PAP* model. If the slope is zero, these same tables show that the intercept may be interpreted as an *AC* strength parameter for *DAT*. In other words, an intercept larger than one would support the *DAT* model, while a slope greater then zero would support he *PAP* model.

Monte Carlo Verification

The expressions shown in Table 1 are representations which arise from simple algebraic manipulations of the basic mathematical assumptions of the *PAP* and *DAT* models. To verify that these expressions give the expected results, we used a published pseudo random number generator (Lewis, 1975) with well-understood properties to produced data that mimicked the results under three models (i.e., *MCE*, *PAP* and *DAT*). Our standard implementation of the PRNG allows the integers in the range $(0,2^{15}-1]$ as potential seeds. For the sequence lengths *100, 500, 1000,* and *5000*, we computed Z-scores for all possible seeds with an effect size of *0.0* to simulate *MCE* and an effect size of *0.03* to simulate *PAP*. We used the fact that if the effect size varies as $1/\sqrt{n}$, *DAT* and PAP are degenerate to simulate *DAT*. For this case we used effect sizes of *0.030, 0.0134, 0.0095*, and *0.0042* for the above sequence lengths, respectively. Figures 2a-c show the results of 100 trials, which were chosen randomly from the appropriate Z-score data sets, at each of the sequence lengths for each of the models. In each Figure, MCE is indicated by a horizontal solid line a $Z^2 = 1$.

The slope of a least squares fit computed under the *MCE* simulation was $-(2.81 \pm 2.49) \times 10^{-6}$, which corresponded to a p-value of 0.812 when tested against zero, and the intercept was 1.007 ± 0.005, which corresponds to a p-value of 0.131 when tested against one. Under the *PAP* model, an estimate of the effect size using the expression in Table 1 was $\varepsilon_{AP} = 0.0288 \pm 0.002$, which is in good agreement with 0.03, the value that was used to create the data. Similarly, under *DAT* the slope was $-(2.44 \pm 57.10) \times 10^{-8}$, which corresponded to a p-value of 0.515 when tested against zero, and the intercept was 1.050 ± 0.001, which corresponds to a p-value of 2.4×10^{-4} when tested against one.

Thus, we are able to say that the Monte Carlo simulations confirm the simple formulation shown in Table 1.

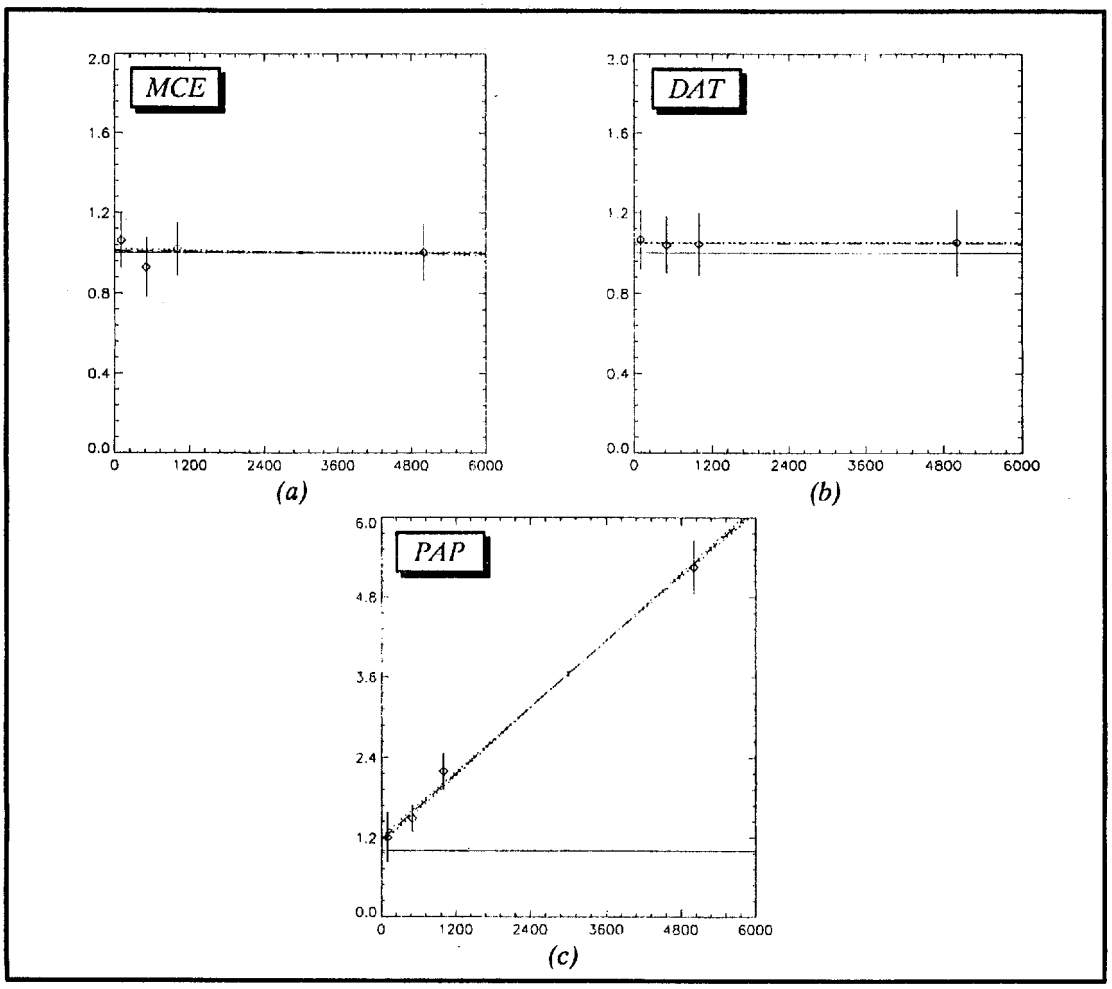

Figure 2. Z^2 vs n for Monte Carlo Simulations of *MCE, PAP,* and *DAT.*

Historical Binary RNG Database

Radin and Nelson (1989) analyzed the complete literature (i.e., over 800 individual studies) of consciousness-related anomalies in random physical systems. They eloquently demonstrated that a robust statistical anomaly exists in that database. Although they analyzed this data from a number of perspectives, they report an average $Z / \sqrt{n}$ effect size of approximately 3×10^{-4}, regardless of the analysis type. Radin and Nelson did not report p-values, but they quote a mean Z of 0.645 and a standard deviation of 1.601 for 597 studies. We compute a single-mean t-score of 9.844, $df = 596$ ($p \leq 3.7 \times 10^{-23}$).

We returned to the original publications of all the binary RNG studies from those listed by Radin and Nelson and identified 128 studies in which we could compute, or were given, the average Z-score, the number of runs, N, and the sequence length, n, which ranged from 16 to 10,000. For each of these studies we computed:

$$\overline{Z^2} = \mu_z^2 + \left(\frac{N-1}{N}\right) s_z^2. \tag{1}$$

Since we were unable to determine the standard deviations of the Z-scores from the literature, we *assumed* that $s_z = 1.0$ for each study. In addition, if aim directions were reported, we inverted the Z-score

for the negative aim since the *DAT* analysis is fundamentally two-tailed. We see from Table 1 that under *MCE* the expected variance is *2.0* so that the estimated standard deviations for the $\overline{Z}^2$ from Equation 1 is $\sqrt{2.0/N}$.

Figure 3 shows a portion of the 128 data points $(\overline{Z}^2, n)$. *MCE* is shown as a solid line (i.e., $Z^2 = 1$), and the expected best-fit lines for two assumed *AP* effect sizes of *0.01* and *0.003*, respectively, are shown as short dashed lines. We calculated a weighted (i.e., using *N/2.0* as the weights) least squares fit to an $a + b*n$ straight line for the 128 data points and display it as a long-dashed line. For clarity, we have offset and limited the Z^2 axis and have not shown the error bars for the individual points, but the weights and all the data were used in the least squares fit. We found an intercept of $a = 1.036 \pm 0.004$. The 1-standard error for the intercept is small and is shown in Figure 3 in the center of the sequence range. The *t*-score for the intercept being different from *1.0* (i.e., $t = 9.1$, $df = 126$, $p \leq 4.8 \times 10^{-20}$) is in good agreement with that derived from Radin and Nelson's analysis. Since we set standard deviations for all the *Z*'s equal to one; and since Radin and Nelson showed that the overall standard deviation was 1.6, we would expect that our analysis would be more conservative than theirs.

The important result, however, was that the slope of the best-fit line was $b = (1.73 \pm 3.19) \times 10^{-6}$ ($t = 0.543$, $df = 126$, $p \leq 0.295$), which is not significantly different from zero. The 1-σ standard error for the slope encompasses zero. Even though a very small *AP* effect size might fit the data at large sequence lengths, it is clear in Figure 3 what happens at small sequence lengths; an $\varepsilon_{AP} = 0.003$, suggests a linear fit that is significantly *below* the actual fit.

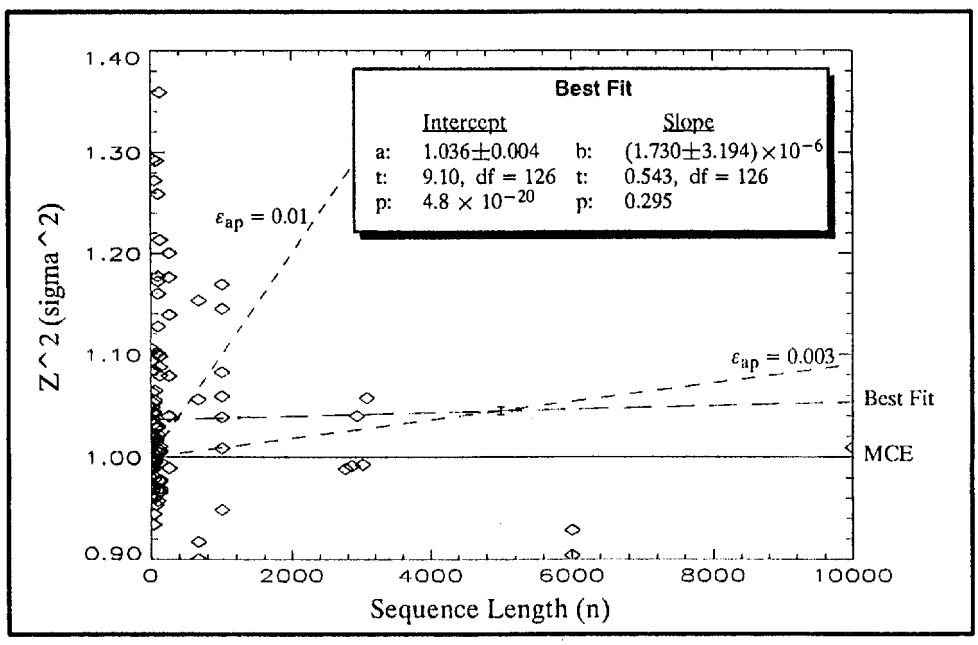

Figure 3. Binary RNG Database: Slope and Intercept for Best Fit Line

The sequence lengths from this database are not symmetric nor are they distributed and do contain ourliers (i.e., median = 64, average = 566). Figure 4 shows that the lower half of the data, however, is symmetric and nearly uniformly distributed (i.e., median = 35, average = 34). Since the criteria for a valid retrospective test is that *n* should be uniform, we analyzed the two median halves independently. The intercept for the weighted best-fit line for the uniform lower half is $a = 1.022 \pm 0.006$ ($t = 3.63$, $df = 62$, $p \leq 2.9 \times 10^{-4}$), and the slope is $b = (-0.034 \pm 3.70) \times 10^{-4}$ ($t = -0.010$, $df = 62$, $p \leq 0.504$). The

fits for the upper half yield $a = 1.064 \pm 0.005$ ($t = 13.47$, $df = 62$, $p \leq 1.2 \times 10^{-41}$) and $b = (-4.52 \pm 2.38) \times 10^{-6}$ ($t = -1.903$, $df = 62$, $p \leq 0.969$), for the intercept and slope, respectively.

Since the best retrospective test for *DAT* is one in which the distribution of n contains no outliers, the statistically zero slope for the fit to the lower half of the data is inconsistent with a simple *AP* model. Although the same conclusion could be reached from the fits to the database in its entirety (i.e., Figure 3), we suggest caution in that this fit could possibly be distorted by the distribution of the sequence lengths. That is, a few points at large sequence lengths can easily influence the slope. Since the slope for the upper half of the data is statistically slightly negative, it is problematical to assign an imaginary *AP* effect size to these data. More likely, the results are distorted by a few outliers in the upper half of the data.

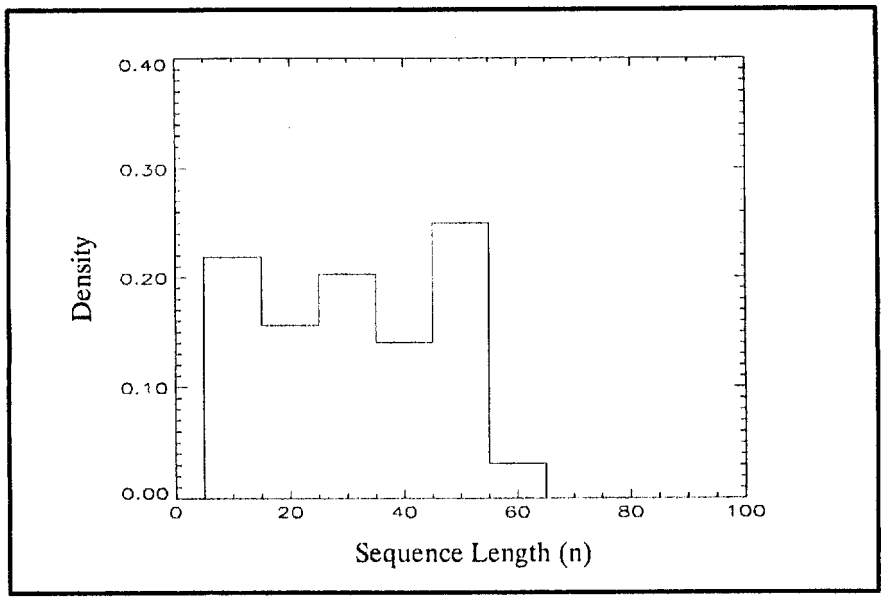

Figure 4. Historical Database: Distribution of Sequence Lengths < 64.

From these analyses, it appears that Z^2 does not linearly depend upon the sequence length; however, since the scatter is so large, even a linear model is not a good fit (i.e., $X^2 = 171.2$, $df = 125$, $p \leq 0.0038$), where X^2 is a goodness-of-fit measure in general given by:

$$X^2 = \sum_{j=1}^{\nu} \frac{1}{\sigma_j^2}(y_j - f_j)^2,$$

where the σ_j are the errors associated with data point y_j, f_j is the falue of the fitted function at point j, and ν is the number of data points.

A "good" fit to a set of data should lead to a non-significant X^2. The fit is not improved by using higher order polynomials (i.e., $X^2 = 170.8$, $df = 124$; $X^2 = 174.1$, $df = 123$; for quadratics and cubics, respectively). If, however, the *AP* effect size was *any* monotonic function of n other than the degenerate case where the *AP* effect size is exactly proportional to $1 / \sqrt{n}$, it would manifest as a non-zero slope in the regression analysis.

Within the limits of this retrospective analysis, we conclude for RNG experiments that we must reject all causal models of *AP* which propose a shift of the mean of the parent distribution.

Princeton Engineering Anomalies Research Laboratory RNG Data

The historical database we analyzed does not include the extensive RNG data from the Princeton Engineering Anomalies Research (PEAR) laboratory since their total number of bits exceeds the total amount in the entire historical database. For example, in a recent report Nelson, Dobyns, Dunne, and Jahn (1991) analyze 5.6×10^6 trials all at $n = 200$. In this section, we apply *DAT* retrospectively to their published work where they have examined other sequence lengths; however, even in these cases, they report over five times as much data as in the historical database.

Jahn (1982) reported an initial RNG data set with a single operator at $n = 200$ and $2{,}000$. Data were collected both in the *automatic* mode (i.e., a single button press produced 50 trials at n) and the *manual* mode (i.e., a single button press produced one trial at n). From a *DAT* perspective, data were actually collect at four values of n (i.e., 200, 2000, $200 \times 50 = 10{,}000$, and $2000 \times 50 = 100{,}000$). Unfortunately data from these two modes were grouped together and reported only at 200 and $2{,}000$ bit/trial. It would seem, therefore, we would be unable to apply *DAT* to these data. Jahn, however, reports that the different modes "...give little indication of importance of such factors in the overall performance." This qualitative statement suggests that the *PAP* model is indeed not a good description for these data, because, under *PAP*, we would expect stronger effects at the longer sequence lengths.

Nelson, Jahn, and Dunne (1986) describe an extensive RNG and pseudorandom RNG (PRNG) database in the manual mode only (i.e., over 7×10^6 trials); however, whereas Jahn provide the mean and standard deviations for the hits, Nelson et al. report only the means. We are unable to apply *DAT* to these data, because any assumption about the standard deviations would be highly amplified by the massive data set.

As part of a cooperative agreement in 1987 between PEAR and the Cognitive Sciences Program at SRI International, we analyzed a set of RNG data from a single operator.* Since they supplied the raw data for each button press, we were able to analyze this data at two extreme values of n. We combined the individual trial Z-scores for the high and low aims by inverting the sign of the low-aim scores, because our analysis is two-tailed, in that we examine Z^2.

Given that the data sets at $n = 200$ and $100{,}000$ were independently significant (Stouffer's Z of 3.37 and 2.45, respectively), and given the wide separation between the sequence lengths, we used *DAT* as a retrospective test on these two data points.

Because we are examining only two values of n, we do not compute a best-fit slope. Instead, as outlined in May, Utts, and Spottiswoode (1994), we compare the *PAP* prediction to the actual data at a single value of n.

At $n = 200$, 5918 trials yielded $\overline{Z} = 0.044 \pm 1.030$ and $\overline{Z^2} = 1.063 \pm 0.019$. We compute a *proposed AP* effect size $\overline{Z} / \sqrt{200} = 3.10 \times 10^{-3}$. With this effect size, we computed what would be expected under the *PAP* model at $n = 100{,}000$. Using the theoretical expressions in Table 1, we computed $\overline{Z^2} = 1.961 \pm 0.099$. The 1-sigma error is derived from the theoretical variance divided by the actual number of trials (597) at $n = 100{,}000$. The *observed* values were $\overline{Z} = 0.100 \pm 0.997$ and $\overline{Z^2} = 1.002 \pm 0.050$. A t-test between the observed and expect values of $\overline{Z^2}$ gives $t = 8.643$, $df = 1192$. Considering this t as equivalent to a Z, the data at $n = 100{,}000$ fails to meet what would be expected under the causal *PAP* model by

* We thank R. Jahn, B. Dunne, and R. Nelson for providing this raw data for our analysis in 1987.

8.6-σ. Suppose, however, that the effect size observed at $n = 100{,}000$ (3.18×10^{-4}) better represents the *AP* effect size. We computed the predicted value of $\overline{Z}^2 = 1.00002 \pm 0.018$ for $n = 200$. Using a *t*-test for the difference between the observed value and this predicted one gives $t = 2.398$, $df = 11{,}834$. The *PAP* model fails in this direction by more than *2.4-σ*. *DAT* predicts that $\overline{Z}^2$ would be statistically equivalent at the two sequence lengths, and we find that to be the case ($t = 1.14$, $df = 6513$, $p \leq 0.127$).

Jahn (1982) indicates in their 1982 PEAR RNG data that "Traced back to the elemental binary samples, these values imply directed inversion from chance behavior of about one or one and a half bits in every one thousand... ." Assuming *1.5* excess bits/1000, we calculate an *AP* effect size of *0.003*, which is consistent with the observed value in their $n = 200$ data set. Thus, we are forced to conclude that this data set from PEAR is inconsistent with the simple *PAP* model, and that Jahn's statement is not a good description of the anomaly.

We urge caution in interpreting these calculations. As is often the case in a retrospective analysis, some of the required criteria for a meaningful test are violated. These data were not collected when the operators were blind to the sequence length. Secondly, these data represent only a fraction of PEAR's RNG database.

A Prospective Test of *DAT*

In developing a methodology for future tests, Radin and May (1986) worked with two operators who had previously demonstrated strong *AP* ability in RNG studies. They used a pseudorandom number generator (PRNG), which was based on a shift-register algorithm by Kendell and has been shown to meet the general criteria for "randomness" (Lewis, 1975), to create the binary sequences so that timing considerations could be examined.

The operators were blind to which of nine different sequences (i.e., $n = 101, 201, 401, 701, 1001, 2001, 4001, 7001, 10001$ bits)* were used in any given trial, and the program was such that the trials lasted for a fixed time period and feedback was presented only after the trial was complete. Thus, the criteria for a valid test of *DAT* had been met, except that the source of the binary bits was a PRNG.

We re-analyzed the combined data from this experiment with the current *Z*-score formalism of *DAT*. For the 200 individual runs (i.e. 10 at each of the sequence lengths for each of the two participants) we found the best fit line to yield a slope $= 4.3 \times 10^{-8} \pm 1.6 \times 10^{-6}$ ($t = 0.028$, $df = 8$, $p \leq 0.489$) and an intercept $= 1.16 \pm 0.06$ ($t = 2.89$, $df = 8$, $p \leq 0.01$). The slope easily encompasses zero and is not significantly different from zero, the significance level is consistent with what Radin and May reported earlier.

Since the PRNG seeds and bit streams were saved for each trial, it was possible to determine if the experiment sequences exactly matched the ones produced by the shift register algorithm; they did. Since our UNIX-based Sun Microsystems workstations were synchronized to the system clock, *any* momentary interruption of the clock would "crash" the machine, but no such crashes occurred. Therefore, we believe *no* casual interaction occurred.

To explore the timing aspects of the experiment Radin and May reran each run with PRNG seeds ranging from −5 to +5 clock ticks (i.e., 20 ms/tick) from the actual seed used in the run. We plot the resulting

* The original *IDS* analysis required the sequence lengths to odd because of the logarithmic formalism.

run effect sizes, which we computed from the experimental F-ratios (Rosenthal, 1991), for operator 531 in Figure 5. The estimated 1-standard errors are the same for each seed shift and equal 0.057.

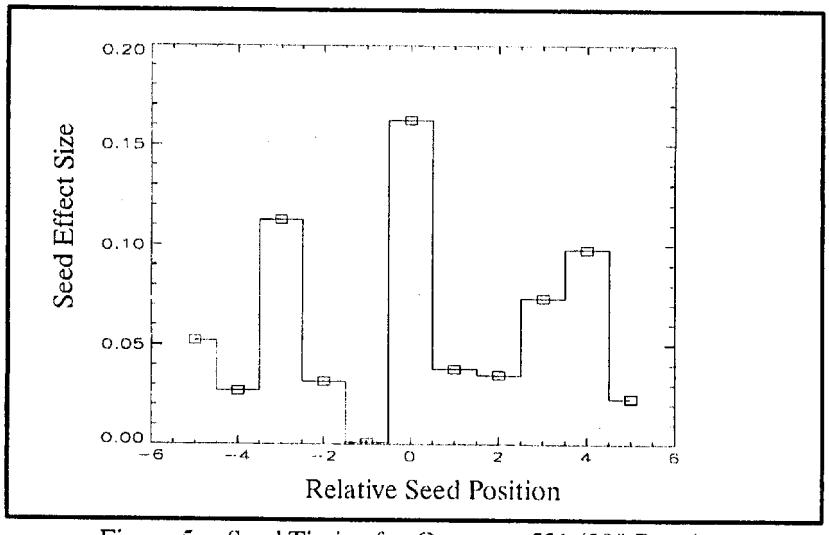

Figure 5. Seed Timing for Operator 531 (298 Runs).

Radin and May erroneously concluded that the significant differences between zero and adjacent seed positions was meaningful, and that the *DAT* ability was effective within 20 milliseconds. In fact, the situation shown in Figure 5 is *expected*. Differing from true random number generators in which slight changes in timing produce essentially the same sequence, PRNGs produced totally different sequences as a function of single digit seed changes. Thus, it would be surprising if the seed-shift display produced anything but a spike at seed shift zero. We will return to this point in our analysis of some of the published remarks on our theory.

From this prospective test of *DAT*, we conclude that for PRNGs it is possible to select a proper entry point into a bit stream to produce significant deviations from MCE that are independent of sequence length.

The Literature: Review and Comment

We have identified six published articles that have commented upon the Intuitive Data Sorting (IDS) theory, the earlier name for *DAT*. In this section, we chronologically summarize and comment on each report.

Walker – September 1987

In his first of two criticisms of IDS, Walker (1987) suggested that his Monte Carlo simulations did not fit the predictions of the IDS model. He generated a single deviant set of 100 bits (i.e., $Z = 2.33, p \leq 0.01$), and he inserted this same sequence as the first 100 bits of 400 otherwise randomly generated sequences ranging from 100 to 10^6 bits in length. Walker's analysis of these sequences did not yield a least square's slope of -0.5 as predicted under the *IDS* formalism. Walker concluded that the *IDS* theory was incorrect. Walker's sequences, however, are not the type that are generated in *AP* experiments or the type for which the *DAT/IDS* model is valid.

May et al. (1985) were explicit about the character of the sequences that fit the IDS model. Specifically, Walker quotes May et al. "*Using psi-acquired information, individuals are able to select locally deviant subsequences from a large random sequence.*" (Italics are used in the original May paper.) The very next sentence on page 249 of the reference says, "Such an ability, if mediated by precognition, would allow individuals (subjects or experimenters) to initiate a collection unit of continuous samples (this has been reported as a trial, a block, a run, etc.) in such a way as to *optimize the final result.* (Italics added here for emphasis.) Walker continued, "Indeed, the only way the subject can produce results that agree with the data is to wait for an extra-chance run that matches the experimental run length." In the final analysis, Walker actually supported our contention that individuals select deviant subsequences. Both from our text and the formalism in our 1985 paper, it is clear that what we meant by a "large random sequence," was large compared to the trial length, n.

In his second criticism of IDS in the same paper, Walker proposed that individuals would have to exhibit a physiologically impossible control over timing (e.g., when to press a button). As evidence apparently in favor of such an exquisite timing ability, he referred to the data presented by Radin and May (1986) that we have discussed above. (Please see Figure 5.) Walker suggested that Radin and May's result, therefore, supported his quantum mechanical observer theory. It is beyond the scope of this paper to critique Walker's quantum mechanical models, but we would hope they do *not* depend upon his analysis of Radin and May's results. The enhanced hitting at zero seed and the suppressed values $\pm$ one 20 ms clock tick that we show in Figure 5 is the expected result based upon the well-understood properties of PRNG's and does not represent the precision of the operator's reaction time.

We must consider how it is possible with normal human reactions to obtain significant scores, which can only happen in 20 ms windows. In typical visual reaction time measurements, Woodworth and Schlosberg (1960) found a standard deviation of 30 ms. If we assume these human reactions are typical of those for *AC* performance and are normally distributed, we compute the *maximum* probability of being within a 20 ms window, which is centered about the mean, of *23.5%*. For the worst case, the operators must "hit" significant seeds less often than 23.5% of the time. Radin and May do not report the number of significant runs, so we provide a worst-case estimate. Given that they quote a p-value of *0.005* for *500* trials, we find that 39 trials must be independently significant. That is, the accumulated binomial probability is *0.005* for *39* hits in *500* trials with an event probability of *0.05*. This corresponds to a hitting rate (i.e., 39/500) of only 7.8%, a value well within the capability of human reaction times. We recognize that it is not a requirement to hit only on significant seeds; however, all other seeds leading to positive Z-scores are less restrictive then the case we have presented.

The zero-center "spike" in Figure 5 misled Walker and others into thinking that exceptional timing was required to produce the observed deviations. As we have shown this is not the case, and, therefore, Walker's second criticism of the *IDS* theory is not valid.

Bierman – 1988

Bierman (1988) attempted to test the *IDS* model with a gifted subject. His experimental design appeared to meet most of the criteria for a valid test of the model; however, Bierman found no evidence for *AMP* and stated that no conclusions could be drawn from his work. We encourage Bierman to continue with this design and to be specific with what he would expect to see if *DAT* were the correct mechanism compared to if it were not.

Braud and Schlitz – 1989

Braud and Schlitz (1989) conducted an electrodermal *PK* experiment specifically to test the *IDS* model. They argued that if the mechanism of the effect were informational, then allowing participants more opportunities to select locally deviant values of the dependent variable should yield stronger effects. In their experiment, 12 electrodermal sampling epochs were either initiated individually by a press of a button, or all 12 were determined as a result of the first button press. Braud and Schlitz hypothesized that under *IDS*, they would expect to see a larger overall effect in the former condition. They found that the single button press data yielded a significant result; whereas the multiple press data scored at chance ($t_{single}[31] = 2.14$, $t_{multi}[31] = -0.53$). They concluded, therefore, that their data were more consistent with an influence mechanism than with an informational one.

In both button-press conditions, the dependent variable was averaged over all 12 epochs; therefore, the formalism discussed in this paper cannot be applied because the data should have been averaged over at least two different values. The idea of multiple decision points, however, is still valid. As stated in their paper, the timing of the epochs was such that 20 seconds of the 30 second epoch was independent of the button-press condition and could not, therefore be subjected to a *DAT*-like selection. To examine the consequence of this overlap, we computed the effect size for the single button case as *0.359* (Rosenthal, 1991, Page 19, Equation 2.16). Since data for the 20 seconds is the same in each condition, the operator can only make *AC*-mediated decisions for the first 10 seconds of the data. If we assume that on the average the remaining 20 seconds meets mean chance expectation and the effect size is constant then we would expect an effect size of $(0.359 + 0 + 0) / 3 = 0.119$.* The measured effect size was *0.095*, which is consistent with this prediction.

Braud and Schlitz's idea was good and provides a possible way to use *DAT* effectively. Because of the epoch timing and the consistency of the effect sizes, however, we believe they have interpreted their results in favor of causal mechanism prematurely. Aside from the timing issues, their protocol complicates the application of *DAT* further. To observe an enhanced effect because of multiple decision points, *Z*-scores should be computed for each decision and combined as a Stouffer's *Z* where the denominator is the square root of the number of decision points. In their protocol, they only combine the dependent variable.

Vassy – 1990

Vassy (1990) used a similar timing argument to refute the *IDS* model as did Walker (1987). Vassy generated PRNG single bits at a rate of one each 8.7 ms. He argued that if *IDS* were operating, that a subject would be more likely to identify bursts of ones rather than single ones given the time between consecutive bits. While he found significant evidence for the primary task of "selecting" individual bits, he found no evidence that these hits were imbedded in excess clusters of ones.

We compute that the maximum probability of a hit within an *8.7* ms window centered on the mean of the normal reaction curve with a standard deviation of *30* ms (Woodworth and Schlosberg, 1960) is *11.5%*. Vassy quotes an overall *Z*-score for *100* runs of *2.39*. From this, we compute a mean *Z* of *0.239* for each run of 36 bits. To obtain this result requires an *excess* hitting of *0.717* bits, which corresponds to an ex-

* As a function of n, *DAT* predicts a $1/\sqrt{n}$ dependency in the effect size; however, at a fixed n, as in this case, the effect size should be constant.

cess hitting rate of *2%*. Given that *11.5%* is the maximum one can expect with normal human reaction times, Vassy's results easily allow for individual bit selection, and, thus, cannot be used to reject the *DAT* model on the basis of timing.

Braud – 1990

In a cooperative effort with SRI International, Braud (1990) conducted a biological *AP* study with human red blood cells as the target system. The study was designed, in part, as a prospective test of *DAT*, so all conditions for a valid test were satisfied. Braud found that a significant number of individuals were independently able to "slow" the rate of hemolysis (i.e., the destruction of red blood cells in saline solution) in what he called the "protect" mode. Using data from the nine significant participants, Braud found support in favor of *PAP* over *DAT*. Figure 6 shows the results of our re-analysis of all of Braud's raw data using our more modern formalism of *DAT*.

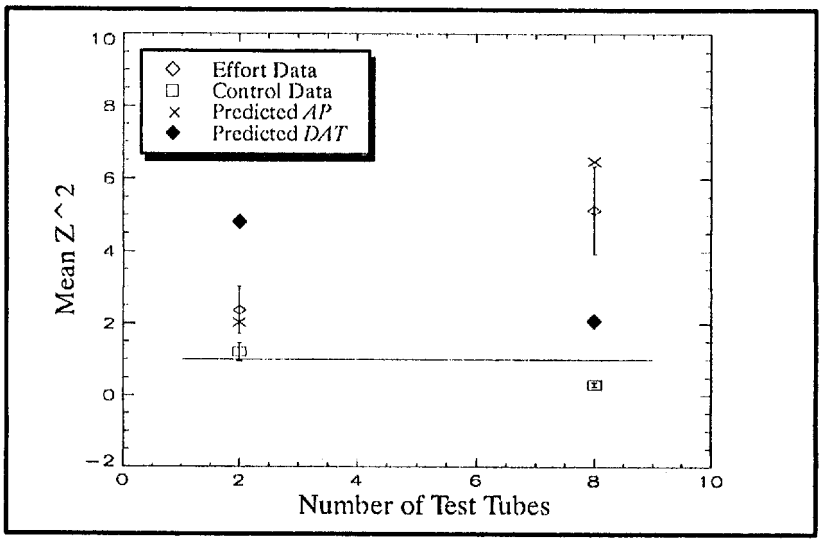

Figure 6. *DAT* Analysis of Hemolysis Data.

The solid line indicates the theoretical expected value for MCE. The squares are the mean values of Z^2 for the control data, and the error bars indicate the 1-standard error for the 32 trials in the study. We notice that the control data with eight test tubes is significantly below MCE ($t = -2.79$, $df = 62$, $p \leq 0.996$). Compared to the MCE line, the effort data is significant ($t = 4.04$, $df = 31$, $p \leq 7.6 \times 10^{-5}$) for eight test tubes and nearly so for $n = 2$ ($t = 2.06$, $df = 31$, $p \leq 0.051$). The $\times$ at $n = 8$ indicates the calculated value of the mean of Z^2 assuming that the effect size at $n = 2$ was entirely because of *AP*; similarly, the $\times$ at $n = 2$ indicates the calculated value assuming that the effect size, which was observed at $n = 8$, was totally due to *AP*. These *AP* predictions are not significantly different from the observed data ($t = 0.156$, $p \leq 0.431$, $df = 62$ and $t = 0.906$, $p \leq 0.184$, $df = 62$, at $n = 2$ and 8, respectively). Whereas *DAT* predicts no differences between the data at the end points for n, we find a significant difference ($t = 2.033$, $p \leq 0.023$, $df = 62$). That is, to a statistical degree the data at $n = 8$, cannot be explained by selection alone. Thus, we concur with Braud's original conclusion; these results indicate a possible causal relationship between mental intent and biological consequences.

It is difficult to conclude from our analysis of a single study with only 32 trials that *AP* is part of nature; nonetheless, this result is very important. It is the first data we have encountered that supports the *PAP*

hypothesis, which is in direct opposition to the substantial support against *PAP* resulting from the analysis of the RNG data sets. In addition, May and Vilenskaya (1993) and Vilenskaya and May (1994) report that the preponderance of *AMP* research in the Former Soviet Union is the study of *AP* on biological systems. Their operators, as do ours, report their internal experiences as being a causal relationship between them and their biological targets.

Dobyns – 1993

Dobyns (1993) presents a method for comparing what he calls the "influence" and "selection" models, corresponding to what we have been calling *DAT* and *PAP*. He uses data from 490 "tripolar sets" of experimental runs at PEAR. For each set, there was a high aim, a baseline and a low aim condition.

The three values produced were then sorted into which one was actually highest, in the middle, and lowest for each set. The data were then summarized into a 3×3 matrix, where the rows represented the three intentions, and the columns represented the actual ordering. If every attempt had been successful, the diagonal of the matrix would consist of the number of tripolar sets, namely 490. We present the data portion of Dobyns' Table from page 264 of the reference as our Table 2:

Table 2.

Scoring Data From Dobyns (1993)

Actual	Intention			
	High	Middle	Low	Total
High	180	167	143	490
Baseline	159	156	175	490
Low	151	167	172	490
Total	490	490	490	

Dobyns computes an aggregate likelihood ratio of his predictions for the *DAT* and *PAP* models and concludes in favor the the influence model with a ratio of 28.9 to one.

However, there are serious problems with the methods used in Dobyns' paper. In this paper we simply outline only two of the difficulties. To fully explain them would require a level of technical discussion not suitable for a short summary such as this.

One problem is in the calculation of the likelihood ratio function using Equation 6, which we reproduce from page 265 of the reference:

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

where p and q are the predicted rank frequencies for each aim under the influence and selection models, respectively, and the n are the observed frequencies for each aim. We agree that this relationship correctly gives the likelihood ratio for comparing the two models for one row of Table 2. However, immediately following that equation, Dobyns writes, "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 will simply be the product of factors such as (6) above for each of the three intentions."

That statement is simply incorrect. A combined likelihood is found by multiplying the individual likelihoods only if the random variables are independent of each other (DeGroot, 1986, p. 145). Clearly, the rows of the table are not independent. In fact, if you know any two of the rows, the third is determined exactly! The correct likelihood ratio needs to build that dependence into the formula.*

A second technical problem with the conclusion that the data support the influence model is that the method itself strongly supports the influence model. As noted by Dobyns, "In fact, applying the test to data sets 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 (page 268)." The actual data in his paper yielded odds of 28.9 to one in favor of the influence model; however, this value is well within the reported limits from his "influence-less" Monte Carlo data.

Under *DAT* it is possible that *AC*-mediated selection might occur at the protocol level, but the primary way is through timing—initiating a run to capitalize upon a locally deviant subsequence. How this might work in dynamic RNG devices is clear; wait until such a deviant sequence is in your immediate future and initiate the run in time to capture it. With "static" devices, such as PEAR's random mechanical cascade (RMC) device, how timing enters in is less obvious. Under closer inspection, however, even with the RMC device there is a statistical variation among unattended control runs. That is, there is never a series of control runs that give exactly the same mean. Physical effects, such as Browian motion, temperature gradients, etc., can account for the observed variance in the absence of human operators. Thus, *when* a run is initiated to capture favorable local "environmental" factors, even for "static" devices, remains the operative issue with regard to *DAT*. Dobyns does not consider this case at all in his analysis. If *DAT* enters in at the protocol selection, as it probably does, it is likely to be a second-order contribution because of the limited possibilities from which to select (i.e., six in the tripolar case).

Finally, a major problem with Dobyns' conclusion, which was pointed out when he first presented this paper at a conference (May, 1990), is that the likelihood ratio supports the influence model even for their PRNG data. Dobyns dismisses this finding (page 268) all too easily given the preponderance of evidence that suggest that no influence occurs during PRNG studies (Radin and May, 1986).

Aside from the technical flaws in Dobyns' likelihood ratio arguments, and even ignoring the problem with the PRNG analysis, we reject his conclusions simply because they hold in favor of influence even in Monte Carlo-constructed and *unperturbed* data.

Circumstantial Evidence Against an *AP* Model for RNG Data

Experiments with hardware RNG devices are not new. In fact, the title of Schmidt's very first paper on the topic (1969) portended our conclusion, "Precognition of a Quantum Process." Schmidt lists *PK* as a third option after two possible sources for precognition, and remarks, "The experiments done so far do not permit a distinction (if such a distinction is at all meaningful) between the three possibilities." From

* Dobyns agrees on this point—private communication.

1969 onward, the RNG research has been strongly oriented toward a *PK* model. The term *micro-PK*, itself, embeds the force concept further into the lexicon of RNG descriptions.

In this section, we examine a number of RNG experimental results that provide circumstantial evidence against the *AP* hypothesis. Any single piece of evidence could be easily dismissed; however, taken together, they demonstrate a substantial case against *AP*.

Internal Complexity of RNG Devices and Source Independence

Schmidt (1974) conducted the first experiment to explore potential dependencies upon the internal workings of his generators. Since by definition *AP* implies a force or influence, it seemed reasonable to expect that an influence should depend upon the details of the target system. In this study, one generator produced individual binary bits, which were derived from the β-decay of ^{90}Sr, while the other "binary" output was a majority vote from 100 bits, each of which were derived from a fast electronic diode. Schmidt reports individually significant effects with both generators, yet does not observe a significant difference between the generators.

This particular study is interesting, quite aside from the timing and majority vote issues; the binary streams were derived from fundamentally different physical sources. Radioactive β-decay is governed by the weak nuclear force, and electronic devices (e.g., noise diodes) are governed by the electromagnetic force. Schematically speaking, the electromagnetic force is approximately 1,000 times as strong as the weak nuclear force, and modern high-energy physics has shown them to be fundamentally different after about 10^{-10} seconds after the big bang (Raby, 1985). Thus, a putative *AP*-force would have to interact equally with these two forces; and since there is no mechanism known that will cause the electromagnetic and weak forces to interact with each other, it is unlikely that *AP* will turn out to be the first coupling mechanism. The lack of difference between β-decay and noise diode generators was confirmed years later by May et al. (1980).

We have already commented upon one aspect of the timing issue with regard to Radin and May's (1986) experiment and the papers by Walker (1987) and Vassy (1990). May (1975) introduced a scheme to remove any first-order biases in binary generators that also is relevant to the timing issue. The output of his generator was a match or anti-match between the random bit stream and a target bit. One mode of the operation of the device, which May describes, included an oscillating target bit—one oscillation per bit at approximately 1 MHz rate.* May and Honorton (1975) and Honorton and May (1975) reported significant effects with the RNG operating in this mode. Thus, significant effects can be seen even with devices that operate in the microsecond time domain, which is three orders of magnitude faster than any known physiological process.

Effects with Pseudorandom Number Generators

Pseudorandom number generators are, by definition, those that depend upon an algorithm, which is usually implemented on a computer. Radin (1985) analyzed all the PRNGs commonly in use and found that they require a starting value (i.e., a seed), which is often derived from the computer's system clock. As we noted above, Radin and May (1986) showed that the bit stream, which proved to be "successful" in a PRNG study, was bit-for-bit identical with the stream, which was generated later, but with the same

* Later, this technique was adopted by Jahn (1982) for use in the RNG devices at PEAR.

seed. With that generator, at least, there was no change from the expected bit stream. Perhaps it is possible that the seed generator (i.e., system clock) was subjected to some *AP* interaction. We propose two arguments against this hypothesis:

(1) Even one cycle interruption of a computers' system clock will usually invoke a system crash; an event not often reported in PRNG experiments.

(2) Computers use crystal oscillators as the basis for their internal clocks. Crystal manufacturers usually quote errors in the stated oscillation frequency of the order of 0.001 percent. That translates to 500 cycles for a 50 MHz crystal, or to 10 μs in time. Assuming that the quoted error is a 1-σ estimate, and that a putative *AP* interaction acts at within the ± 2-σ domain, then shifting the clock by this amount might account for only one seed shift in Radin and May's experiment. By Monte Carlo methods, we determined that, given a random entry into seed-space, the average number of ticks to reach a "significant" seed is 10; therefore, even if *AP* could shift the oscillators by 2-σ, it cannot account for the observed data.

Since computers in PRNG experiments are not reported as "crashing" often, it is safe to assume that PRNG results are only due to *AC*. In addition, since the results of PRNG studies are statistically inseparable from those reported with true RNGs, it is also reasonable to assume that the mechanisms are similarly *AC*-based.

Precognitive *AC*

Using the tools of modern meta-analysis, Honorton reviewed the precognition card-guessing database (Honorton and Ferarri, 1989). This analysis included 309 separate studies reported by 62 investigators. Nearly two million individual trials were contributed by more the 50,000 subjects. The combined effect size was $\bar{\varepsilon} = 0.020 \pm 0.002$, which corresponds to an overall combined effect of 11.4σ. Two important results emerge from Honorton's analysis. First, it is often stated by critics that the best results are from studies with the least methodological controls. To check this hypothesis, Honorton devised an eight-point quality measure (e.g., automated recording of data, proper randomization techniques) and scored each study with regard to these measures. There was no significant correlation between study quality and study score. Second, if researchers improved their experiments over time, one would expect a significant correlation of study quality with date of publication. Honorton found $r = 0.246$, $df = 307$, $p \leq 2 \times 10^{-7}$. In brief, Honorton concludes that a statistical anomaly exists in this data that cannot be explained by poor study quality or a large variety of other hypotheses; therefore, a potential mechanism underlying *DAT* has been verified.

SRI International's RNG Experiment

May, Humphrey, and Hubbard (1980) conducted an extensive RNG study at SRI International in 1979. They applied state-of-the-art engineering and methodology to construct two true RNGs, one based on the β-decay of ^{137}Pm and the other based on an MD-20 noise diode from Texas Instruments. It is beyond the scope of this paper to describe, in detail, the intricacies of this experiment; however, we will discuss those aspects, which are pertinent to this discussion.

Technical Details

Each of the two sources were battery operated and optically coupled to a Digital Equipment Corporation LSI 11/23 computer. Fail-safe circuitry would disable the sources if critical physical parameters

(e.g., battery voltages and currents, temperature) exceed preset ranges. Both sources were subjected to environmental testing which included extreme temperature cycles, vibration tests, E&M and nuclear gamma and neutron radiation tests. Both sources behaved as expected, and the critical parameters, such as temperature, were monitored and their data stored along with the experimental data.

A source was sampled at 1 KHz rate. After eight milliseconds, the resulting byte was sent to the computer while the next byte was being obtained. In this way, a continuous stream of 1 ms data was presented to the computer. May et al. had specified, in advance, that bit number 4 was the designated target bit. Thus each byte provided 3 ms of bits prior to the target and 4 ms of bits after the target bit.

A trial was defined as a definitive outcome from a sequential analysis of bit four from each byte. In exchange for not specifying the number of samples in advance, sequential analysis requires that the Type I and Type II errors, and the chance and extra-chance hitting rate be specified in advance. In May et al.'s two-tailed analysis, $\alpha = \beta = 0.05$ and the chance and extra-chance hitting rate was 0.50 and 0.52, respectively. The expected number of samples to reach a definitive decision was approximately 3,000. The outcome from a single trial could be in favor of a hitting rate of 0.52, 0.48, or at chance of 0.50, with the usual risk of error in accordance with the specified Type I and Type II errors.

Each of seven operators participated in 100 trials of this type. For an operator's data to reach independently statistical significance, the operator had to produce 16 successes in 100 trials, where a success was defined as extra-chance hitting (i.e., the exact binomial probability of 16 successes for 100 trials with an event probability of 0.10 is 0.04 where one less success is not significant). Two operators produced 16 and 17 successful trials, respectively.

Temporal Analysis

We analyzed the 33 trials from the two independently significant operators from May et al.'s experiment. Each of the 33 trials consisted of approximately 3,000 bits of data with -3 bits and $+4$ bits of 1 ms/bit temporal history surrounding the target bit. We argue that if the significance observed in the target bits was because of AP, we would expect a large correlation with the target bit's immediate neighbors, which are only ± 1 ms away. As far as we know, there is no known physiological process that can be cognitively, or in any other way, manipulated within a millisecond. We might even expect a 100% correlation under the complete AP model.

We computed the linear correlation coefficients between bits 3 and 4, 4 and 5, and 3 and 5. For binary data:

$$N\phi^2 \sim X^2(df = 1),$$

where ϕ is the linear correlation coefficient and N is the number of samples. Since we examined three different correlations for 33 trials, we computed 99 different values of $N\phi^2$. Four of them produced X^2s that were significant—well within chance expectation. The complete distribution is shown in Figure 7. We see that there is excellent agreement of the 99 correlations with the X^2 distribution for one degree of freedom, which is shown as a smooth curve.

We conclude, therefore, that there was no evidence beyond chance to suggest that the target bit neighbors were affected even when the target bit analysis produced significant evidence for an anomaly. So,

knowing the physiological limitations of the human systems, we further concluded that the observed effects could not have arisen due to a human-mediated force (i.e., AP).

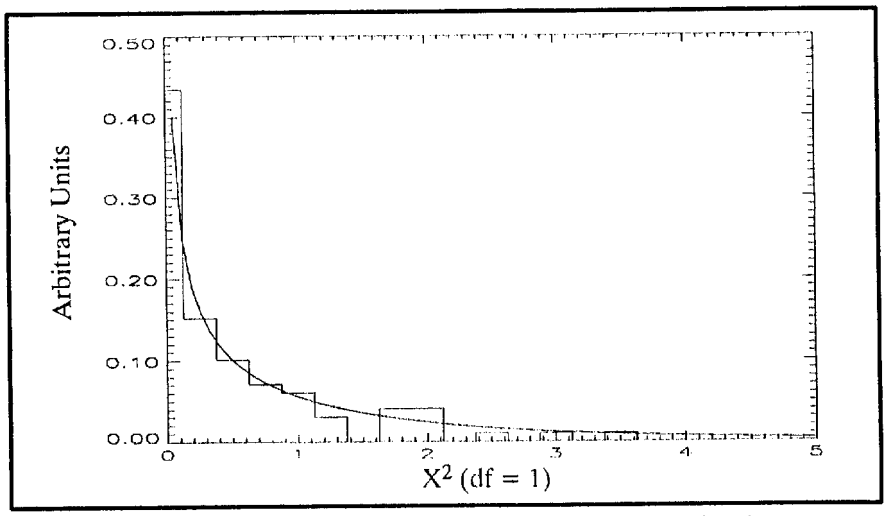

Figure 7. Observed and Theoretical Correlation Distributions.

Mathematical Model of the Noise Diode

Because of the unique construction parameters of Texas Instrument's MD-20 noise diode, May et al. were able to construct a quantum mechanical model of the detailed workings of the device. This model contained all known properties of the material and it's construction parameters. For example, the band gap energy in Si, the effective mass of an electron or hole in the semiconductor, and the impurity concentration were among the parameters for the model. The model was successful at calculating the diode's known and measured behavior as a function of temperature. May et al. were able to simulate their RNG experiment down to the quantum mechanical details of the noise source. They hoped that by adjusting the model's parameters so that the computed output agreed with the experimental one, that they could gain insight as to where the causal influence "entered" the device.

May et al. were not able to find a set of model parameters that mimicked their RNG data. For example, changing the band gap energy for short periods of time; increasing or reducing the electron's effect mass; or redistributing or changing the impurity content produced no unexpected changes in the device output. The only device behavior that could be effected was its known function of temperature.

Because of the construction details of the physical RNG, this result could have been anticipated. The changes that could be simulated in the model were all in the microsecond domain because of the details of the device. Both with the RNG and in its model, the diode's multi-MHz output was filtered by a 100-KHz wide bandwidth filter. Thus, any microsecond changes would not pass through the filter. In short, because of this filtering, the RNG was particularly insensitive to potential changes of the physical parameters of the diode.

Yet solid statistical evidence for an anomaly was seen by May et al. Since the diode device was shown mathematically and empirically to be insensitive to environmental and physical changes, these results must have been as a result of AC rather than any causal AP. In fact, this observation coupled with the bit

timing argument, which we have described above, lead May et al. to question causality in RNG studies in general.

Summary of Circumstantial Evidence Against *AP*

We have identified six circumstantial arguments that, when taken together, provide increasingly difficult requirements that must be met by a putative *AP* force. In summary, the RNG database demonstrates that:

(1) Data are independent of internal complexity of the hardware RNG device.

(2) Data are independent of the physical mechanism producing the randomness (i.e., weak nuclear or electromagnetic).

(3) Effects with pseudorandom generators are statistically equivalent to those observed with true hardware generators.

(4) Reasonable AP models of mechanism do not fit the data.

(5) In one study, bits which are $\pm$ 1 ms from a "perturbed" target bit are themselves unperturbed.

(6) A detailed model of a diode noise source, which includes all known physics of the device, could not simulate the observed data streams.

In addition, *AC*, which is a mechanism to describe the data, has been confirmed in non-RNG experiments. We conclude, therefore, an *AP* force that is consistent with the database must

- Be equally coupled to the electromagnetic and weak nuclear forces.
- Be mentally mediated in times shorter than one millisecond.
- Follow a $1/\sqrt{n}$ law.

For these to be true, an *AP* force would be at odds with an extensive amount of verified physics and common behavioral observables. We are *not* saying, therefore, that *AP* cannot exist; rather, we are suggesting that instead of having to force ourselves to invent a whole new science, we should look for ways in which *AP* might fit into the present world view. In addition, as *DAT* tries to accomplish, we should invent non-causal and testable alternate mechanisms for the experimental observables.

Conclusions

We have shown that *DAT* can determine whether a causal or informational explanation is more consistent with a given set of anomalous statistical data. In applying *DAT* to the substantial physical RNG database, we showed that an informational mechanism is strongly favored over a large class of perturbational ones. Given that one possible information mechanism (i.e., precognitive *AC*) can, and has been, independently confirmed in the laboratory, and given the weight of the empirical, yet circumstantial, arguments taken together against *AP*, we conclude that the anomalous results from the RNG studies arise not because of a mentally mediated force, but rather because of a human ability to be a mental opportunist by making *AC*-mediated decisions to capitalize on the locally deviant circumstances.

Our recent results in the study of anomalous cognition (May, Spottiswoode, and James, 1994) suggest the the quality of AC is proportional to the change in Shannon entropy. Following Vassy (1990), we compute the change in Shannon entropy for an extra-chance, binary sequence of length n. The total change of entropy is given by:

$$\Delta S = S_0 - S,$$

where for an unbiased binary sequence of length n, $S_o = n$, and S is given by:

$$S = -np_1 log_2 p_1 - n(1 - p_1) log_2(1 - p_1).$$

Let $p_1 = 0.5\ (1 + \varepsilon)$ and assume that ε, the effect size, is small compared to one (i.e., typical RNG effect sizes are of the order of 3×10^{-4}). Using the approximation:

$$ln(1 + \varepsilon) = \varepsilon - \frac{\varepsilon^2}{2},$$

we find that S is given by:

$$S = n - n\frac{\varepsilon^2}{2ln2},$$

or that the total change of entropy for a biased binary sequence is given by;

$$\Delta S = S_0 - S = n\frac{\varepsilon^2}{2ln2}.$$

Since our analysis of the historical RNG database shows that the effect size is proportional to $1 / \sqrt{n}$ (i.e., Z-score is independent of n), the total change of Shannon entropy becomes a constant that is independent of the sequence length. Thus, if entropy is related to what is being sensed by anomalous cognition, then this argument suggests that informational processes are responsible for the RNG anomaly.

The one exception to this conclusion is Braud's study of *AP* on red blood cells. It may be that there is something unique about living systems that can account for this observation. On the other hand, it is the only biological *AP* study we could find that could be analyzed by *DAT*, and the perturbation hypothesis is only weakly favored over the selection one (i.e., $p \leq 0.023$). Before we would be willing to declare that *AP* is a valid mechanism, more than a single, albeit well designed and executed, study is needed.

Generally, we suggest that future RNG, PRNG, and biological *AP* studies be designed in such a way that the criteria, as outlined in this paper and in May, Utts, Spottiswoode (1994), are adhered to for a valid *DAT* analysis. Our discipline has evolved to the point where we can no longer be satisfied with yet one more piece of evidence of a statistical anomaly. We must identify the sources of variance as suggested by May, Spottiswoode, and James (1994); limit them as much as possible; and apply models, such as *DAT*, which can begin to shed light on the physical, physiological, and psychological mechanisms of anomalous mental phenomena.

References

Bierman, D. J. (1988). Testing the IDS model with a gifted subject., *Theoretical Parapsychology,* **6**, 31-36.

Braud, W. G. and Schlitz, M. J. (1989). Possible role of Intuituve Data Sorting in electrodermal biological psychokinesis (Bio-PK). *The Journal of the American Society for Psychical Research,* **83**, No. 4, 289-302.

Braud, W. G. (1990). Distant mental influence of rate of hemolysis of human red blood cells. *The Journal of the American Society for Psychical Research,* **84**, No. 1, 1-24.

DeGroot, Morris H. (1985). *Probability and Statistics, 2nd Edition*. Reading, MA: Addison-Wesley Publishing Co.

Dobyns, Y. H. (1993). Selection versus influence in remote REG anomalies. *Journal of Scientific Exploration*. **7**, No. 3, 259-270.

Honorton, C. and May, E. C. (1975). Volitional control in a psychokinetic task with auditory and visual feedback. *Research in Parapsychology*, 1975, 90-91.

Honorton, C. and Ferrari, D. C. (1989) "Future Telling:" A meta-analysis of forced-choice precognition experiments, 1935-1987. *Journal of Parapsychology*, **53**, 281-308.

Jahn, R. G. (1982). The persistent paradox of psychic phenomena: an engineering perspecitve. *Proceedings of the IEEE.* **70**, No. 2, 136-170.

Lewis, T. G. (1975). Distribution Sampling for Computer Simulation. Lexington, MA: Lexington Books.

May, E. C. (1975). PSIFI: A physiology-coupled, noise-driven random generator to extend PK studies. *Research in Parapsychology*, 1975, 20-22.

May, E. C. and Honorton, C. (1975). A dynamic PK experiment with Ingo Swann. *Research in Parapsychology*, 1975, 88-89.

May, E. C., Humphrey, B. S., Hubbard, G. S. (1980). Electronic System Perturbation Techniques. *Final Report.* SRI International Menlo Park, CA.

May, E. C., Radin, D. I., Hubbard, G. S., and Humphrey, B. S. (1985). Psi experiments with random number generators: an informational model. *Proceedings of Presented Papers Vol I.* The Parapsychological Association 28th Annual Convention, Tufts University, Medford, MA, 237-266.

May, E. C. (1990). As chair for the session at the annual meeting of the Society for Scientific Exploration in which this original work was presented, I pointed out the problem of the likelihood ratio for the PRNG data from the floor of the convention.

May, E. C., Spottiswoode, S. James P., and James, C. L. (1994). Shannon entropy as an Intrinsic Target property: Toward a reductionist model of anomalous cognition. Submitted to *The Journal of Parapsychology*.

May, E. C., Utts, J. M., Spottiswoode, S. J. (1994). Decision augmentation theory: Toward a model of anomalous mental phenomena. Submitted to *The Journal of Parapsychology*.

May, E. C. and Vilenskaya, L. (1994). Overview of current parapsychology research in the former Soviet union. *Subtle Energies.* **3**, No 3. 45-67.

Nelson, R. D., Jahn, R. G., and Dunne, B. J. (1986). Operator-related anomalies in physical systems and information processes. *Jorunal of the Society for Psychical Research*, **53**, No. 803, 261-285.

Nelson, R. D., Dobyns, Y. H., Dunne, B. J., and Jahn, R. G., and (1991). Analysis of Variance of REG Experiments: Operator Intention, Secondary Parameters, Database Structures. PEAR Laboratory Technical Report 91004, School of Engineering, Princeton University.

Raby, S. (1985). Supersymmetry and cosmology. in *Supersymmetry, Supergravity, and Related Topics*. Proceedings of the XVth GIFT International Seminar on Theoretical Physics, Sant Feliu de Guixols, Girona, Spain. World Scientific Publishing Co. Pte. Ltd. Singapore, 226-270.

Radin, D. I. (1985). Pseudorandom Number Generators in Psi Research. *Journal of Parapsychology.* **49**, No 4, 303-328.

Radin, D. I. and May, E. C. (1986). Testing the Intuitive Data Sorting mode with pseudorandom number generators: A proposed method. The Proceedings of Presented Papers of the 29th Annual Convention of the Parapsychological Association, Sonoma State University, Rohnert Park, CA, 539-554.

Radin, D. I. and Nelson, R. D. (1989). Evidence for consciousness-related anomalies in random physical systems. *Foundations of Physics.* **19**, No. 12, 1499-1514.

Rosenthal, R. (1991). Meta-analysis procedures for social research. Applied Social Research Methods Series, Vol. 6, Sage Publications, Newbury Park, CA.

Schmidt. H. (1969). Precognition of a quantum process. *Journal of Parapsychology*. **33**, No. 2, 99-108.

Schmidt. H. (1974). Comparison of PK action on two different random number generators. *Journal of Parapsychology*. **38**, No. 1, 47-55.

Vassy, Z. (1990). Experimental study of precognitive timing: Indications of a radically noncausal operation. *Journal of Parapsychology*. **54**, 299-320.

Vilenskaya, L. and May, E. C. (1994). Anomalous mental phenomena research in Russia and the Former Soviet Union: A follow up. Submitted to the 1994 Annual Meeting of the Parapsychological Association.

Walker, E. H. (1987). A comparison of the intuitive data sorting and quantum mechanical observer theories. *Journal of Parapsychology,* **51**, 217-227.

Woodworth, R.S. and Schlosberg H. (1960). *Experimental Psychology.* Rev ed. New York Holt. New York.