

Overview of Several Theoretical Models on PEAR Data

YORK H. DOBYNS

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

Abstract—Existing PEAR datasets on microelectronic experiments are reviewed for their consistency with several theoretical models. The analysis includes a comparison of the observed data with predictions from bitwise-effect (BIT), Decision Augmentation Theory (DAT), time-normalized (TIM), and teleological (TEL) models for the phenomena. Methods for constructing a model comparison are discussed, together with the merits of relative *versus* absolute tests and the sensitivity of various test representations. Final results are presented in both frequentist (p -value) and Bayesian (hypothesis odds) formats.

Keywords: human/machine interaction—Decision Augmentation Theory—modeling methods—analysis methods

1. Terminology

The following terms will be encountered frequently:

Z score. This is a normalized score used as a statistical figure-of-merit for the degree to which an observation differs from a theoretical prediction. In general, a Z score is defined by $Z = (x_o - x_t)/\sigma_t$, where x_o is the observed value, x_t is the theoretically predicted value, and σ_t is the theoretical uncertainty (standard deviation) of the measurement. If the measurement error is normally distributed and the theory correctly describes the experimental situation, one expects Z to follow the standard normal distribution with mean 0, variance 1. A large value of Z is evidence against the theoretical model under which it is calculated. When different models make different predictions for the same observation, a Z score can be calculated for each model.

Effect size. If some number N of identical, independent observations is made, the mean of those observations has a theoretical standard deviation σ_t/\sqrt{N} , where σ_t is the theoretical standard deviation of a single observation. Therefore the Z score of the collected database will tend to grow as \sqrt{N} , provided the same effect (*i.e.*, the same quantitative degree of departure from the theoretical prediction) is present in each observation. This means the Z score is not directly useful for comparisons between experiments because differences in Z may simply reflect different amounts of data in situations where the same phenomenon is observed. To get around this problem one customarily defines

an effect size $\epsilon = Z / \sqrt{N}$. It is obvious that ϵ will remain constant as N increases, provided the repeated observations remain mutually consistent.

Dataset and database. In general *database* refers to the entire collection of experimental data under consideration, while a *dataset* is a subset of the total database extracted according to some analytical criterion. For example, a parameter relevant to a particular theory, which takes on several discrete values in the database as a whole, may be used to segregate the data into several datasets for analysis.

2. Introduction

The current analysis focuses on a subset of data from PEAR published in the *Journal of Scientific Exploration* (Jahn *et al.*, 1997). Of those data, this analysis will examine only those obtained from microelectronic random sources; in the interests of a consistent database, we shall not include the data from deterministic sources or those from the macroscopic, mechanical experiments. A partial reproduction of Table 2 in Jahn *et al.* (1997), presenting the relevant data, is given as Table 1 below.

The primary consideration in an anomalies experiment is, of course, the question of whether an anomaly is present. Our main tool for examining this will be a Z score against the null-hypothesis prediction. However, different models of the nature of the phenomenon lead to different formulas for calculating the composite Z score across several datasets, for reasons that will be discussed below. Thus, when there are competing explanations, it may be necessary to evaluate them against each other before attempting to make a definitive, quantitative evaluation against the null hypothesis. We shall examine four hypotheses that have been proposed as models for the observed phenomena.

Any database of this sort is necessarily a combination of efforts from many individual experimental sessions. There are variables that cannot be controlled effectively in experiments in which the data must be generated by human volunteers at their own convenience. A single experiment combines data from many individuals with different personal characteristics, at different times of day and year, and so forth. In attempting to compare overall models of the phenomena, one must assume that such uncontrollable variables are not conse-

TABLE 1
Data by Experiment

Experiment	Series	Bits	Z
Original	522	3.35×10^8	3.809
Remote	212	1.83×10^8	2.214
Alternate source	46	4.94×10^7	2.765
Co-operator	45	3.62×10^7	1.635
REG2000	44	3.25×10^8	2.718
REG20	20	1.64×10^6	-0.956

quential to the phenomenon, or at least that their effects are averaged out over the accumulation of a large database. When we say that a model predicts that some measure or other should be “constant across all experiments,” we naturally mean this to include the assumption that there are no consequential factors that systematically change between one dataset and another. In fact, we will illustrate one instance in which the presence of such a confound can lead to an incorrect model evaluation if it is not properly taken into account.

All of the models discussed here are phenomenological; that is, they discuss *what* takes place under certain experimental conditions, without addressing the question of *how* it takes place. For example, the bitwise-effect model explains the anomalous observation as the result of a change in binary probabilities; that change is assumed, and no explanation is offered for how a human consciousness could alter the probability of a physical outcome. It seems advisable to compare such proposed models, even though they offer no deep theoretical insight, because it is unlikely that the deep questions can be addressed adequately while we are still uncertain about the phenomenology.

3. The Models

Bitwise effect. The observed phenomenon consists of a shift in the mean output value of a binary random number generator, in accordance with the human operator's prestated intention. This can be modeled as a change in the probability of the random process, which favors the outcome in accordance with the operator's intention. The bitwise effect model simply states that this is in fact what happens; the probability of a binary outcome is changed by some small amount in accordance with intention. The natural unit for the bitwise model is the bit. The effect size measure, under the bitwise model, is $Z/\sqrt{N_b}$, where N_b is the number of bits. The bitwise model predicts that this measure should be constant across experiments. For brevity, we will use the mnemonic BIT to refer to this model.

Decision Augmentation Theory. This theory, commonly abbreviated DAT even by its authors, suggests that no change to the performance of the experimental apparatus actually takes place (May, Utts, and Spottiswoode, 1995). Instead, they suggest that the human operator is able to foresee, unconsciously and by anomalous means, the outcome of an experimental run. The operator thus can choose a moment to start the run when random fluctuations will produce a favorable outcome. The natural unit for this model is the “DAT event” of a single start-of-data-collection by the operator. The effect size measure is $Z/\sqrt{N_D}$, where N_D is the number of DAT events in a single dataset; DAT predicts that this should be constant across experiments.

Time normalization. This model suggests that the critical variable in an anomalous experiment is the amount of time, or equivalently the amount of effort, contributed by the human operator (Nelson, 1994). The natural effect size measure is $Z/\sqrt{N_s}$, where N_s is the number of seconds spent generating data.

The time-normalization model predicts that this should be constant across experiments. The three-letter mnemonic we will use for this model is TIM.

Teleological. Several models have been discussed internally at PEAR that predict a decline over the course of a single experiment, specifically that the effect size should decline as $1/\sqrt{N_a}$, where N_a is the total amount of accumulated data in the experiment to date. These models are collectively called “teleological” since the first such proposal was based on teleology at the level of individual experiments: It proposed that any experiment rapidly attains a “characteristic” overall Z score, which then holds more or less steady no matter how much further accumulation of data takes place. Unlike the previous models, this one does not have a “natural” effect size associated with it; rather, it predicts that the *locally* measured effect should decline in a specific way as data accumulate. It also does not make predictions across experiments, but it can be evaluated only by examining the sequential evolution within a single experimental dataset. The mnemonic for this model is TEL.

4. Comparing Models

It is easiest to compare models in pairwise fashion. In general, such a comparison requires identifying a variable such that (a) data with different values of this variable are present in the database and (b) the models being compared make different predictions for the functional dependence of the effect on the variable.

We shall make all our pairwise comparisons between theoretical models by comparing the BIT model against another model. This is primarily a matter of convenience and historical preference: The original data collection systems used an implicit bitwise model in their assumptions, and this is still the most convenient representation for analyses.

For the comparison of BIT against DAT, the key variable is the *sequence length*, or number of bits per DAT event. The data under consideration have eight different sequence lengths, ranging from 20 to 200,000. We shall refer to this sequence length variable as n ; it is related to N_b , the number of bits in a given dataset, and N_D , the number of DAT events, by the obvious $N_b = nN_D$.

As noted above, BIT and DAT identify different increments of data as fundamental, and the specific functional dependence upon n will depend on which of these is used. BIT predicts that the bitwise effect size, $\epsilon_b = Z/\sqrt{N_b}$, is constant across datasets, while DAT predicts the same of the DAT effect size $\epsilon_D = Z/\sqrt{N_D}$. A comparison of these formulas will immediately show that $\epsilon_b = \epsilon_D/\sqrt{n}$. Thus, if we adopt ϵ_b as our measure, BIT predicts that the effect is constant, and DAT predicts that the effect varies as $1/\sqrt{n}$. Conversely, if we measure effects by ϵ_D , then DAT predicts constancy and BIT predicts an effect that increases as \sqrt{n} . We will have some further comments on this duality of representation later.

A precisely similar situation applies between BIT and TIM except that the key variable is now \mathcal{B} , the number of bits per second, forming the connection

$N_b = \mathcal{B}N_s$. As above, each model predicts that its own measure is constant; TIM predicts that $\epsilon_b \propto 1/\sqrt{\mathcal{B}}$, while BIT predicts $\epsilon_s \propto \sqrt{\mathcal{B}}$. For the comparison of BIT versus TIM, we have three available bit rates: 20, 200, and 2,000 bits per second. The data in the first and last of these categories are somewhat scanty, so we may expect that this issue will be resolved much less clearly than the previous comparison.

TEL is somewhat different in that it does not present an alternative fundamental unit for the local interpretation of data; rather, it predicts a specific course of historical change over the accumulation of data in a particular experiment. This prediction applies to any of the effect size measures discussed in the previous paragraphs. For a pairwise test of BIT versus TEL, the obvious choice is to use ϵ_b ; then BIT predicts that the value should be constant, while TEL predicts $\epsilon_b \propto 1/\sqrt{N_a}$.

Finally, there is the question of a null-hypothesis comparison. The null hypothesis simply predicts that there is no effect: $\epsilon_b = \epsilon_D = \epsilon_s = 0$. Nonetheless, because the different models weight contributing datasets differently, the confidence with which one can assert $\epsilon \neq 0$ differs from model to model. The procedure we will develop for the pairwise comparison between models includes, as a byproduct, the generation of a Z score against the null hypothesis for each model involved. Our ultimate candidate for an interpretive test against the null hypothesis should be the preferred model emerging from the pairwise comparisons.

5. Comparison Procedure

In each of the pairwise comparisons, we are confronted with a set of empirical data and a key independent variable: One model predicts that the data should be independent of that variable, and the other predicts that there should be a specific functional dependence. The simplest evaluative procedure at this point is a goodness-of-fit test. We know, for example, that BIT predicts $\epsilon_b = k_b$, while DAT predicts $\epsilon_b = k_D/\sqrt{n}$, where k_b and k_D are undetermined constants of proportionality. We can find empirical values for these constants by minimizing the total squared error between the model and the observed data, and then calculate the χ^2 of the residual differences between the observations and the fitted model.

This is not, however, the most sensitive test available in the current situation. A standard and often-used procedure for such tests is to linearize the model and then perform a linear regression to identify the slope. In this particular case, “linearizing” means not discarding higher-order terms, but constructing a new variable such that the model predictions are linear functions of that variable.

We will continue to use the BIT *versus* DAT comparison as an initial example. If we work in the ϵ_b representation, then the DAT model predicts a dependence on $1/\sqrt{n}$. Therefore, let us define $x = 1/\sqrt{n}$; then the BIT prediction

described above remains unchanged, and the DAT prediction becomes $\epsilon_b = k_D x$.

These models then can be compared directly with a weighted linear regression performed on the observations of ϵ_b versus x . Indeed, because we must perform a least-squares fit to find the empirical constants of proportionality, each of the model fits is a special case of a constrained linear regression. The general linear regression will find parameters b_1 , a slope, and b_0 , an intercept, such that the line $y = b_0 + b_1 x$ minimizes the overall squared error $\sum(\epsilon - y)^2$. The constant model is a special case of a linear least-squares fit with the slope constrained to be 0, and the model proportional to x is a special case of a linear model with intercept 0. The parameter values found by the linear regression, as optimal fits to the data, can be compared with the values found for the constrained models to establish which model is the better fit for the data.

To compare the sensitivity of the linear-regression test with the goodness-of-fit test, a Monte Carlo procedure was used comparing the efficacy of both tests on synthetic data generated from both constant and linear models. For the linear model, a Z score for the slope against a nominal value of 0 was used. This was found to be correctly distributed in constant models and to grow with the size of the effect in linear models. The χ^2 goodness-of-fit test also correctly identified which model had been used to generate the data but generally achieved a lesser degree of statistical significance than the regression test on the same data. This loss of significance was equivalent to that produced by a 17% deflation of the Z score in the regression test. It therefore was decided that the linear regression tests would be used as the primary evaluator for these hypothesis comparisons.

The χ^2 goodness-of-fit tests, however, retain a secondary utility. The regression test must always report *some* values for its parameters, whether the data are actually linear or not. Comparing these with the model predictions can identify a preferred model but will not give a warning if neither model is a good fit to the data. The χ^2 tests, on the other hand, can identify cases in which neither model fits the data well. If χ^2 is large for both models, some further source of variation, not well described by either model, is present.

5.1 Formulas

Each subdivision of the overall database produces a group of datasets, each with its own Z_i and N_i , where the N s may count bits, DAT events, or seconds as appropriate to the effect size measure in use. The i th dataset thus has an effect size observation $\epsilon_i = Z_i / \sqrt{N_i}$. In the description of Z scores above, it was noted that Z is always normalized to the standard deviation, or measurement uncertainty, of the observation in question. Thus, a Z score must by construction have a standard deviation of 1. The uncertainty in the effect size ϵ_i is therefore $\sigma_{\epsilon_i} = 1 / \sqrt{N_i}$.

In addition, there is a set x_i of values of the key model-discriminating

parameter. After suitable definition of x , we are always comparing two models, one of which predicts ϵ_i constant, while the other predicts $\epsilon_i \propto x_i$. The constant model is

$$\hat{\epsilon} = k_c \quad \text{where } k_c = \frac{\sum \epsilon_i / \sigma_i^2}{\sum 1 / \sigma_i^2} \text{ and } \sigma(k_c) = \frac{1}{\sqrt{\sum 1 / \sigma_i^2}}. \quad (1)$$

Here $\sigma(k_c)$ denotes the overall uncertainty on the value of k_c that propagates from the individual measurement uncertainties of the ϵ_i . The alternative model always predicts a linear dependence $\hat{\epsilon} \propto x$; we will call this the *slope* model to distinguish it from the linear model produced by the empirical two-parameter regression. The formulas for this model are:

$$\hat{\epsilon} = k_s x \quad \text{where } k_s = \frac{\sum x_i y_i / \sigma_i^2}{\sum x_i^2 / \sigma_i^2} \text{ and } \sigma(k_s) = \frac{1}{\sqrt{\sum x_i^2 / \sigma_i^2}}. \quad (2)$$

The general two-parameter regression, on the other hand, involves formulas that are rather complicated and can be understood more readily in an incremental presentation. If we first define the intermediate quantities,

$$\begin{aligned} m_0 &= \sum 1 / \sigma_i^2, & m_1 &= \sum x_i / \sigma_i^2, & m_2 &= \sum x_i^2 / \sigma_i^2, \\ c_{0,i} &= \frac{m_2 / \sigma_i^2 - m_1 x_i / \sigma_i^2}{m_0 m_2 - m_1^2}, & \text{and } c_{1,i} &= \frac{m_0 x_i / \sigma_i^2 - m_1 / \sigma_i^2}{m_0 m_2 - m_1^2}, \end{aligned} \quad (3.A)$$

we then can simply define the linear parameters:

$$\begin{aligned} \hat{\epsilon} &= b_0 + b_1 x, \\ \text{where } b_0 &= \sum_i c_{0,i} \epsilon_i \quad \text{and} \quad b_1 = \sum_i c_{1,i} \epsilon_i; \\ \sigma(b_0) &= \sqrt{\sum c_{0,i}^2 \sigma_i^2}; \quad \sigma(b_1) = \sqrt{\sum c_{1,i}^2 \sigma_i^2}. \end{aligned} \quad (3.B)$$

5.2 Problem with the Simplest Approach

It might seem that the linear regression provides a direct and immediate test between the two models; because one predicts a slope and the other does not, we need only examine the slope parameter (b_1) for statistical significance. This approach has even been used in other work comparing linear models (May, Utts, and Spottiswoode, 1995). However, Figure 1 illustrates a disconcerting feature of such a direct comparison.

In Figure 1a, the eight datasets at different values of the DAT sequence length n are plotted against $1 / \sqrt{n}$; the vertical axis is the bitwise effect size ϵ_b . Figure 1b uses the dual representation, plotting ϵ_d against \sqrt{n} . The two dotted lines in each graph show the two models; the solid slanting line shows the regression fit. In the first graph, the slope of the regression line seems rather am-

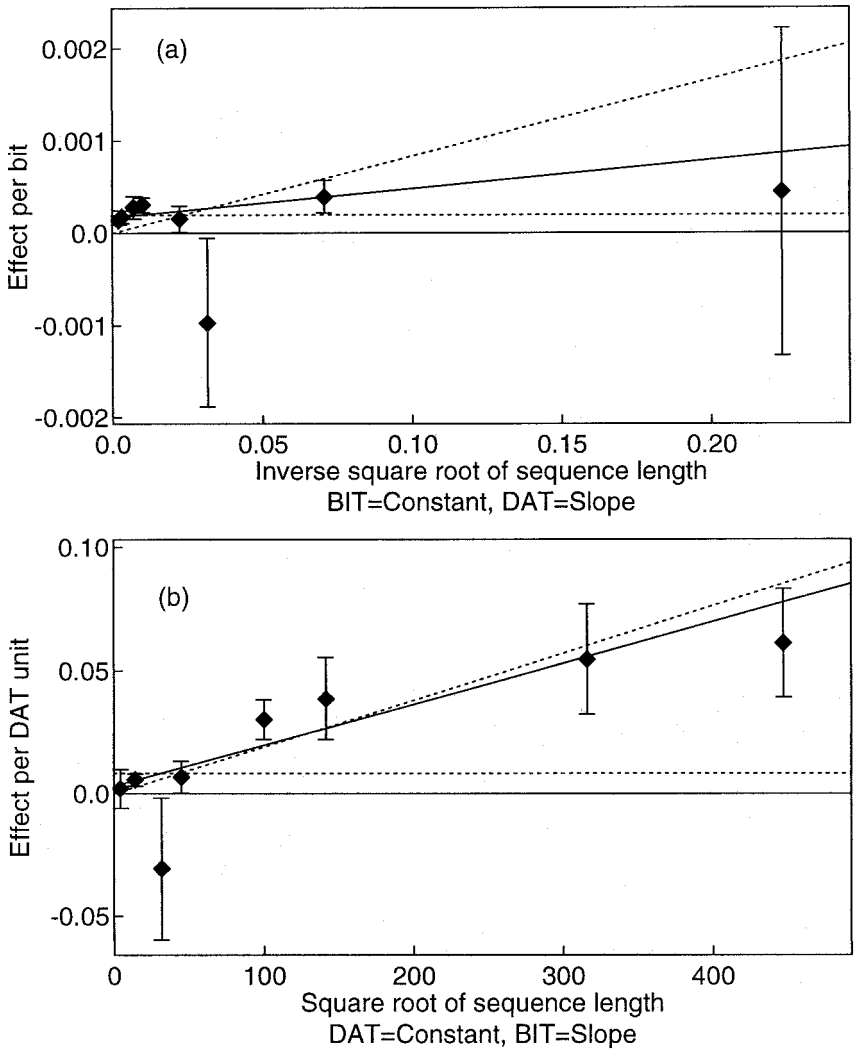


Fig. 1. Model comparison in two representations.

biguous between the two predictions; in the second, it is extremely close to one of the predictions. It would seem that the representation in the first graph allows no very strong preference between models, while the representation in the second graph definitely supports one. But it seems very odd to arrive at such different conclusions, when one considers that we are examining the same two models on exactly the same data; only the representations differ.

Some thought will demonstrate that a simple comparison, whether visual or numeric, between the slopes of the regression line and the slope model is suspect. Both slopes are obtained from a least-squares fit to the same data; a com-

parison between them, using the error figures in the formulas above, is liable to overestimate the variance of the difference and thus underestimate the significance of that difference.

That this is indeed the case was established by a systematic Monte Carlo evaluation of the slope-comparison procedure. The normal statistical interpretation of a Z score requires that it be normally distributed with mean 0 and variance 1 when the model it is testing is true. So to determine the validity of a slope-comparison Z score, the Monte Carlo program constructed synthetic datasets according to the requirements of a slope model.

Figure 2 shows the result of a 10,000-iteration Monte Carlo evaluation of the slope-comparison Z score $(b_1 - k_s)/\sigma(b_1)$ for data spaced as in the BIT versus DAT comparison. The points with error bars show the empirical population density, the solid curve shows the corresponding normal fit, and the dotted curve shows the standard normal probability density. We can see that the Z score remains normally distributed, but its standard deviation is much too small, about .48 rather than 1.

5.3 Vector Formulation

How, then, can we obtain reliable tests of the two models? A certain amount of analytical development is needed to demonstrate the proper tests. Let us, first, reprise the situation: For any model comparison, after identifying the key discriminator variable in the database, we have a set of observations ϵ_i of the effect size, each with uncertainty σ_i , obtained at a key variable value x_i . We have, in essence, four models to consider. The null hypothesis contends that

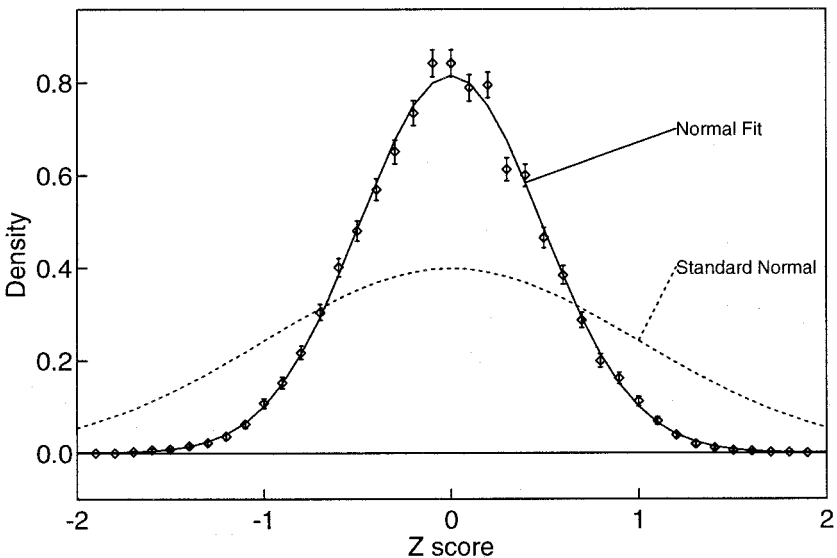


Fig. 2. Monte Carlo population density for slope-comparison Z score.

there is no effect and that all of the observations differ from 0 only by noise. The constant and slope hypotheses have been discussed in the previous section, and the linear regression formulae were given in Equation 3 amount to the assumption of a two-parameter model. By explicitly adding a noise term v_i , assumed to be normally distributed with mean 0 and standard deviation σ_i , we may express the models in terms of predictions for individual observations:

$$\begin{aligned}
 (\text{Null}) &: \epsilon_i = v_i \\
 (\text{Constant}) &: \epsilon_i = k_c + v_i \\
 (\text{Slope}) &: \epsilon_i = k_s x_i + v_i \\
 (\text{Regression}) &: \epsilon_i = b_0 + b_1 x_i + v_i.
 \end{aligned} \tag{4}$$

Now let us change notation by dividing each of these equations by σ_i . Defining the new variables,

$$z_i \equiv \frac{\epsilon_i}{\sigma_i}; \quad \xi_{c,i} \equiv \frac{1}{\sigma_i}; \quad \xi_{s,i} \equiv \frac{x_i}{\sigma_i}; \quad \varepsilon_i \equiv \frac{v_i}{\sigma_i}, \tag{5}$$

allows us to re-express Equation 4:

$$\begin{aligned}
 (\text{Null}) &: z_i = \varepsilon_i \\
 (\text{Constant}) &: z_i = k_c \xi_{c,i} + \varepsilon_i \\
 (\text{Slope}) &: z_i = k_s \xi_{s,i} + \varepsilon_i \\
 (\text{Regression}) &: z_i = b_0 \xi_{c,i} + b_1 \xi_{s,i} + \varepsilon_i.
 \end{aligned} \tag{6}$$

Note that the new noise terms, ε_i , all have variance 1, from their definition in Equation 5. Also, because we assumed that the initial noise terms, v_i , were normally distributed, the normality of the error terms will continue to propagate through subsequent analyses. (This assumption of normality is quite accurate for the empirical data presently under consideration.)

To alleviate an excess of subscripts, we may take advantage of the fact that the model equations can readily be written in terms of vectors rather than individual indexed equations. Thus, for example, \vec{z} is the vector whose individual elements are the z_i corresponding to individual observations: $\vec{z} = \{z_i\} = \{z_1, z_2, \dots, z_m\}$, where there are a total of m observations. With similar notations for the other indexed variables we may write

$$\begin{aligned}
 (\text{Null}) &: \vec{z} = \vec{\varepsilon} \\
 (\text{Constant}) &: \vec{z} = k_c \vec{\xi}_c + \vec{\varepsilon} \\
 (\text{Slope}) &: \vec{z} = k_s \vec{\xi}_s + \vec{\varepsilon} \\
 (\text{Regression}) &: \vec{z} = b_0 \vec{\xi}_c + b_1 \vec{\xi}_s + \vec{\varepsilon}.
 \end{aligned} \tag{7}$$

Before proceeding further, let us recall the basic operations of vector arithmetic. If we have two vectors $\vec{x} = \{x_i\}$ and $\vec{y} = \{y_i\}$, the vector sum, $\vec{x} + \vec{y} = \{x_i + y_i\}$, is the vector obtained by adding the components of the two vectors.

The product of a scalar (simple number) with a vector, as in $k\vec{x} = \{kx_i\}$, is the vector obtained by multiplying each element of \vec{x} individually by k . The inner product, $\vec{x} \cdot \vec{y} = \sum_i x_i y_i$, is the result of multiplying each element of \vec{x} by the corresponding element of \vec{y} and adding up all the resulting products. The length of a vector $|\vec{x}|$ can be found by taking the square root of its inner product with itself: $\vec{x} \cdot \vec{x} = |\vec{x}|^2$. One more useful result is not a vector identity but follows from applying these formulas. Because the z_i defined in Equation 5 have been normalized to the observational uncertainty, they all have a variance 1. The vector \vec{z} used in Equation 7 therefore has a variance 1 in each element: $\sigma^2[\vec{z}] = \vec{1}$. Such a vector has the property that the variance of its inner product with any constant (*i.e.*, nonrandom or zero-variance) vector \vec{x} is

$$\sigma^2[\vec{z} \cdot \vec{x}] = |\vec{x}|^2. \quad (8)$$

This can readily be verified by explicitly writing out the inner product and applying the standard rules for variances of sums and products.

All of the formulas in Equations 1 through 3 can be reexpressed in this vector notation, with exactly the same values for the model parameters, k_c , k_s , b_0 , and b_1 . Let us begin by considering a least-squares fit to the constant model. We note from Equation 7 that both the constant model and the slope model have exactly the same functional form, $\vec{z} = k\vec{\xi} + \vec{\epsilon}$, differing only in a change of subscript; therefore, a single derivation will do for both. Because the noise term, $\vec{\epsilon}$, is unknown, we can solve for it as the error between the model prediction and the empirical value: $\vec{\epsilon} = \vec{z} - k\vec{\xi}$. The optimal value of k is the one that minimizes the length, or equivalently minimizes the squared length, of this error. This squared length is

$$|\vec{\epsilon}|^2 = (\vec{z} - k\vec{\xi}) \cdot (\vec{z} - k\vec{\xi}) = |\vec{z}|^2 - 2k\vec{z} \cdot \vec{\xi} + k^2|\vec{\xi}|^2. \quad (9)$$

To find the minimizing value of k , we take the partial derivative $\partial(|\vec{\epsilon}|^2)/\partial k$ and set it to zero:

$$\begin{aligned} \frac{\partial(|\vec{\epsilon}|^2)}{\partial k} &= -2\vec{z} \cdot \vec{\xi} + 2k|\vec{\xi}|^2 = 0 \\ \implies k|\vec{\xi}|^2 &= \vec{z} \cdot \vec{\xi} \\ \implies k &= \frac{\vec{z} \cdot \vec{\xi}}{|\vec{\xi}|^2}. \end{aligned} \quad (10)$$

By substituting the definitions of Equation 5 for z_i and $\xi_{c,i}$, we find that this gives us Equation 1 for k_c ; by instead substituting $\xi_{s,i}$, we recover Equation 2 for k_s . By applying the variance identity (Equation 8) to the last line of equation 10 above, we find

$$\begin{aligned}\sigma^2[k] &= \sigma^2 \left[\frac{\vec{z} \cdot \vec{\xi}}{|\vec{\xi}|^2} \right] = \frac{\sigma^2[\vec{z} \cdot \vec{\xi}]}{|\vec{\xi}|^4} = \frac{|\vec{\xi}|^2}{|\vec{\xi}|^4} = \frac{1}{|\vec{\xi}|^2} \\ &\implies \sigma(k) = \frac{1}{|\vec{\xi}|},\end{aligned}\quad (11)$$

which because $|\vec{\xi}_c| = \sqrt{\sum_i 1/\sigma_i^2}$ and $|\vec{\xi}_s| = \sqrt{\sum_i x_i^2/\sigma_i^2}$, appropriately reproduces Equation 1 for $\sigma(k_c)$ or Equation 2 for $\sigma(k_s)$, depending on whether we insert $\vec{\xi}_c$ or $\vec{\xi}_s$ for $\vec{\xi}$ in the last line of Equation 11.

The vector formulas for b_0 and b_1 are likewise identical, whether one rewrites Equation 3 in terms of \vec{z} , $\vec{\xi}_c$, and $\vec{\xi}_s$ or performs a least-squares-fit calculation directly on the model formula in Equation 7. Because the vector notation avoids the tedious enumeration of indices and summation signs, we present the vector forms directly rather than through the intermediate step of the $c_{0,1}$ coefficients:

$$\begin{aligned}b_0 &= \frac{[|\vec{\xi}_s|^2 \vec{\xi}_c - (\vec{\xi}_c \cdot \vec{\xi}_s) \vec{\xi}_s] \cdot \vec{z}}{|\vec{\xi}_c|^2 |\vec{\xi}_s|^2 - (\vec{\xi}_c \cdot \vec{\xi}_s)^2}; & \sigma(b_0) &= \frac{|\vec{\xi}_s|}{\sqrt{|\vec{\xi}_c|^2 |\vec{\xi}_s|^2 - (\vec{\xi}_c \cdot \vec{\xi}_s)^2}}; \\ b_1 &= \frac{[|\vec{\xi}_c|^2 \vec{\xi}_s - (\vec{\xi}_c \cdot \vec{\xi}_s) \vec{\xi}_c] \cdot \vec{z}}{|\vec{\xi}_c|^2 |\vec{\xi}_s|^2 - (\vec{\xi}_c \cdot \vec{\xi}_s)^2}; & \sigma(b_1) &= \frac{|\vec{\xi}_c|}{\sqrt{|\vec{\xi}_c|^2 |\vec{\xi}_s|^2 - (\vec{\xi}_c \cdot \vec{\xi}_s)^2}}.\end{aligned}\quad (12)$$

Because the inner product is commutative ($\vec{x} \cdot \vec{y} = \vec{y} \cdot \vec{x}$), we may note from Equation 12 that b_0 and $\sigma(b_0)$ differ from b_1 and $\sigma(b_1)$ solely through the systematic exchange of c and s subscripts.

There is one more set of quantities of interest that can be computed directly from Equations 10 through 12. These equations give values and uncertainties for the four parameters k_c , k_s , b_0 , and b_1 . For each parameter, we can therefore calculate a Z score against the hypothesis that its value is 0. From Equations 10 and 11, we can calculate that

$$Z_k = \frac{k}{\sigma_k} = \frac{\vec{z} \cdot \vec{\xi}}{|\vec{\xi}|},\quad (13)$$

which can be applied to either k simply by attaching the appropriate subscript. The values of Z_0 and Z_1 corresponding to tests of nonzero value on b_0 and b_1 are from Equation 12:

$$Z_0 = \frac{b_0}{\sigma_{b_0}} = \frac{[|\vec{\xi}_s|^2 \vec{\xi}_c - (\vec{\xi}_c \cdot \vec{\xi}_s) \vec{\xi}_s] \cdot \vec{z}}{|\vec{\xi}_s| \sqrt{|\vec{\xi}_c|^2 |\vec{\xi}_s|^2 - (\vec{\xi}_c \cdot \vec{\xi}_s)^2}};$$

$$Z_1 = \frac{b_1}{\sigma_{b_1}} = \frac{[|\vec{\xi}_c|^2 \vec{\xi}_s - (\vec{\xi}_c \cdot \vec{\xi}_s) \vec{\xi}_c] \cdot \vec{z}}{|\vec{\xi}_c| \sqrt{|\vec{\xi}_c|^2 |\vec{\xi}_s|^2 - (\vec{\xi}_c \cdot \vec{\xi}_s)^2}}. \quad (14)$$

As we might expect from the earlier observation on Equation 12, $Z_0 \leftrightarrow Z_1$ upon substituting $\vec{\xi}_c \leftrightarrow \vec{\xi}_s$.

5.4 Abstract Modeling Problem

Let us temporarily take our leave of the complicated least-squares-fit formulas and consider a very abstract representation of modeling data in multiple dimensions. We start with the assumption that we have a set of observations, which have been normalized such that each observation has unit variance. We summarize the set of observations as a vector \vec{z} with as many dimensions as we have observations. Because \vec{z} has been constructed to have unit variance, we can express it as two components, $\vec{z} = \langle \vec{z} \rangle + \vec{\epsilon}$, with a deterministic part $\langle \vec{z} \rangle$ and a random component $\vec{\epsilon}$ defined as in the previous section: $\langle \vec{\epsilon} \rangle = 0$, $\sigma^2[\vec{\epsilon}] = \sigma^2[\vec{z}] = \vec{1}$.

Let us presume further that we have two vectors, \vec{v}_1 and \vec{v}_2 , which are selected on theoretical grounds as being plausible models for the phenomena. The observed results might be explained by the first model, by the second, or by some linear combination of the two. We can summarize all three possibilities, along with the null hypothesis, by these four formulas for $\langle \vec{z} \rangle$:

$$\begin{aligned} H_0 : \langle \vec{z} \rangle &= \vec{0}; \\ H_1 : \langle \vec{z} \rangle &= \alpha_1 \vec{v}_1; \\ H_2 : \langle \vec{z} \rangle &= \alpha_2 \vec{v}_2; \\ H_3 : \langle \vec{z} \rangle &= \beta_1 \vec{v}_1 + \beta_2 \vec{v}_2. \end{aligned} \quad (15)$$

Figure 3 illustrates a two-dimensional example of this formalism, with the observation vector \vec{z} and the theory vectors \vec{v}_1 and \vec{v}_2 shown in solid black, while one speculative possibility for the model vector $\langle \vec{z} \rangle$ and the noise vector $\vec{\epsilon}$ are drawn in gray.

The equations (Equation 15) form a hierarchy, in the sense that lower-numbered models are contained within higher-numbered ones as special cases: H_0 is a special case of either H_1 or H_2 , while either of the latter is a special case of H_3 . In each case the simpler model is produced by setting a parameter of the more complex model to 0. The questions we wish to answer are all of the form of asking whether this in fact holds true. Within the context of H_1 or H_2 , is $\alpha = 0$ so that H_0 is actually true? Within the context of the two-parameter model H_3 , is one of the β parameters actually 0 so that the model reduces to H_1 or H_2 ?

It is relatively straightforward to find an appropriate test for the null hypothesis H_0 within the context of either H_1 or H_2 . First, noting that H_1 and H_2 are

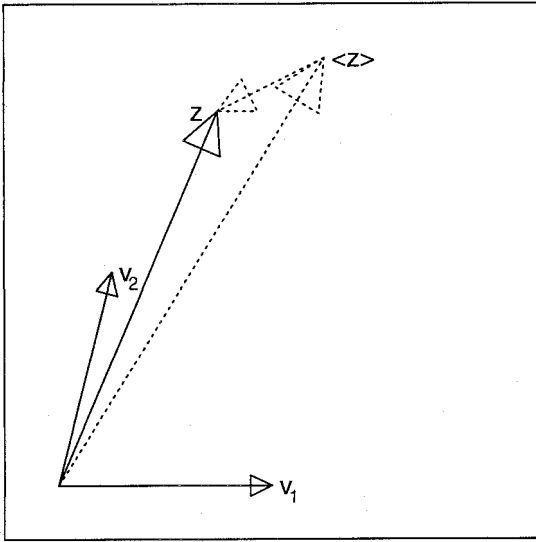


Fig. 3. Example of theory and observation vectors.

identical in functional form, differing only in subscript, we may for purposes of the derivation drop the subscript and simply work with $\langle \vec{z} \rangle = \alpha \vec{v}$. Since the observation vector \vec{z} is constructed to have unit variance, a properly distributed Z score for H_0 can be constructed by taking the *parallel projection* of \vec{z} along the hypothesis vector: that is, the component of the length $|\vec{z}|$ in the direction defined by the hypothesis vector \vec{v} . This projection is

$$Z = \frac{\vec{z} \cdot \vec{v}}{|\vec{v}|}. \quad (16)$$

We can readily verify that this formula gives us a valid Z score for H_0 . The expectation and variance of Equation 16 are given by:

$$\begin{aligned} \langle Z \rangle &= \left\langle \frac{\vec{z} \cdot \vec{v}}{|\vec{v}|} \right\rangle = \frac{\langle \vec{z} \rangle \cdot \vec{v}}{|\vec{v}|} = \frac{\alpha \vec{v} \cdot \vec{v}}{|\vec{v}|} = \alpha |\vec{v}|; \\ \sigma^2[Z] &= \sigma^2 \left[\frac{\vec{z} \cdot \vec{v}}{|\vec{v}|} \right] = \frac{\sigma^2[\vec{z} \cdot \vec{v}]}{|\vec{v}|^2} = \frac{|\vec{v}|^2}{|\vec{v}|^2} = 1, \end{aligned} \quad (17)$$

making use of the variance identity (Equation 8) and the model substitution $\langle \vec{z} \rangle = \alpha \vec{v}$. We see that the variance of Z is 1, as required, while the expectation of Z is proportional to α , so that it is 0 when H_0 is true and non-zero when H_0 is false.

The question of evaluating whether the general two-parameter model H_3 reduces to a one-parameter model is more involved. If one simply looks at the projection of \vec{z} along one of the hypothesis vectors, one is repeating Equation

16 and computing a test statistic for H_0 within H_1 or H_2 , rather than computing the desired test statistic for H_1 or H_2 within H_3 . The fundamental problem is that we have no guarantee that the vectors \vec{v}_1 and \vec{v}_2 are orthogonal. (Indeed, in our intended application, they are not.) Given nonorthogonal hypothesis vectors, simple projections of \vec{z} along \vec{v}_1 and \vec{v}_2 will *both* have nonzero expectation if *either* has a nonzero coefficient in reality. If, let us say, the actual model is H_1 , so that $\langle \vec{z} \rangle = \alpha_1 \vec{v}_1$ by hypothesis, the expected value of $Z_1 = \vec{z} \cdot \vec{v}_1 / |\vec{v}_1|$ is $\langle Z_1 \rangle = \alpha_1 |\vec{v}_1|$. The expected value of Z_2 , on the other hand, is $\langle \vec{z} \rangle \cdot \vec{v}_2 / |\vec{v}_2| = \alpha_1 \vec{v}_1 \cdot \vec{v}_2 / |\vec{v}_2|$; this vanishes only if $\vec{v}_1 \cdot \vec{v}_2 = 0$, *i.e.*, if the two hypothesis vectors are orthogonal. For the general case with nonorthogonal hypotheses, both $\langle Z_1 \rangle \neq 0$ and $\langle Z_2 \rangle \neq 0$ regardless of which hypothesis is true. Clearly neither Z_1 nor Z_2 can be a satisfactory test against the hypothesis that one of the β parameters in H_3 is 0.

For definiteness, let us consider the case in which we wish to test whether $\beta_2 = 0$, so that H_3 reduces to H_1 . (Since the notation is completely symmetric, we can do our derivations once and then get the formulas for the other case by interchanging $1 \leftrightarrow 2$.) To obtain a Z score against this possibility, we need to know the extent to which the observation \vec{z} requires a component *not* simply proportional to \vec{v}_1 ; that is, we need the projection of \vec{z} not along \vec{v}_2 , but along that component of \vec{v}_2 orthogonal to \vec{v}_1 . Let us label this orthogonal component \tilde{v}_2 ; its value is given by

$$\tilde{v}_2 = \vec{v}_2 - \frac{\vec{v}_1 \cdot \vec{v}_2}{|\vec{v}_1|^2} \vec{v}_1. \quad (18)$$

One can readily verify that $\vec{v}_1 \cdot \tilde{v}_2 = 0$, so that \tilde{v}_2 is indeed, as we require, orthogonal to \vec{v}_1 .

The Z score against $\beta_2 = 0$ is then just the projection of \vec{z} in the \tilde{v}_2 direction:

$$Z_{\beta_2} = \frac{\vec{z} \cdot \tilde{v}_2}{|\tilde{v}_2|}. \quad (19)$$

Let us verify that this does in fact have variance 1 and expectation 0 if, and only if, $\beta_2 = 0$.

$$\begin{aligned} \langle Z_{\beta_2} \rangle &= \left\langle \frac{\vec{z} \cdot \tilde{v}_2}{|\tilde{v}_2|} \right\rangle = \frac{\langle \vec{z} \rangle \cdot \tilde{v}_2}{|\tilde{v}_2|} = \frac{(\beta_1 \vec{v}_1 + \beta_2 \vec{v}_2) \cdot \tilde{v}_2}{|\tilde{v}_2|} = \left(\frac{\vec{v}_2 \cdot \tilde{v}_2}{|\tilde{v}_2|} \right) \beta_2. \\ \sigma^2 [Z_{\beta_2}] &= \sigma^2 \left[\frac{\vec{z} \cdot \tilde{v}_2}{|\tilde{v}_2|} \right] = \frac{\sigma^2 [\vec{z} \cdot \tilde{v}_2]}{|\tilde{v}_2|^2} = \frac{|\tilde{v}_2|^2}{|\tilde{v}_2|^2} = 1. \end{aligned} \quad (20)$$

We see that the proposed Z score against $\beta_2 = 0$ has a variance 1 always and has an expectation value proportional to β_2 ; therefore, it is Z -distributed if and only if $\beta_2 = 0$, as required.

If we choose to expand \tilde{v}_2 in terms of its definition (Equation 18), so as to express all quantities in terms of the original hypothesis vectors \vec{v}_1 and \vec{v}_2 , the form of Z_{β_2} from Equation 19 becomes

$$\begin{aligned}
 Z_{\beta_2} &= \frac{\vec{z} \cdot \vec{v}_2}{|\vec{v}_2|} = \frac{\vec{z} \cdot \left[\vec{v}_2 - \frac{\vec{v}_1 \cdot \vec{v}_2}{|\vec{v}_1|^2} \vec{v}_1 \right]}{\sqrt{\left(\vec{v}_2 - \frac{\vec{v}_1 \cdot \vec{v}_2}{|\vec{v}_1|^2} \vec{v}_1 \right) \cdot \left(\vec{v}_2 - \frac{\vec{v}_1 \cdot \vec{v}_2}{|\vec{v}_1|^2} \vec{v}_1 \right)}} \\
 &= \frac{\left[\vec{v}_2 - \frac{\vec{v}_1 \cdot \vec{v}_2}{|\vec{v}_1|^2} \vec{v}_1 \right] \cdot \vec{z}}{\sqrt{|\vec{v}_2|^2 - \frac{(\vec{v}_1 \cdot \vec{v}_2)^2}{|\vec{v}_1|^2}}}. \tag{21}
 \end{aligned}$$

If we multiply the numerator and denominator of the last line of Equation 21 by $|\vec{v}_1|^2$, we finally obtain

$$Z_{\beta_2} = \frac{[|\vec{v}_1|^2 \vec{v}_2 - (\vec{v}_1 \cdot \vec{v}_2) \vec{v}_1] \cdot \vec{z}}{|\vec{v}_1| \sqrt{|\vec{v}_1|^2 |\vec{v}_2|^2 - (\vec{v}_1 \cdot \vec{v}_2)^2}}, \tag{22}$$

with a corresponding Z_{β_1} obtained by interchanging $1 \leftrightarrow 2$ in Equation 22:

$$Z_{\beta_1} = \frac{[|\vec{v}_2|^2 \vec{v}_1 - (\vec{v}_1 \cdot \vec{v}_2) \vec{v}_2] \cdot \vec{z}}{|\vec{v}_2| \sqrt{|\vec{v}_1|^2 |\vec{v}_2|^2 - (\vec{v}_1 \cdot \vec{v}_2)^2}}. \tag{23}$$

Equations 22 and 23 should look very familiar in form. The equations in Equation 15 were deliberately laid out to suggest the form of the equations in Equation 7; H_0 , H_1 , H_2 , and H_3 are identical to the null, constant, slope, and regression models, respectively, if one identifies the symbols:

$$\alpha_1 = k_c; \quad \vec{v}_1 = \vec{\xi}_c; \quad \alpha_2 = k_s; \quad \vec{v}_2 = \vec{\xi}_s; \quad \beta_1 = b_0; \quad \beta_2 = b_1. \tag{24}$$

These identifications make it very clear that Equation 16, giving the Z score for a one-parameter model against the null hypothesis, is the same as Equation 13 derived from least-squares parameter evaluation, and that likewise Equations 22 and 23 for the Z scores of the parameters of the two-component model are identical to Equation 14 for the Z scores of the regression parameters.

These identities, particularly those equating Equations 22 and 23 with Equation 14, are critically important. We have already derived and demonstrated the fact that Z_{β_1} and Z_{β_2} are correctly distributed, properly normalized, and independent Z scores for the evaluation of their respective hypotheses. We now see that the least-squares method of parameter evaluation automatically generates the extra, orthogonalizing terms required to make sure that Z_{β_1} tests only the parameter β_1 regardless of the value of β_2 , and vice versa. Because we made no particular assumptions about the values of \vec{v}_1 or \vec{v}_2 in the derivation from Equations 15 through 23, this derivation of the validity of the Z formulas is entirely general; in particular, it holds despite the fact that our regression formulas (Equation 3) use the origin of coordinates, rather than the mean of the x_i values, as the origin of the regression.

One small subtlety needs to be kept in mind when evaluating these models.

The Z scores of Equations 22 and 23 are Z scores against the null hypothesis that their respective parameters are 0. To consider Z_{β_2} (which is Z_1 in the notation of Equation 14), this is a Z score against the hypothesis that $\beta_2 = 0$ ($b_1 = 0$), in which case H_3 (regression) reduces to H_1 (constant). Thus, the Z score of the *slope* is the test that can refute the *constant* model, by demonstrating a nonzero value of the parameter that must be 0 if the constant model is true. Likewise, the Z score of the intercept (Z_{β_1} , Z_0) is the Z score against H_2 (slope model).

Despite the power of the vector formalism, clarity of visualization requires that the original notation of ϵ_i , x_i , and σ_i be retained for illustrative purposes. Because graphs on paper are inherently two-dimensional, and most readers are three-dimensional, it grows very difficult to render graphically vectors in spaces of four or more dimensions. It is true that any two arbitrary vectors can define a plane, so that a representation such as Figure 3 can show the relations of the theory vectors to each other and to the projection of the data into their common plane. However, such a compressed representation makes it impossible to visualize the actual data for most analyses. Therefore, the analyses in subsequent sections will retain the familiar graphs in x_i , ϵ_i space as well as the vector representation developed here.

5.5 Resolution of the Quandary

With this new approach to calculating the Z score between the linear fit and each model, we find a fully symmetric duality between the two representations of the data in any model comparison. If we switch from one model's definition of the fundamental effect size to the other, the role of constant model and slope model are exchanged. But all of the statistical parameters also change places in a symmetrical fashion. The Z scores of the model parameters against the null hypothesis, of the linear regression against each model, all change places so that the same statistical figure-of-merit is associated with the same model. Even a χ^2 goodness-of-fit test against the residuals from the model displays the same behavior. All of the statistical measures are associated with a specific model, regardless of the data representation used (see Table 2).

TABLE 2
Duality of Representations

Parameter	$Z/\sqrt{N_b}$ effect	$Z/\sqrt{N_D}$ effect
X axis	$1/\sqrt{n}$	\sqrt{n}
Constant model	BIT	DAT
Slope model	DAT	BIT
Constant Z vs. null	5.80	3.97
Slope Z vs. null	3.97	5.80
Regression Z vs. constant	1.30	4.43
Regression Z vs. slope	4.43	1.30
χ^2_7 vs. constant	6.68	24.59
χ^2_7 vs. slope	24.59	6.68

This exact equivalence between representations is, of course, further evidence that we are justified in using the techniques discussed in the preceding section: The comparison between two models and a set of data *should* be independent of the choice of representation. Having established that both representations are exactly equivalent when properly analyzed, we are now free to choose either representation without concern that our choice biases the conclusion. The remaining figures therefore will display the BIT representation exclusively.

6. Model Test Outcomes

Although we have already seen some of the DAT versus BIT comparison results, since it was the example used in the previous section on developing the comparison methodology, we will cover the matter in more detail here. First, there is the matter of the raw data. The experimental summary in Table 1 is inadequate, because several of the experiments enumerated collected data at more than one sequence length, while in other cases more than one experiment used the same sequence length. Table 3 shows the result of collecting the data according to sequence length, the key variable for the DAT versus BIT comparison.

When one calculates the statistical parameters for the comparison between models—which, as noted in the previous section, are the same whether one

TABLE 3
Data for DAT vs. BIT

n	N_b	N_D	Z
20	3.2×10^5	1,600	0.2510
200	3.267×10^7	163,350	2.2335
1,000	1.2×10^6	1,200	-1.0626
2,000	4.868×10^7	24,340	1.0428
10,000	1.5182×10^8	15,182	3.7203
20,000	7.044×10^7	3,522	2.2989
100,000	2.04×10^8	2,040	2.4660
200,000	4.212×10^8	2,106	2.8082

Note: DAT = Decision Augmentation Theory; BIT = bitwise effect.

TABLE 4
BIT versus DAT Comparison

Comparison	Score	p
BIT vs. null (Z)	5.80	3.3×10^{-9}
DAT vs. null (Z)	3.97	3.6×10^{-5}
BIT vs. regression (Z)	1.30	.097
DAT vs. regression (Z)	4.43	4.7×10^{-6}
χ^2_7 on BIT	6.68	.46
χ^2_7 on DAT	24.60	9.0×10^{-4}

Note: BIT = bitwise effect; DAT = Decision Augmentation Theory.

uses $Z/\sqrt{N_b}$ or $Z/\sqrt{N_D}$ as one's measure of effect size—one finds the values presented in Table 2, represented here in Table 4 for more convenient model identification.

These outcomes are summarized graphically in Figures 4 and 5. Figure 4a shows the familiar regression graph, and 4b shows the various model parameters directly so as to facilitate comparison.

Figure 4b plots model slope against model intercept, in units of the empirical uncertainty associated with each measurement. In other words, it plots the

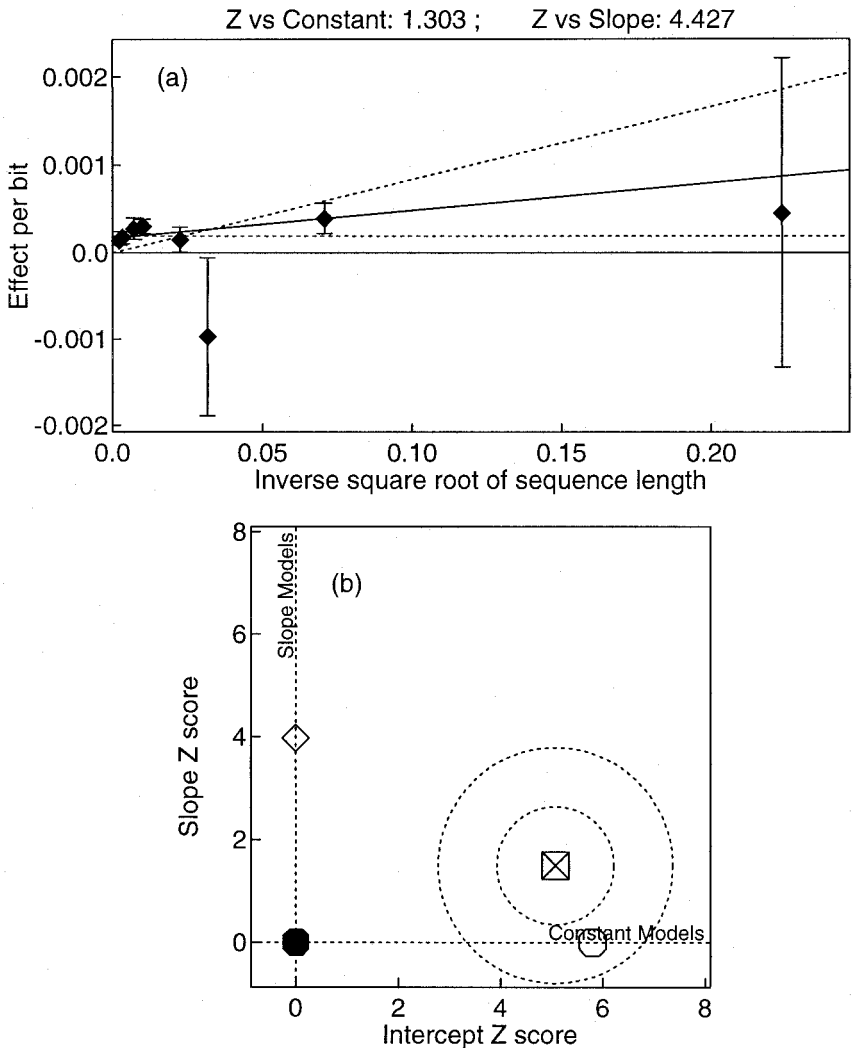
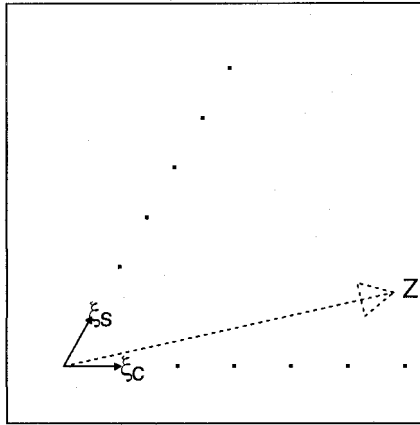


Fig. 4. Model comparison: BIT = constant and DAT = slope.

(a): Projection of Z In Theory Plane



(b): Normalized Residuals

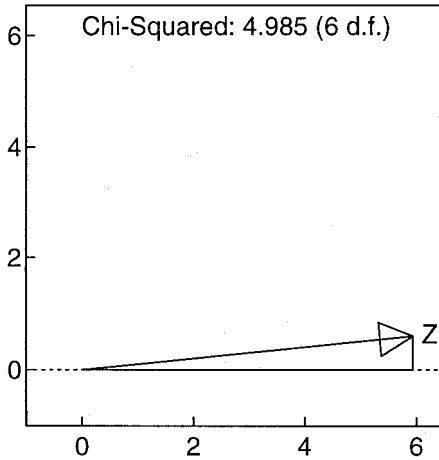


Fig. 5. BIT constant and DAT slope as vectors.

Z scores of the two models. The null hypothesis is represented by the filled dot at slope 0, intercept 0. The horizontal dotted line represents the family of all constant (zero-slope) models. Recall that in the current data representation, these are BIT models. The open circle on this line shows the actual model in this family that is the best fit to the data. The vertical line shows the family of all DAT-like, zero-intercept models. The open diamond on this line shows the actual DAT model that best fits the data.

The two-parameter linear regression fit also is shown, but *not* directly as a Z score. To provide a valid visual comparison, its slope and intercept are plotted to the same scale as the slope of the slope model and the intercept of the con-

stant model. Because the two-parameter regression in general has larger uncertainties on each parameter than the one-parameter models, this means that the position of the regression fit cannot simply be read off the axes to show a Z score. To provide a visual cue to the degree of uncertainty in the regression, 1σ and 2σ contours have been drawn around the fit point. We can clearly see that the general linear regression is quite close to the constant, BIT model, and very different from the entire DAT family of models.

Figure 5 illustrates the vectorial representation of the problem. As noted in section 5, any two (nonparallel) vectors define a plane, however high the dimensionality of the space in which the vectors themselves exist. Figure 5a presents a view of this “theory plane” defined by $\vec{\xi}_c$ for BIT and $\vec{\xi}_s$ for DAT. The two theory vectors themselves are drawn at unit length (the absolute scaling of this space is irrelevant to the theoretical comparison), with dots drawn at every unit increment of the vector length to help the eye project the vector's direction. Shown as a closely dotted line is the observation vector \vec{z} , or rather its two-dimensional projection into the theory plane. This representation, free of the visual biases induced by a regression plot, allows an immediate appreciation of how much more closely the observation falls along the $\vec{\xi}_c$ direction than the $\vec{\xi}_s$ direction.

Finally, Figure 5b deals with the residual vector $\vec{R} = \vec{z} - (b_0\vec{\xi}_c + b_1\vec{\xi}_s)$, that part of the observation vector \vec{z} that is not accommodated by either model at all. A few moments of vector algebra will allow one to verify that \vec{R} is perpendicular to both $\vec{\xi}_c$ and $\vec{\xi}_s$, and is therefore perpendicular to the entire theory plane. The viewer is assumed to be sighting along the theory plane, so that it forms the dotted line at $y = 0$. The solid line embedded in that dotted line is the projection of \vec{z} into the theory plane, while the vertical line and the vector \vec{z} itself are shown extending slightly into a perpendicular dimension. If one pretends, for a moment, that Figure 5a is a three-dimensional object rather than a flat plot on paper, the view in Figure 5b is what one would achieve by rotating the page slightly clockwise, so as to make the vector \vec{z} shown be level rather than inclined, and then tilting the page away from oneself so that one is sighting along the paper, rather than looking down on it.

It must be noted, however, that this lower graph is labeled as showing the “normalized residuals,” rather than simply the “residuals.” The reason for this is to avoid a visual deception arising from our accustomed three-dimensional experience. In the three-dimensional space of our everyday experience, a given plane has exactly one perpendicular axis. This is not so in more dimensions. The eight-dimensional space that Figure 5 attempts to summarize permits six mutually perpendicular directions all of which are also perpendicular to the theory plane. Under the hypothesis that $\langle \vec{z} \rangle$ lies in the theory plane, the actual observation \vec{z} is expected to depart by a Z -distributed amount in *each* of those six orthogonal directions, so the squared length of the residual vector, $\vec{R} \cdot \vec{R}$, is clearly a χ^2 with a number of degrees of freedom equal to the number of dimensions in which \vec{R} exists. Therefore, the visual presentation normalizes this length. Figure 5b reports the χ^2 value, and its degrees of freedom, explicitly:

but, because the eye only expects one perpendicular to a plane, rather than six, the vertical distance shown is the Z score that has the same two-tailed p -value as the χ^2 calculation. The vertical axis can be read off directly as a Z score; it represents the length that the residual vector *would* have if the entire vector space were confined to the three dimensions we are accustomed to visualizing.

Turning now to the comparison of BIT with TIM, Table 5 presents the raw data when segregated according to the key variable of bits per second. These figures result in the model comparisons summarized in Table 6.

Figures 6 and 7 illustrate the model comparison between BIT and TIM in the same way as Figures 4 and 5 do for BIT versus DAT. In Figure 6a, the x axis is, of course, $1/\sqrt{B}$. As it happens, we have only three data points for this plot, and one of them has rather large error bars. This is the reason for the very wide error contours visible on the lower plot; despite the fact that both the BIT and TIM models are almost 6σ away from the origin, the ability of the regression to discriminate is very much weaker, so that the 2σ contour covers almost the whole plot. Nevertheless, the regression fit is visibly much closer to the BIT model than to the TIM model. Figure 7a illustrates the same point: The acute angle between the two theory vectors shows why \vec{z} has little power to distinguish between them, despite being much closer to one. Figure 7b once again shows no consequential residual vector. It is worth noting that because we actually have only three dimensions in this vector space, the normalization of Figure 7b is one-to-one and we see the residual vector at its true length.

The TEL model refers to a historical evolution *within* a single experiment, rather than across experiments. The model parameter becomes the amount of

TABLE 5
BIT vs. TIM Comparison Data

B	N_b	N_s	Z
20	1.64×10^6	8.2×10^4	-0.9558
200	6.0409×10^8	3.02045×10^6	5.2477
2,000	3.246×10^8	1.623×10^5	2.7179

Note: BIT = bitwise effect; TIM = time normalization.

TABLE 6
BIT vs. TIM Comparison Results

Test	Score	p
BIT vs. null (Z)	5.79	3.5×10^{-9}
TIM vs. null (Z)	5.50	1.7×10^{-8}
BIT vs. regression (Z)	0.52	.30
TIM vs. regression (Z)	1.89	.029
χ_2^2 vs. BIT	2.27	.32
χ_2^2 vs. TIM	5.57	.062

Note: BIT = bitwise effect; TIM = time normalization.

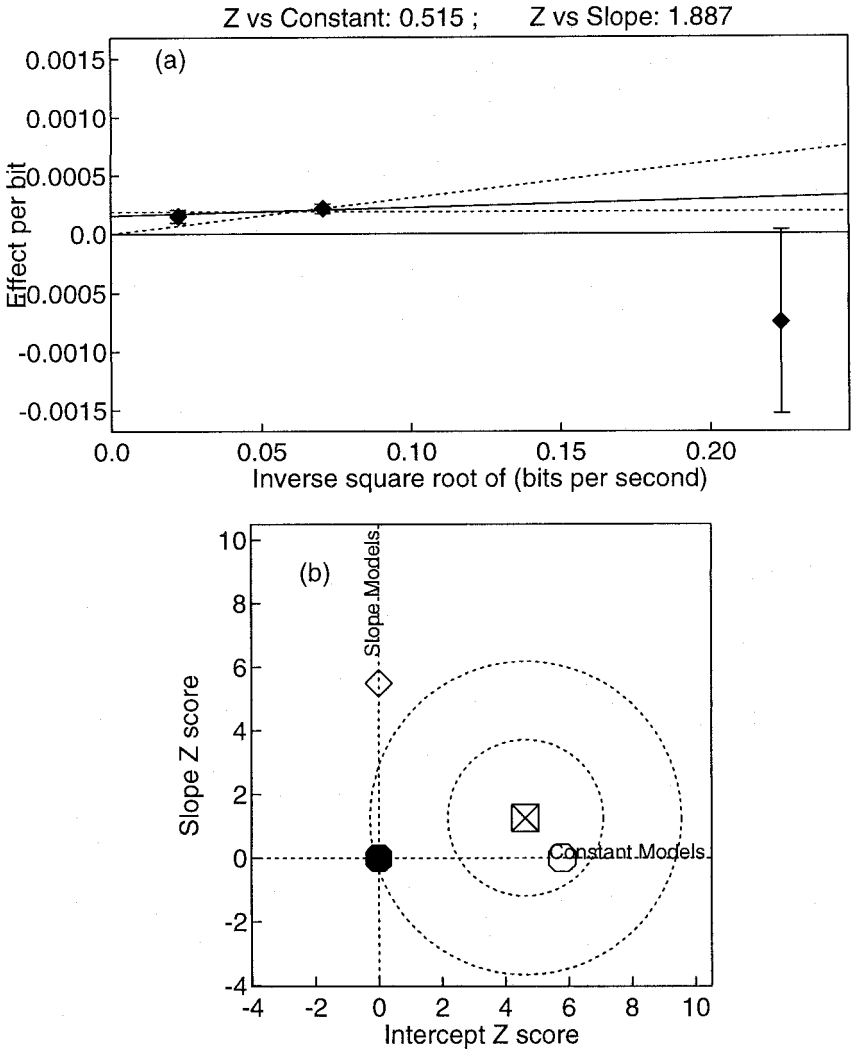
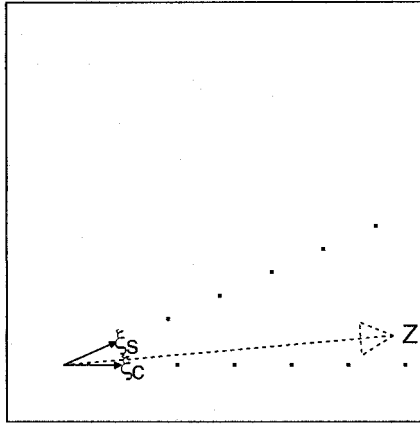


Fig. 6. Model comparison: BIT = constant and TIM = slope.

data accumulated in the experiment to date. To test this model, we therefore must perform evaluations within single experiments. Furthermore, because the hypothesis predicts a progressive diminution in effect from series to series, a tabular presentation of the raw data would be excessively cumbersome, with as many lines as the total of series enumerated in Table 1. Therefore, Table 7 presents only analysis results: the Z score versus the null hypothesis, the Z score against the linear regression, and the χ^2 fit test, for both BIT and TEL models, for the six experiments summarized in Table 1.

(a): Projection of Z In Theory Plane



(b): Normalized Residuals

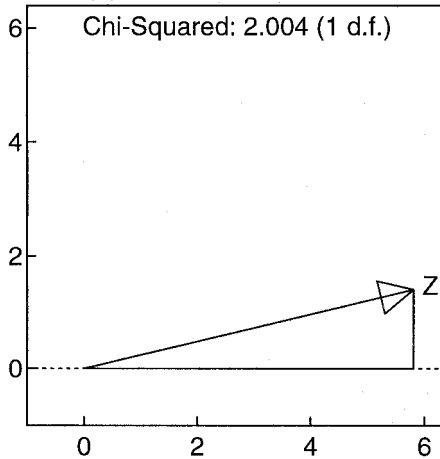


Fig. 7. BIT constant and TIM slope as vectors.

Here we have a rather disconcerting lack of unity. Some of the experimental datasets show a strong preference for TEL and some show a modest preference for BIT. The result cannot be attributed entirely to statistical power issues, as the two largest datasets show oppositely inclined preferences.

The largest of these datasets, furthermore, contains a known confound. As was noted in section 2, an analysis of this sort proceeds on the implicit assumption that conditions affecting the experimental yield, which are not related to the hypothesis under consideration, are uniformly or at least randomly distributed across the subsets defined by the hypothesis parameter. This is

TABLE 7
BIT *versus* TEL Summaries

Experiment	Z_{nb}	Z_{nt}	Z_{rb}	Z_{rt}	$\chi_b^2(df)$	$\chi_t^2(df)$
Original	3.809	5.405	3.886	-0.626	571.6 (521)	555.9 (521)
Remote	2.214	1.615	-0.244	1.534	200.5 (211)	202.8 (211)
Alternative source	2.765	3.205	1.624	0.072	57.12 (45)	54.49 (45)
Co-operator	1.635	1.525	0.218	0.628	39.68 (44)	40.03 (44)
REG2000	2.718	3.410	2.125	-0.523	45.41 (43)	41.17 (43)
REG20	-0.956	-0.229	1.498	-1.762	18.31 (19)	19.17 (19)

Note: BIT = bitwise effect; TEL = teleological; Z_{nb} = BIT Z score *versus* null; Z_{nt} = TEL Z score *versus* null; Z_{rb} = BIT Z score *versus* regression fit; Z_{rt} = TEL Z score *versus* regression fit; $\chi_b^2(df)$ = Goodness-of-fit χ^2 for BIT (degrees of freedom). $\chi_t^2(df)$ = Goodness-of-fit χ^2 for TEL (degrees of freedom).

known not to be the case for a specific subset of the “Original REG” data with respect to the TEL parameter.

As has been discussed in fuller detail elsewhere (Dobyns and Nelson, 1998), the earliest REG experiments were run with a slightly different experimental protocol, which distinguishes itself by a much larger effect size than the subsequent continuation of the experiment. Because the change in protocols is a historical one and distinguishes part of the earliest data from the remainder of the dataset, it is directly confounded with the measure of historical progression that defines the TEL hypothesis.

Merely identifying the presence of a confounding factor, however, cannot by itself determine the status of the two hypotheses. The fact that the difference between the early data (designated as the X protocol) and the subsequent data was well established long before the TEL hypothesis was considered prompts an inclination to consider the support for TEL a confounding result of the difference in protocols, but this reasoning is spurious. It is equally plausible that the unexplained difference in the protocols is driven by the hitherto-unrecognized effects of teleological decline.

To resolve this ambiguity, we would like ideally to compare data generated under the two protocols at the same teleological status. This unfortunately is impossible, because “teleological status” is a historical parameter that is distinct for any two series. We can do the next best thing, however, by segregating the two conditions and examining each independently for teleological effects.

There is a small subtlety involved in this segregation, however. When we remove the X protocol, we are taking away the earliest series in the Original REG dataset. Did the protocol change “restart” the teleological “clock,” so that the new data represented a new beginning? Or should the subsequent data be evaluated under the assumption that they are continuing a teleological decline started in the early data? We do not know enough about the constraints of the TEL model to answer this question, so we have no alternative but to ana-

TABLE 8
 BIT vs. TEL Summaries, Reduced Datasets

Experiment	Z_{nb}	Z_{nt}	Z_{rb}	Z_{rt}	$\chi_b^2(df)$	$\chi_r^2(df)$
X only	3.519	4.886	4.331	-2.696	46.19 (15)	34.69 (15)
Continue	3.102	2.734	0.034	1.466	516.9 (505)	519.0 (505)
Restart	3.102	2.286	-0.124	2.101	516.9 (505)	521.3 (505)

Note: BIT = bitwise effect; TEL = teleological.

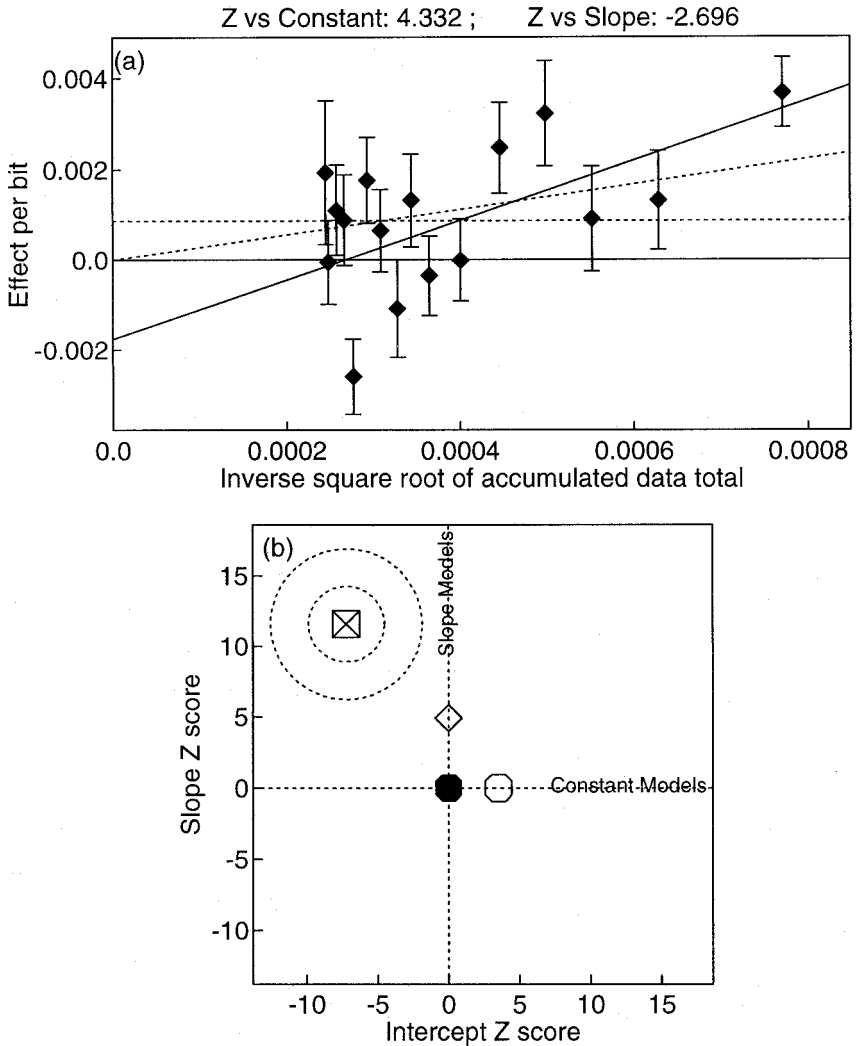


Fig. 8. BIT constant *versus* TEL slope for X protocol only.

lyze the “remainder” data in both ways. The results of such analysis are presented in Table 8.

It is plainly evident that while the X protocol shows a preference for TEL internally, *neither* model is a very good fit to the data in this unique dataset. Both χ^2 values are quite large, with p -values of 4.96×10^{-5} and 2.72×10^{-3} , respectively. Neither Z score is within the range that would be considered plausibly consistent with model variation. This strongly suggests that this dataset contains some structure other than that explained by either model. Conversely, the data without the X protocol no longer show any support for TEL, regardless of the status of the “restart” assumption.

Figures 8 and 9 display the X protocol data in the format used for Figures 4 and 5 and Figures 6 and 7. In Figure 8a, we note what appears to be almost a bifurcation in the experimental data, with the data points below $x = 0.0004$ seeming at least visually to cluster into two vertically separated populations. It is inevitable that even two-parameter linear regression fails to fit this dataset well, because no single-valued functional fit can possibly reproduce such a feature. In Figure 8b, the regression fit is conspicuous for its wide departure from both families of theoretical model. Figure 9, illustrating the vector space representation, drives home the same point. In the theory plane of Figure 9a, the observation vector is well outside the acute angle subtended by the two theory vectors, differing more from either theoretical prediction than the theories differ from each other. The residual graph in Figure 9b shows us a vector lying far outside the theory plane, with an equivalent Z score well over 2.

Figures 10 and 11 give the same reports for the non-X protocol data, using the “restart” assumption. We can see that the results are now a very good fit to BIT, whether we look at the parameter plot in Figure 10 or the vectors of Figure 11. We also see a completely inconsequential residual in Figure 11b.

The overall conclusion remains ambiguous. Computing a weighted Z across all of the datasets reported in Table 7, with the “Original REG” further broken down as summarized in Table 8, leads to the unenlightening results presented in Table 9.

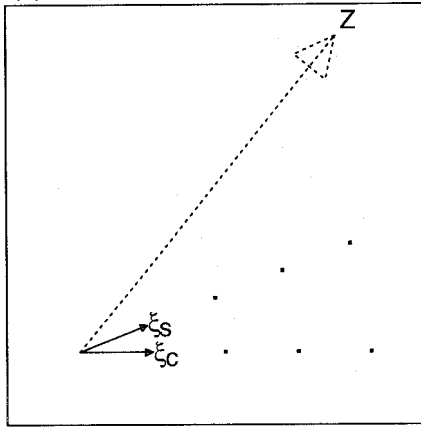
As in Table 6, two results are reported, depending on whether the transition from the X protocol to the standard protocol is considered a continuation of the Original REG experiment, or a transition to a new experiment. We are left with the conclusion that we have no good grounds for preferring either BIT or TEL on the basis of the current data.

TABLE 9
Overall BIT vs. TEL

X Treatment	Z vs. BIT	Z vs. TEL
Continue	1.603	1.270
Restart	1.484	1.750

Note: BIT = bitwise effect; TEL = teleological.

(a): Projection of Z In Theory Plane



(b): Normalized Residuals

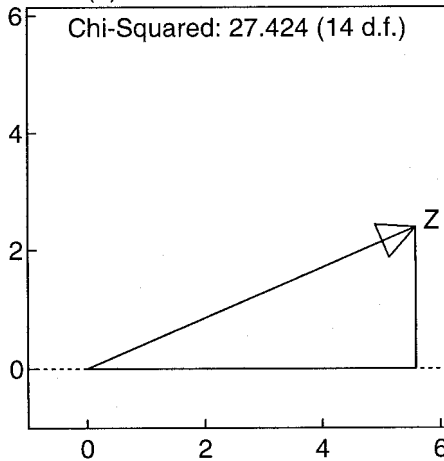


Fig. 9. BIT constant and TEL slope, as vectors, X protocol data only.

7. Conclusions

We noted above that the choice of model can affect even the the primary conclusion about the existence of an anomaly. One reason for this follows from the differences in the definition of the fundamental data unit. Let us suppose that we have several datasets, each of which contains N_i observations, and has a Z score Z_i , where $i = 1, \dots, m$ for $m =$ the number of datasets. What is the overall Z score of the aggregate database? It is provably correct that this composite score is $Z = (\sum Z_i \sqrt{N_i}) / \sqrt{\sum N_i}$, in the sense that this reproduces

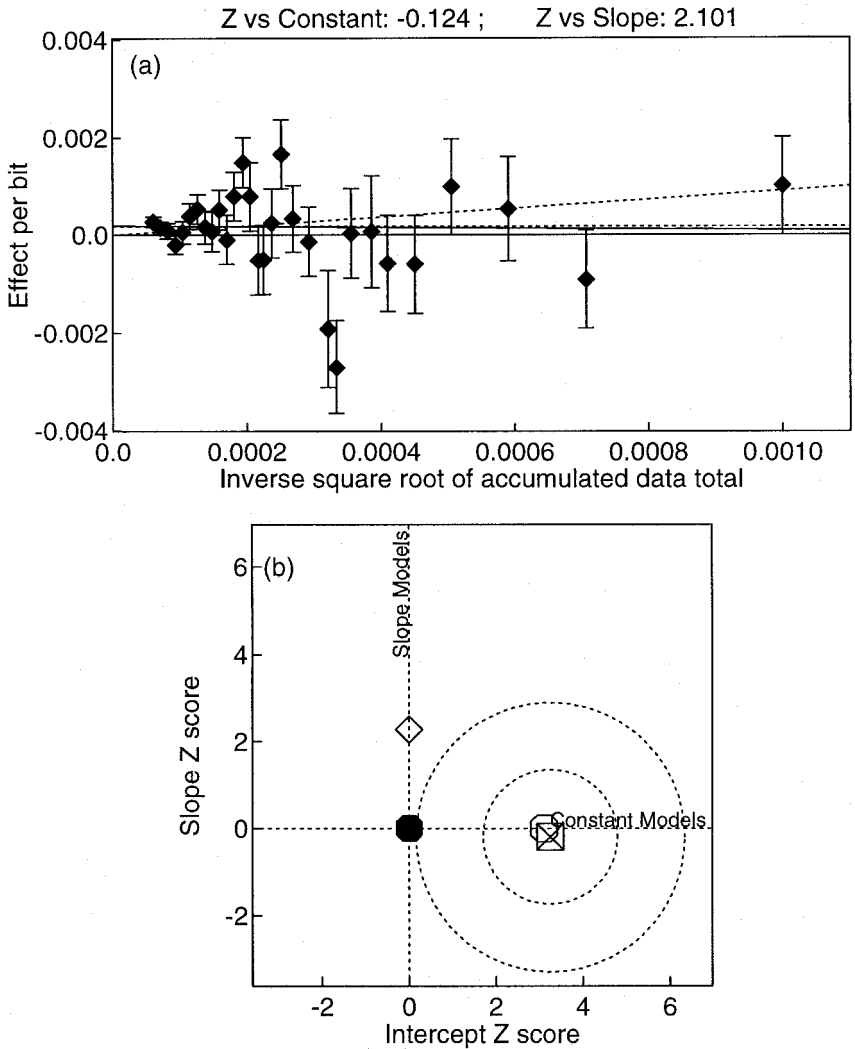
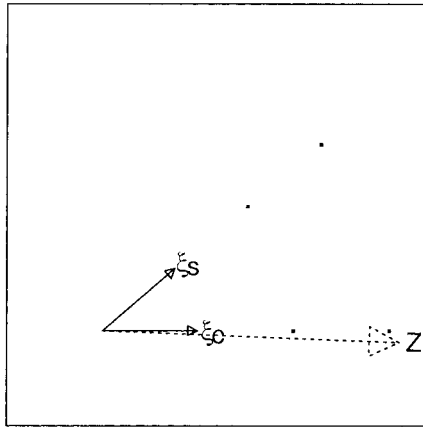


Fig. 10. BIT *versus* TEL with X data removed, restart version.

the Z score that would be calculated from pooling all of the raw data. We have noted, however, that different models of the phenomenon may disagree on the definition of the fundamental data unit and therefore may assign different values of N_i to the various subsets. In general, this changes the value of the composite Z. This is the reason for the different Z scores against the null hypothesis displayed by the different models in section 6.

Therefore, when competing models of the effect have been proposed, it seems advisable to select the model that is the best fit to the data before at-

(a): Projection of Z In Theory Plane



(b): Normalized Residuals

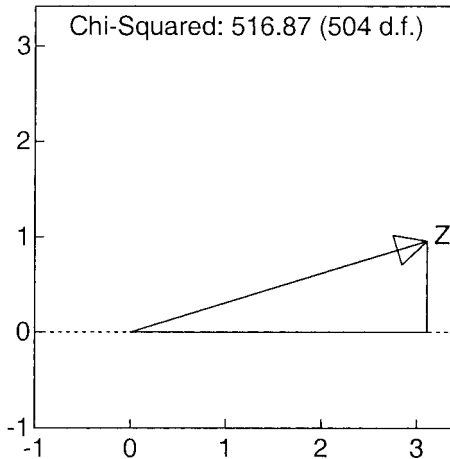


Fig. 11. BIT constant and TEL slope as vectors, X protocol data removed.

tempting an absolute test for the existence of the effect. We have done this, above, for the four models under consideration and find in each case that either the BIT model is preferred, or that no preference can be established.

If the BIT model is thus taken as the best candidate model of a real effect, we note that its Z score against the null hypothesis is 5.8, with a p -value of 3.3×10^{-9} . We may be confident that the database examined here shows a real effect and that the hypothesis that this effect comprises a constant bitwise probability shift serves at least as well as does any alternative thus far considered.

7.1 Bayesian Calculations

The foregoing material has employed a “frequentist” approach to the interpretation of the statistical scores, by the presentation of p -values. It would be inappropriate to reprise the entire theoretical debate between frequentist and Bayesian statistics here, and rather than take sides in this dispute, PEAR has found it more productive to present results in both formalisms. The Z scores calculated above can be put to work immediately in a Bayesian hypothesis-testing framework.

Where possible, we prefer to present results as an odds ratio, or odds adjustment factor, which represents the *relative* support that a piece of empirical data provides to two well-defined hypotheses. Because this describes the proportional change that must be made in going from prior to posterior probabilities on the hypotheses concerned, it can be computed without reference to the prior probabilities. Table 10 summarizes the odds ratios for the pairwise comparisons made above.

We see that we have strong grounds for supporting BIT against DAT, and much weaker but still positive support for preferring it to TIM. The two conflicting Bayes factors, both close to 1, for the “continuation” versus “restart” interpretation of the X protocol transition, reflect the lack of conclusive information discussed in the previous section.

The Bayesian evaluation against the null hypothesis becomes slightly more complicated. In the pairwise comparisons above, the two-parameter linear regression is simply compared with two families of one-parameter models; no prior information about the model parameters is required, since they are handled in a perfectly symmetric fashion. However, once one of these models (in this case, BIT) is chosen as the favored alternative for comparison with the null, the actual odds adjustment one computes will depend on the prior assumptions one makes for the actual parameter *values*, rather than merely their constraint equations.

Because we are using a BIT model for our comparison, the variable under consideration is the probability that a binary decision chooses the option targeted by human intention. The null hypothesis predicts that this probability should be precisely 0.5. A completely uninformed BIT alternative might suppose that we have no grounds for constraining the effect of human intention on

TABLE 10
Pairwise Odds Adjustments

Comparison	Odds Adjustment Favoring BIT
BIT vs. DAT	7,700
BIT vs. TIM	5.2
BIT vs. TEL (continue)	0.62
BIT vs. TEL (restart)	1.54

Note: BIT = bitwise effect; DAT = Decision Augmentation Theory; TIM = time normalization; TEL = teleological.

physical processes and that the probability might take on any value between 0 and 1. In other words, the effect of intention is a probability shift Δp in the range $-\frac{1}{2} < \Delta p < \frac{1}{2}$. This is in many ways a maximally conservative evaluation. Nonetheless, when this is used as the alternative hypothesis, one finds an odds adjustment factor of 1658 in favor of BIT.

More informed priors produce better odds in favor of the effect. For example, one can almost casually constrain the effect of human intention on random events by making the observation that casinos make steady profits despite the presumptive (and often quite obvious) intention and desire of gamblers to win. We may therefore presume that whatever the typical scale of intentional effects on binary probabilities, it is smaller than the half-percent house advantage, which seems to be the smallest value for a commonly played casino game. (Blackjack, played by optimal strategy but without card counting, has this house advantage level.) Taken as the bitwise alternative, this means a uniform prior probability density $-0.005 < \Delta p < 0.005$, which produces an odds adjustment of 1.66×10^5 in favor of BIT.

Finally, and most apropos, a previous meta-analysis of microelectronic PK experiments (Radin and Nelson, 1989) surveyed a substantial body of prior work in the field. Unfortunately this survey includes some of the data in the current analysis; however, when the overlapping data are removed from the previous observation, the remainder provides a prior probability for the bitwise alternate that leads to an odds adjustment of 5.5×10^6 in favor of the effect.

Acknowledgments

The author gratefully acknowledges the advice and assistance of Werner Ehm of the Institut für Grenzgebiete für Psychologie und Psychohygiene (IGPP) in developing the analytical formalisms of section 5. Please direct reprints requests to York H. Dobyms, C-131 E-Quad, Princeton University, Princeton, NJ 08544-5263; e-mail ydobyms@princeton.edu. The PEAR laboratory is funded by donations from the IGPP, the Lifebridge Foundation, the Ohrstrom Foundation, and Laurance Rockefeller.

References

- Dobyms, Y. H., & Nelson, R. D. (1998). Evidence against decision augmentation theory. *Journal of Scientific Exploration, 12*, 231–258.
- Jahn, R. G., Dunne, B. J., Nelson, R. D., Dobyms, Y. H., & Bradish, G. J. (1997). Correlations of random binary sequences with pre-stated operator intention: A review of a 12-year program. *Journal of Scientific Exploration, 11*, 345–367.
- May, E. C., Utts, J. M., & Spottiswoode, S. J. P. (1995). Decision augmentation theory: Applications to the random number generator database. *Journal of Scientific Exploration, 9*, 453–488.
- Nelson, R. D. (1994). Effect size per hour: A natural unit for interpreting anomalies experiments. Technical Note PEAR 94003.
- Radin, D. I., & Nelson, R. D. (1989). Evidence for consciousness-related anomalies in random physical systems. *Foundations of Physics, 19*, 1499–1514.