

## COMMENTARY

### Comments on Freedman, Jeffers, Saeger, Binns, and Black: “Effects of Frontal Lobe Lesions on Intentionality and Random Physical Phenomena”

YORK H. DOBYNS

*Princeton Engineering Anomalies Research (PEAR)  
C-131 Engineering Quad  
Princeton University  
Princeton NJ 08544-5263  
ydobyns@princeton.edu*

**Abstract**—Freedman *et al.* present research results, one portion of which they claim to constitute a failed replication of the experimental work of Princeton Engineering Anomalies Research (PEAR) on the effects of human intentionality on physical processes, and which they suggest shows no effect due to improved experimental controls. It is shown here that the methodology used and recommended by Freedman *et al.* in fact weakens the experimental controls, rather than improving them. For all classes of potential artifactual influences, their methodology either is no better controlled, or is more susceptible to contamination, than the methodology used at PEAR. Moreover, their experimental design differs in important theoretical and operational ways from that of PEAR, and therefore does not provide a replication study. Furthermore, the statistical power of their experiment is far too small to sustain the claims they derive from it. Nevertheless, the intentional effect size generated by their normal subjects, which is the only part of their experiment relevant to the PEAR work, is actually larger than that seen at PEAR, so that if their experiment qualified as a replication, it should be considered to have reproduced the claimed effect. In addition, their own replication of a significant anomalous result by one of their brain-damaged subjects clearly strengthens the argument for some correlation of the random event generator (REG) output with the neurophysiological status of the operator. One curious omission from their report is the lack of any quantitative index of the degree of brain damage presented by their disadvantaged subjects that might be compared with their respective experimental performances.

#### Part 1: Local Control vs. Differential Tests

The assertion is made by Freedman *et al.* (2003), referencing a claim by Jeffers (in press), that the absence of immediately adjacent controls in Princeton

Engineering Anomalies Research (PEAR) experimental sessions renders their conclusions suspect. Freedman *et al.* go on to assert that a procedure that immediately follows each active experimental session, in each intention, with a control run of equal size, is superior, and to report the results of an experiment conducted with this allegedly superior methodology.

In fact the methodology recommended and followed by Freedman *et al.* can rigorously be proven *inferior* to the PEAR methodology, regardless of assumptions concerning external confounding or artifactual contributions. That is to say,

- the statistical power of the Freedman *et al.* method is uniformly less than that of the PEAR method;
- for no model of possible artifacts does the Freedman *et al.* method provide superior control to the PEAR method; and
- for some models of possible artifacts, the Freedman *et al.* method provides poorer control than the PEAR method.

### *1.1: Specification of Data Analysis and Reduction*

Freedman *et al.* spell out their data collection method in considerable detail. One session consists of ten blocks of 100 intentional trials, run with the experimental subject present, usually with the experimenter also present, followed by a 1000-trial block of control data with no one present in the experiment room. Participation was spread across several sessions, Freedman *et al.* state; since they are explicit that 1000 trials in one intention (plus the subsequent 1000 controls) completed one session, it must follow that different intentions were examined in different sessions, quite possibly with considerable time lags between sessions. A pre-specified pattern of intentions was followed for each subject, assuring that the total amount of data collected in each intention would be balanced, provided that the total number of sessions was divisible by three. (In fact, the numbers of trials per intention reported in their Table 4 makes it clear that this was not achieved with all subjects.)

The PEAR tripolar protocol, in comparison, requires that a single session collect a balanced dataset with equal numbers of high, low, and baseline trials. (Exceptions to this occurred in a few early series, where the baseline was regarded as a calibration condition rather than a third intention, and in the first version of the instructed protocol, which randomly set the intention as high or low without regard to numerical balance between the two. The quantitative imbalances induced by these earlier techniques are minor on the scale of the full database, and at no time did the protocol admit of a completely unipolar session as apparently is the norm of the Freedman *et al.* methodology.)

The fundamental statistic used by Freedman *et al.* is a Student's *t* score between the populations of intentional and control data. The fundamental statistic used by PEAR is a *Z* score between the observed mean and its theoretical value.

### 1.2: Effect of Artifact

The PEAR protocol did not involve taking calibration data in close temporal proximity to the active experiments, so Freedman *et al.* express concern about possible unknown confounding factors that could influence the output data mean, rendering a  $Z$  score based on the theoretical distribution invalid.

Although Freedman *et al.* are concerned that control data should be collected in close temporal proximity to the intentional data, their recommended methodology involves a span of approximately thirty minutes between the start of intentional data collection and the end of control data collection, with all intentional data collected in the first part of the session and all control data collected in the later part. It is rather obvious that this method of collecting controls cannot protect from artifacts that vary on timescales comparable to the session length. The consequences of artifacts with short-term variation will be discussed in Part 2. The current analysis will be restricted to artifacts that remain constant during an experimental session, since this is the only kind of artifact for which, even in principle, the Freedman *et al.* methodology might provide superior control.

The fundamental measure of effect predefined in the PEAR REG experiments, as always has been quoted in PEAR publications (Dobyns, 2000; Dunne & Jahn, 1992; Jahn, 1982; Jahn & Dunne, 1987; Jahn *et al.*, 1987, 1997; Nelson *et al.*, 2000) is the difference between the high-intention and low-intention outputs. Furthermore, as noted above, the PEAR protocol requires that the opposed intentions be run in close temporal proximity, as part of the same experimental session.

Let us consider now the data collected under the two experimental protocols. The Freedman *et al.* technique prescribes that a population of  $n_i$  intentional trials be collected, which will have some observed mean value  $m_i$  and sample standard deviation  $s_i$ . Also collected are  $n_c$  control trials with mean  $m_c$  and standard deviation  $s_c$ .

The  $t$  score for the session is then

$$t = \frac{m_i - m_c}{\sqrt{s_i^2/n_i + s_c^2/n_c}}. \quad (1.1)$$

The  $Z$  score for a single intention in the PEAR protocol makes use of the theoretical mean  $\mu$  and standard deviation  $\sigma$  of the data source, so that

$$Z = \frac{m_i - \mu}{\sigma/\sqrt{n_i}} = \sqrt{n_i} \frac{m_i - \mu}{\sigma}. \quad (1.2)$$

Some simplification may be achieved in Equations 1.1 and 1.2 by noting that the

experimental design of Freedman *et al.* calls for  $n_i = n_c = n$ , and that we are at liberty to consider PEAR results with the same  $n$  as used in Freedman *et al.* While the Student's  $t$  distribution is considerably different in theory from the  $Z$  distribution, it converges on the  $Z$  distribution as the number of degrees of freedom grows large. In the current instance no  $t$  score ever is considered that has fewer than 1998 degrees of freedom; we therefore will regard both  $t$  and  $Z$  as random variables that, under a null hypothesis, should have mean 0 and variance 1. We likewise shall ignore the difference in the higher distribution moments, which grows vanishingly small for the cases of interest.

The following analysis also shall treat the sample standard deviation  $s$  as always taking on its expected value of  $\sigma$ . The motivation for doing so is simplicity, since we thereby avoid computing the moments of ratios of random variables. The justification for doing so is that the error introduced is negligible. The random variation in both  $m$  and  $s$  is quite small relative to their absolute values. For  $m$ , where the magnitude is canceled by the subtraction of either another sample or a theoretical expectation, this smallness is irrelevant; the expectation after cancellation is zero, and no fluctuation is small when compared to zero. The  $s$  terms in Equation 1.1, on the other hand, are added rather than subtracted, and the effect of random variations will remain a very small proportional correction to the  $t$  score. (Indeed, it is this random variation in the  $s$  terms that drives the difference between the  $t$  and  $Z$  distributions, and as already noted this difference is negligible when thousands of degrees of freedom are involved.) If we apply the simplifications  $n_i = n_c = n$  (exact) and  $s_i = s_c = \sigma$  (approximate but with negligible error), we obtain the modified formula for theoretical consideration of  $t$ :

$$t = \frac{m_i - m_c}{\sqrt{2\sigma^2/n}} = \sqrt{\frac{n}{2}} \frac{m_i - m_c}{\sigma}. \quad (1.3)$$

Let us now consider a given measurement of the output mean  $m$ . We will use  $\Delta$  to label any artifactual mean shift that may be present. The symbol  $\epsilon$  will denote the real effect ( $\epsilon = 0$  under the null hypothesis), and all random variation will be collected into a noise term  $v$ . With these definitions we may write:

$$m_i = \mu + \epsilon + \Delta + v_i; \quad m_c = \mu + \Delta + v_c. \quad (1.4)$$

The noise terms are subscripted since they will vary between any pair of measurements. Any noise term is, by hypothesis, a random variable with expectation 0 and variance  $\sigma^2/n$ . The remaining terms on the right sides of Equation 1.4 are constants, not random variables, although both  $\epsilon$  and  $\Delta$  are unknown and possibly 0. Inserting Equation 1.4 into Equations 1.3 and 1.2 yields:

$$Z = \sqrt{n} \frac{-\epsilon + \Delta + v_i}{\sigma}; \quad t = \sqrt{\frac{n}{2}} \frac{\epsilon + v_i - v_c}{\sigma}. \tag{1.5}$$

It is tedious but straightforward to calculate the moments of these test statistics  $Z$  and  $t$  considered as random variables, *i.e.*,

$$\begin{aligned} \langle Z \rangle &= \sqrt{n} \frac{-\epsilon + \Delta}{\sigma}, \\ \langle Z^2 \rangle &= \frac{n}{\sigma^2} (\epsilon + \Delta)^2 + 1, \\ \sigma_Z^2 \oslash \langle Z^2 \rangle - \langle Z \rangle^2 &= \frac{n}{\sigma^2} (\epsilon + \Delta)^2 + 1 - \left[ \sqrt{n} \frac{-\epsilon + \Delta}{\sigma} \right]^2 = 1, \\ \langle t \rangle &= \sqrt{\frac{n}{2}} \frac{\epsilon}{\sigma}, \\ \langle t^2 \rangle &= \frac{n\epsilon^2}{2\sigma^2} + 1, \\ \sigma_t^2 \oslash \langle t^2 \rangle - \langle t \rangle^2 &= 1. \end{aligned} \tag{1.6}$$

Note that for a single intention in isolation, the  $Z$  score is confounded by  $\Delta$  exactly as we would expect. As was noted above, however, the standard used in the published PEAR database for an existence claim is a differential shift between oppositely directed intentions. Therefore, the statistic of interest for PEAR data is:

$$Z_d \oslash \frac{Z_+ - Z_-}{\sqrt{2}}, \tag{1.7}$$

where  $Z_+$  and  $Z_-$  are  $Z$  scores from oppositely directed intentional runs in which the effect sizes are by hypothesis  $\epsilon$  and  $-\epsilon$ , respectively. The normalizing denominator  $\sqrt{2}$  follows from the fact that each individual  $Z$ , per Equation 1.6, has variance 1. The analogous formula for the Freedman *et al.* analysis is:

$$t_d \oslash \frac{t_+ - t_-}{\sqrt{2}}. \tag{1.8}$$

By construction,  $Z_d$  and  $t_d$  have variance 1. Their expectation values are:

$$\begin{aligned} \langle Z_d \rangle &= \left[ \frac{Z_+ - Z_-}{\sqrt{2}} \right] = \frac{\sqrt{n}}{\sigma \sqrt{2}} [\epsilon + \Delta - (-\epsilon + \Delta)] = \frac{\sqrt{2n}}{\sigma} \epsilon, \\ \langle t_d \rangle &= \left[ \frac{t_+ - t_-}{\sqrt{2}} \right] = \frac{1}{\sqrt{2}} \sqrt{\frac{n}{2}} \left[ \frac{\epsilon}{\sigma} - \frac{-\epsilon}{\sigma} \right] = \frac{\sqrt{n}}{\sigma} \epsilon. \end{aligned} \tag{1.9}$$

The artifactual confound  $\Delta$  thus is completely canceled out in the statistic PEAR uses to validate existence claims. The Freedman *et al.* statistic displays no superiority in this regard. Moreover, the expected value of the corresponding test

statistic, for a given effect size, is smaller by a factor of  $\sqrt{2}$  for the Freedman *et al.* method than for the PEAR method. The Freedman *et al.* method thus suffers from a loss of statistical power. Given a real effect of the same scale, the same amount of intentional data in both methodologies will produce a final test statistic only  $1/\sqrt{2}$  times as large in the Freedman *et al.* methodology. To put it another way, the Freedman *et al.* technique must collect twice as much intentional data to have the same probability of  $\beta$  error.

### 1.3: Conclusions from Part 1

For artifacts of the class for which the Freedman *et al.* methodology is alleged to have superior controls, it is in fact no better controlled than the PEAR statistic used for evaluating existence claims. In addition, the Freedman *et al.* statistic suffers a needless loss of  $\sqrt{2}$  in its statistical resolution, leading to a halving of effective statistical power.

## Part 2: General Artifact Model

The methodological issue raised by Freedman *et al.* is that calibration data taken at considerable time separation from the experimental data may not be adequate to control for artifacts. This presupposes that a hypothetical artifactual influence may change with time, since any constant artifact would appear in calibrations and would be detected, regardless of when the calibrations were run. Therefore, to make a general assessment of the potential effects of artifacts driven by unidentified environmental influences, we must consider a time-varying artifact.

### 2.1: Derivation of Artifact Vulnerability Formula

As in the previous analysis, we will be interested only in artifactual shifts of the mean of the output distribution. Although external influences might alter higher moments of the distribution as well, these have a negligible effect on the output statistics, for the reasons given in Part 1. Let us consider, therefore, the effects of a phenomenon that influences the mean of the REG output in some arbitrary time-varying fashion, say  $A(t)$ . Any such function, regardless of the complexity of its structure, can be decomposed into sinusoidal oscillations at different frequencies:

$$A(t) = \int_0^{\infty} d\omega \alpha(\omega) \cos[\omega t + \phi(\omega)] \quad (2.1)$$

where the amplitudes  $\alpha$  and phase shifts  $\phi$  both depend on the frequency  $\omega$ . This is by no means the only possible representation of such a decomposition; for example, one might instead include both sine and cosine functions, in which case one could dispense with the frequency-dependent phase shifts  $\phi$ . The notation of Equation 2.1 is chosen for convenience in the subsequent discussion.

The advantage of this decomposition is that, rather than attempt to consider an arbitrary undefined artifact  $A(t)$ , we may restrict our analysis to a single artifact of the form  $\cos(\omega t + \phi)$ . Since any arbitrary time-dependent artifact can be resolved into such components, examining the effect of such single-frequency contributions, as a function of frequency, will allow us to gauge the relative vulnerability of different control procedures to general artifacts. Note that we do not include the amplitude in this analysis. This means that any figures that we obtain will be ratios, expressing the fraction of the original artifact amplitude that appears in the final analyses.

For both the PEAR and the Freedman *et al.* methodologies, claims are based on a difference between two datasets collected at different times. The Freedman *et al.* method compares an active intentional period with an immediately subsequent control; the active period itself is broken up into short runs with brief intervals between them. The PEAR method compares oppositely directed intentional intervals generated at different times within the same experimental session. Depending on the specific protocol of a PEAR experiment, the data for each intention might be collected in a single long block or in several short blocks within the session; the assignment of intention within the session might be either random or chosen by the operator.

Part 1 demonstrates that both methods completely cancel any time-independent artifact. For the general problem of time-varying artifacts, we must evaluate the difference between the average value of the artifact in two distinct sets of time intervals. The first step is to consider the mean value of the artifact during a single interval, say from  $t = a$  to  $t = b$ :

$$\bar{A} = \frac{\int_a^b \cos(\omega t + \phi) dt}{b - a} = \frac{\sin(\omega b + \phi) - \sin(\omega a + \phi)}{\omega(b - a)}. \tag{2.2}$$

The dependence on the interval boundaries may be separated from the dependence on  $\phi$  by using the identity  $\sin(x + y) = \sin(x) \cos(y) + \cos(x) \sin(y)$ :

$$\bar{A} = \frac{\{\sin(\omega b) - \sin(\omega a)\}}{\omega(b - a)} \cos(\phi) + \frac{\{\cos(\omega b) - \cos(\omega a)\}}{\omega(b - a)} \sin(\phi). \tag{2.3}$$

Next, let us consider the average of the artifact during several nonoverlapping intervals, the first from  $a_1$  to  $b_1$ , the second from  $a_2$  to  $b_2$ , and so forth. The average is clearly the integral of the artifact over all intervals, divided by the total length of all intervals:

$$\bar{A} = \frac{\int_{a_1}^{b_1} \cos(\omega t + \phi) dt + \int_{a_2}^{b_2} \cos(\omega t + \phi) dt + \dots}{(b_1 - a_1) + (b_2 - a_2) + \dots}. \tag{2.4}$$

Let us presume that there are  $n$  intervals, indexed by  $i$ . Integrating and collecting terms in  $\phi$  as in Equation 2.3 leads to the form:

$$\bar{A} = \left\langle \frac{\sum_{i=1}^n [\sin(\omega b_i) - \sin(\omega a_i)]^*}{\omega \sum_{i=1}^n (b_i - a_i)} \right\rangle \cos \phi + \left\langle \frac{\sum_{i=1}^n [\cos(\omega b_i) - \cos(\omega a_i)]^*}{\omega \sum_{i=1}^n (b_i - a_i)} \right\rangle \sin \phi. \tag{2.5}$$

We are concerned with the difference in the artifactual mean between two sets of measurements. Let us assume that there are  $m$  measurements in the second set, and use  $\alpha_i$  and  $\beta_i$  to denote their start and end times. (The index  $i$  is always bound to a summation, so its repeated use causes no problems here.) The mean value of the artifact in the second set of measurements obviously has the same form as Equation 2.5, with the substitution of  $m$  for  $n$  and  $\alpha, \beta$  for  $a, b$ . The difference between the two expressions is then:

$$\mathcal{D} = \left\langle \frac{\sum_{i=1}^n [\sin(\omega b_i) - \sin(\omega a_i)]^*}{\omega \sum_{i=1}^n (b_i - a_i)} - \frac{\sum_{i=1}^m [\sin(\omega \beta_i) - \sin(\omega \alpha_i)]^*}{\omega \sum_{i=1}^m (\beta_i - \alpha_i)} \right\rangle \cos \phi + \left\langle \frac{\sum_{i=1}^n [\cos(\omega b_i) - \cos(\omega a_i)]^*}{\omega \sum_{i=1}^n (b_i - a_i)} - \frac{\sum_{i=1}^m [\cos(\omega \beta_i) - \cos(\omega \alpha_i)]^*}{\omega \sum_{i=1}^m (\beta_i - \alpha_i)} \right\rangle \sin \phi. \tag{2.6}$$

Equation 2.6 applies equally to both the PEAR and the Freedman *et al.* analysis methods; the two differ only in the details of how  $a, b, \alpha,$  and  $\beta$  are determined, and in the numbers  $n$  and  $m$  of each type of interval. Note that as  $\omega \rightarrow 0$ , each individual term vanishes, and so  $\mathcal{D} \rightarrow 0$ . Thus the result of Part 1, for constant artifacts, is confirmed. Also note a common factor  $1/\omega$  in all terms, so that as  $\omega \rightarrow \infty$ ,  $\mathcal{D} \rightarrow 0$ . Thus the effect of an artifact vanishes in the limit of both very slow and very fast variation, so that the domain of interest can only be that of intermediate frequencies such that the period of variation is roughly comparable to the session length.

Equation 2.6 explicitly depends on the unknown phase  $\phi$ ; in terms of this variable, we may note that:

$$\mathcal{D} = X \cos \phi + Y \sin \phi, \tag{2.7}$$

where:

$$X \propto \frac{\sum_{i=1}^n [\sin(\omega b_i) - \sin(\omega a_i)]^*}{\omega \sum_{i=1}^n (b_i - a_i)} - \frac{\sum_{i=1}^m [\sin(\omega \beta_i) - \sin(\omega \alpha_i)]^*}{\omega \sum_{i=1}^m (\beta_i - \alpha_i)},$$

$$Y \propto \frac{\sum_{i=1}^n [\cos(\omega b_i) - \cos(\omega a_i)]^*}{\omega \sum_{i=1}^n (b_i - a_i)} - \frac{\sum_{i=1}^m [\cos(\omega \beta_i) - \cos(\omega \alpha_i)]^*}{\omega \sum_{i=1}^m (\beta_i - \alpha_i)}. \tag{2.8}$$

There are two ways of dealing with the unknown phase factor  $\phi$ . First, we may make a worst-case analysis: find the maximum possible value of  $\mathcal{D}$  for any  $\phi$ , which will occur if the artifactual contribution is at the most unfortunate part of its cycle at the start of the experiment. (Recall that we still are considering only



one frequency component of the artifactual contribution and later will expand the analysis to incorporate all frequencies.) Using the form of Equation 2.7, we easily can find  $\partial\mathcal{D}/\partial\phi$ . Setting this to zero shows that the extrema of Equation 2.7 occur when  $\tan \phi = Y/X$ . Employing the standard identities,  $\sin(\tan^{-1}x) = x/\sqrt{x^2 + 1}$  and  $\cos(\tan^{-1}x) = 1/\sqrt{x^2 + 1}$ , and evaluating Equation 2.7 leads to the result:

$$\mathcal{D}_{max} = \sqrt{X^2 + Y^2}. \tag{2.9}$$

Equation 2.9 gives the largest possible artifactual contribution for a given configuration of experimental intervals and a given time scale for variation of the artifact. An alternative to this worst-case approach is to consider the average impact of the artifact over all possible values of  $\phi$ . Since  $\mathcal{D}$  is a sum of sinusoidal functions of  $\phi$ , its mean value over a full cycle of  $\phi$  vanishes. This is what we expect, of course, since a randomly chosen phase for a variable artifact is just as likely to shift the two datasets being compared in either direction, and the average shift over all possible phases must be zero. This does not mean, however, that the average effect of an artifact can be ignored. A commonly used estimate of the scale of a random variable with zero mean is the root mean square value: that is, the square root of the mean of the squared value. (The standard deviation of a distribution is the RMS value of the difference of individual samples from the mean.) The RMS value of  $\mathcal{D}$  is:

$$\begin{aligned} \sqrt{\langle \mathcal{D}^2 \rangle} &= \sqrt{\langle X^2 \cos^2 \phi + 2XY \cos \phi \sin \phi + Y^2 \sin^2 \phi \rangle} \\ &= \sqrt{\frac{1}{2\pi} \int_0^{2\pi} (X^2 \cos^2 \phi + 2XY \cos \phi \sin \phi + Y^2 \sin^2 \phi) d\phi} \\ &= \sqrt{\frac{1}{2}X^2 + \frac{1}{2}Y^2} = \frac{1}{\sqrt{2}}\sqrt{X^2 + Y^2} \\ &= \frac{1}{\sqrt{2}}\mathcal{D}_{max}. \end{aligned} \tag{2.10}$$

Evidently the quantity  $\sqrt{X^2 + Y^2}$  is crucial, whether we wish to examine the worst-case vulnerability or the typical scale of artifactual contribution.

The form of Equation 2.8 suggests that  $X^2 + Y^2$  evaluates to a rather cumbersome expression. To facilitate the calculation, we may first employ symbolic abbreviations for the two denominators:  $D_n \oslash \omega \sum_{i=1}^n (b_i - a_i)$  and  $D_m \oslash \omega \sum_{i=1}^m (\beta_i - \alpha_i)$ . Breaking up the summations so that each sum covers only a single term,

$$\begin{aligned} X &= \frac{\sum_{i=1}^n \sin(\omega b_i)}{D_n} - \frac{\sum_{i=1}^n \sin(\omega a_i)}{D_n} \\ &\quad - \frac{\sum_{i=1}^m \sin(\omega \beta_i)}{D_m} + \frac{\sum_{i=1}^m \sin(\omega \alpha_i)}{D_m}. \end{aligned} \tag{2.11}$$

As with Equation 2.8,  $Y$  may be obtained from  $X$  by replacing all sines with cosines.

When the product  $X^2$  is formed, the products of the terms produce double sums over products of sine functions. For each such summation, the product  $Y^2$  produces a double sum over exactly the same indices and arguments, but with cosines replacing sines. Collecting the corresponding summations, one finds that each is a sum over terms of the form  $\sin \theta \sin \phi + \cos \theta \cos \phi$ , which is equal to  $\cos(\theta - \phi)$  by a standard trigonometric identity. We thus find:

$$\begin{aligned}
 X^2 + Y^2 = & \frac{\sum_{i=1}^n \sum_{j=1}^n \cos[\omega(b_i - b_j)^*]}{D_n D_n} + \frac{\sum_{i=1}^n \sum_{j=1}^n \cos[\omega(a_i - a_j)^*]}{D_n D_n} \\
 & + \frac{\sum_{i=1}^m \sum_{j=1}^m \cos[\omega(\beta_i - \beta_j)^*]}{D_m D_m} + \frac{\sum_{i=1}^m \sum_{j=1}^m \cos[\omega(\alpha_i - \alpha_j)^*]}{D_m D_m} \\
 & - 2 \frac{\sum_{i=1}^n \sum_{j=1}^m \cos[\omega(b_i - a_j)^*]}{D_n D_n} - 2 \frac{\sum_{i=1}^n \sum_{j=1}^m \cos[\omega(b_i - \beta_j)^*]}{D_n D_m} \\
 & + 2 \frac{\sum_{i=1}^n \sum_{j=1}^m \cos[\omega(b_i - \alpha_j)^*]}{D_n D_m} + 2 \frac{\sum_{i=1}^n \sum_{j=1}^m \cos[\omega(a_i - \beta_j)^*]}{D_n D_m} \\
 & - 2 \frac{\sum_{i=1}^n \sum_{j=1}^m \cos[\omega(a_i - \alpha_j)^*]}{D_n D_m} - 2 \frac{\sum_{i=1}^m \sum_{j=1}^m \cos[\omega(\beta_i - \alpha_j)^*]}{D_m D_m}.
 \end{aligned}
 \tag{2.12}$$

This rather elaborate sum can be rendered conceptually simpler if we note that, aside from the weight factors, it is simply a sum over cosines of all possible phase differences between the state of the artifactual contribution's cycle at each of the starting and ending points of a data-collection segment.

### 2.2: Application of Artifact Vulnerability Formula

To evaluate the worst-case vulnerability for the two protocols, the square root of Equation 2.12 must be calculated for values of  $a$ ,  $b$ ,  $\alpha$ , and  $\beta$  as determined by the protocol and for a wide range of  $\omega$  values. From the protocol description provided by Freedman *et al.*, we constructed an estimated pattern of  $a$  and  $b$  values for the 10 active segments, based on an assumption of an average 30-second inactive period between runs during which the experimenter conducted the brief interview described. In contrast to Freedman *et al.*, the PEAR protocol involves several different run lengths; additionally, the intention of the next run is sometimes decided by the operator, sometimes set by an external random or pseudo-random process. To deal with the effect of this variety of protocols, we calculated vulnerability functions for each major protocol used in the PEAR database and computed their mean.

Figure 1 illustrates the results of this comparison for both methods. The plotted curves show the vulnerability as a function of the period of the artifactual influence, as that period ranges from 1 minute to 120 minutes. (The period  $T$  is

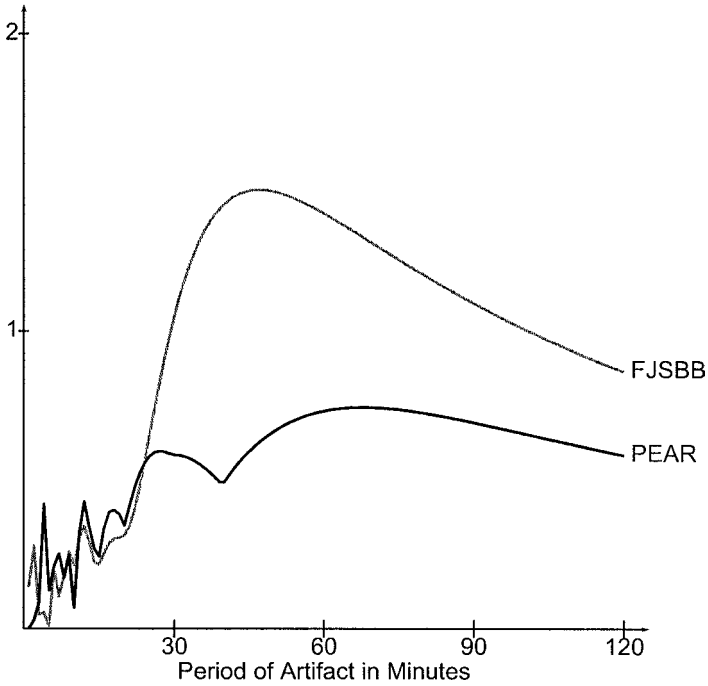


Fig. 1. Worst-case artifact vulnerability.

a somewhat more intuitive measure than the angular frequency  $\omega$ ; the two are related by  $\omega = 2\pi/T$ .) For longer periods, we note that 120 minutes is more than twice the length of either protocol, and we would expect the sensitivity to be declining as  $\omega \rightarrow 0$ , as indeed is suggested by both of the plotted curves.

Figure 1 clearly demonstrates that the Freedman *et al.* protocol, commended by them as giving superior protection against artifacts in the random noise source, is in fact *more* vulnerable to artifacts than the PEAR protocol over a broad range of artifact frequencies. Only for the fastest-varying artifacts do the protocols become comparable in their degree of vulnerability.

The vertical normalization of Figure 1 is the amplitude of the artifactual influence. Its limit is 2, not 1, because the theoretical limiting case occurs when the artifact is at a maximum (+A) during the active run and a minimum (-A) during the control run.

### 2.3: Peripheral Issues

The high-frequency artifacts examined in Figure 1, although included for completeness, should not be taken seriously as a meaningful possibility. Such short-term variations would be visible in calibration data; in fact, the calibration

data show no evidence of such short-term trends, beyond the random variation expected of the source.

One might object that the calibrations are not run in close temporal proximity to the active sessions, and that rapidly varying artifacts might be present only during experimental sessions. This may be disposed of by considering the logically possible cases. If a class of artifact appears preferentially during experimental sessions, either it must be due to the experimental procedure itself or it must be the result of an environmental influence that preferentially appears during experiments. Since experiments are not preferentially run in particular seasons, under particular weather conditions, etc., it is difficult to identify any temporal variable that characterizes experimental sessions save one: they are generally run during or shortly after the normal workday, rather than at night. Enough calibrations have been run during the day, however, to exclude the possibility of a diurnally appearing artifact.

The other possibility, of an artifact driven by the experimental session itself, can be identified most credibly with some undetermined effect of physical presence of the operator (at PEAR) or the subject and perhaps the experimenter as well (in Freedman *et al.*) in the experiment room. In this area it is clear that PEAR's data are better controlled in principle, since Freedman *et al.* have the human participants exit the room before starting the collection of control data; in contrast, PEAR's tripolar protocol not only entails the comparison of high-intention and low-intention data collected in the same session with the operator present, but also permits comparisons with baseline data, which also are run with the operator present.

Finally, one might protest that the formalism culminating in Equation 2.12 above makes an unfair comparison, in that it compares the artifact vulnerability of a single intention in the Freedman *et al.* protocol (with its adjacent control) against the differential effect between oppositely directed intentions measured in a PEAR session. This, however, is an erroneous objection. The fact that the Freedman *et al.* protocol allows intentions to be split between different sessions means that forming a differential test between intentions fails to increase the protection against artifacts. The medium-term artifacts to which the Freedman *et al.* protocol is most vulnerable (cf. Figure 1) may appear with completely different phases in sessions run on different days, or with rest breaks of unspecified length between sessions. The worst-case analysis must take into account the fact that artifacts may appear independently in *each* session of the Freedman *et al.* protocol, and are just as likely to reinforce as to cancel in a differential comparison between intentions.

#### 2.4: Conclusions from Part 2

When time-varying artifacts are considered systematically, we see that the PEAR protocol is in no case more vulnerable to artifactual influences than the Freedman *et al.* protocol, and in most cases is considerably less so. The protocol

recommended by Freedman *et al.* out of their concern for unspecified environmental influences on the random source is in fact more vulnerable to such influences than the original PEAR protocols.

### Part 3: Other Issues

Parts 1 and 2 of this discussion have focused on the issue of the different control methodologies employed by Freedman *et al.* and at PEAR. This one issue appeared to merit such extensive discussion, since it is the center of a methodological criticism of PEAR and has been the topic of an independent article submitted by one of the coauthors (Jeffers, in press). Aside from this theoretical issue, there are some other practical problems with certain interpretations made by the Freedman *et al.* work.

#### 3.1: Replication

Following are two direct quotes from the Discussion section of Freedman *et al.*:

- “We examined the claims from the PEAR program that an individual’s intention can influence the statistical distribution of random physical phenomena.”
- “Although our results did not replicate the findings reported by Jahn and his colleagues in normal subjects (Jahn *et al.*, 1987, 1997), they support their claims that intentionality can alter the output of a random event generator.”

It seems clear that Freedman *et al.* believe their experiments with their control subjects to be, at least in part, a replication of the PEAR REG work, and one that failed to replicate the results. They do note that they find statistically significant and replicable evidence of intentional effects for one of their frontal-lobe subjects, which is an intriguing result; however, it addresses the issue of replication only obliquely, since PEAR’s results were produced by operators of normal brain physiology. The results on frontal-lobe subjects represent an extension of PEAR’s work and provide valuable new information on the phenomenology of intentional effects.

#### 3.2: Methodological Differences

Even with their normal subjects, Freedman *et al.* employed a research protocol considerably at variance with that used at PEAR, quite aside from the issue of experimental controls addressed in Parts 1 and 2. A majority of the normal subjects were research staff, and by inspection of Table 4 in Freedman *et al.*, this majority also was responsible for a large majority of the data; neither of these is the case in PEAR’s operator population. The non-staff subjects of Freedman *et al.*, in an even greater departure from PEAR procedures, were accompanied by an examiner in much the same way as the frontal-lobe patients.

Although we can understand the usefulness of maintaining the same experimental procedure, in considering the normals as controls for the frontal-lobe patients, the presence of a second person in the room during the data collection raises a serious question regarding who is actually the subject in these experiments. PEAR's experimental program finds substantial differences between the performance of individual operators and the joint performance of two or more simultaneous participants (Dunne, 1991).

Even if we grant that the experimenter can avoid participating in the experiment itself, we must still consider the psychological effect of the social dynamic established between the experimenter and the subject. Freedman *et al.* give a description that clearly establishes the examiner as an authority figure who instructs the subject and controls the experimental operations. Furthermore, the instructions given to subjects, quoted in Freedman *et al.*, display a general tenor of skepticism and hostility to the phenomenon. "There are some people who believe that . . .," "We would like to see if there is a possibility . . .," "I want you to try . . . as much as possible."

PEAR has presumed from the outset that a relaxed and supportive atmosphere, in which no air of the bizarre or outré attaches to consciousness-related effects, is psychologically important to the operators' ability to induce those effects. The experimenter who, however graciously, would like to see if there is a possibility that a phenomenon in which some people believe actually occurs is setting a very different attitudinal context for the experimental subject. If this politely incredulous approach in fact describes the opinions of the experimenters, the unaccompanied, research-staff subjects probably will be approaching the task with a similar psychological barrier against anomalies.

### 3.3: *Inadequate Statistical Power*

As noted above, the only portion of the Freedman *et al.* database directly relevant to PEAR's empirical claims is the subset generated by normal subjects. This statistical universe comprises six individuals and, as shown in Table 4 of Freedman *et al.*, a total of 94,000 intentional trials (plus an equal number of non-intentional controls, plus approximately half that number of baselines and baseline-controls that are irrelevant to the evaluation of an intentional mean-shift). The analogous database from PEAR, namely that subset of REG trials run locally with the operator on-site, comprises over 1.6 million intentional trials from 91 operators (Jahn *et al.*, 1997). Simple comparison of the numbers might lead one to suspect a potential problem with statistical power.

Such a suspicion would be entirely justified. The comparison database from PEAR attains a total  $Z$  score of 3.809 for the high-minus-low intentional comparison. If exactly the same effect size were to manifest in a population of only 94,000 trials, the expected  $Z$  score would be only 0.901. But, to make matters worse, as was pointed out in Part 1, the local-control comparison procedure of Freedman *et al.* reduces the statistical power of the test by a factor

of  $\sqrt{2}$ . The real prediction, then, if the normal subjects produced exactly the same intentional effect as PEAR's operators, would be for a composite score for all normal subjects pooled of only 0.638, with so many degrees of freedom that the  $t$  and  $Z$  distributions are effectively indistinguishable. It is straightforward to calculate that the chance of Type II error (mistakenly failing to reject the null hypothesis, when it is in fact false) is 0.843. Let us emphasize the point: if we postulate that the effects seen at PEAR are real, and are in fact present to exactly the same extent in the Freedman *et al.* study despite the methodological differences, *we would expect Freedman et al. to fail to attain statistical significance 84.3% of the time because their database is so small.*

In fact, the statistical power situation is even worse than that, due to the way in which Freedman *et al.* choose to employ Bonferroni corrections. The Type II error rate calculated above is that for a one-tailed 5% significance criterion, applied to the differential test of intentional effect, for the pooled normal subjects only. In contrast, the discussion in Freedman *et al.* pursuant to their Table 3 makes it clear that they examine each intention independently, without regard to direction of intention (thus requiring two-tailed  $p$ -values), and apply a Bonferroni correction for the number of datasets examined (five) and the number of intentions (three). Let us recall that the frontal-lobe subjects represent an extrapolative test of a new hypothesis, for which the PEAR database *makes no prediction whatever*. In the normal subjects, the effect seen at PEAR would predict  $t = 0.451$  in each intention separately. The Bonferroni correction requires  $p = 0.003$ , two-tailed, for any single test to be reckoned significant. From this it can be calculated that the probability of Type II error on either the right or the left intention is 0.994, and the probability that both will jointly fail to detect a real effect is just the square of this:  $\beta = 0.988$ . *The test applied by Freedman et al. to the data presented in Table 3 is 98.8% likely to overlook a real effect in normal subjects of the magnitude seen at PEAR.* The fact that one result in their Table 3 actually manages to achieve statistical significance despite this stringent criterion may be intriguing in its own right, but since it derives from the performance of the frontal-lobe patients, it is irrelevant to the validity of PEAR's findings.

The actual statistic for the intentional performance of all normal subjects, pooled, can be extracted from Table 3 of Freedman *et al.* quite directly. For the pooled normals, they report  $t = -0.8538$  with 95998  $df$  in the right intention, and  $t = 1.1474$  with 91998  $df$  in the left intention. The sign convention of Freedman *et al.* is such that both of these results are in the direction of intention; that is, the right intention is to the right of (numerically larger than) the controls, and the left intention is to the left of (numerically smaller than) the controls. Neglecting the difference (a matter of a few percent) between the amount of data in the two intentions, we may calculate the differential-effect statistic  $t_d$  (Equation 1.8) as  $t_d = (0.8538 + 1.1474)/\sqrt{2} = 1.415$ . Taking the different amounts of data into account corrects this to 1.413. Recall that the predicted  $t_d$  for an effect identical to PEAR's appearing in a database of this size is only 0.638; the Freedman *et al.*

experiment is actually seeing an intentional effect more than twice as large as that seen at PEAR! Granted, this fails to achieve statistical significance due to the small size of the database, and moreover is of questionable validity as a replication due to the protocol differences noted in section 3.2 above. It nevertheless is intriguing to note that if one were to stipulate the validity of the Freedman *et al.* protocol, their normal-subjects database would be interpreted by reasonable analysts as a promising first stage in a *successful* replication. Indeed, from a hypothesis-testing viewpoint it *is* a successful replication: Bayesian evaluation of the Freedman *et al.* data produces a factor of 2.01 in favor of the PEAR alternative over the null hypothesis. That is to say, the Freedman *et al.* result, despite its lack of significance, approximately doubles the posterior probability that the PEAR result is a real effect.

### 3.4: Implications of Data from Brain-Damaged Patients

As noted above, the data from normal subjects are the only part of the database relevant to a replication of PEAR's work. However, there remains a result of considerable interest and value in their work with their brain-damaged subjects, namely the replicated anomalous performance of one of these. Although the total amount of data accumulated from these subjects is modest, leading to statistical power concerns similar to those detailed above for the normal subjects, this particular one nevertheless achieved results that were significant after Bonferroni correction, and replicated those results in a repeat study undertaken after the initial analysis.

Given the low statistical power of the experiment, this result suggests an anomalous effect substantially larger than that seen in PEAR's experiments with normal operators. Although the Freedman *et al.* subject population is far too small to sustain any theoretical generalizations, these results would seem to encourage more extensive empirical studies of this sort. In this context, it is curious that Freedman *et al.* provide us with no quantitative index of the *degree* of frontal lobe damage that characterizes each of their subjects. A correlation of such a measure with the degree of apparent anomalous performance, even if the latter is not independently significant for an individual, might provide far more useful insights into the nature of the phenomenon. It is to be hoped that such a quantitative indicator will accompany their future studies.

### 3.5: Conclusions from Part 3

It is both inappropriate and incorrect for Freedman *et al.* to assert, as they do in their discussion section, that their experiment fails to replicate PEAR's findings on normal subjects. Their methodology differs from that used at PEAR in ways that seem likely to affect the outcome. The statistical power of the Freedman *et al.* database from normal subjects is far too small to constitute a valid test of the PEAR claim, with the probability of Type II error standing at an astounding  $\beta = 0.988$ . Despite their assertion of failed replication, the



Freedman *et al.* database on normals not only is consistent with PEAR's results but in fact contains a larger intentional effect than that observed at PEAR. Finally, while their significant results with brain-damaged subjects are promising, much better use could be made of those REG data if they were accompanied by a quantitative index of the degree as well as the class of damage.

### References

- Dobyns, Y. H. (2000). Overview of several theoretical models on PEAR data. *Journal of Scientific Exploration, 14*, 163–194.
- Dunne, B. J. (1991). *Co-Operator Experiments with an REG Device*. Technical Note PEAR 91005.
- Dunne, B. J., & Jahn, R. G. (1992). Experiments in remote human/machine interaction. *Journal of Scientific Exploration, 6*, 311–332.
- Freedman, M., Jeffers, S., Saeger, K., Binns, M., & Black, S. Effects of Frontal Lobe Lesions on Intentionality and Random Physical Phenomena. *Journal of Scientific Exploration, 17*, 651–668.
- Jahn, R. G. (1982). The persistent paradox of psychic phenomena: An engineering perspective. *Proceedings of the IEEE, 70*, 136–170.
- Jahn, R. G., & Dunne, B. J. (1987). *Margins of Reality: The Role of Consciousness in the Physical World*. Orlando: Harcourt Brace Jovanovich.
- Jahn, R. G., Dunne, B. J., & Nelson, R. D. (1987). Engineering anomalies research. *Journal of Scientific Exploration, 1*, 21–50.
- Jahn, R. G., Dunne, B. J., Nelson, R. D., Dobyns, 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.
- Jeffers, S. (in press). Physics and claims for anomalous effects related to consciousness. *Journal of Consciousness Studies*.
- Nelson, R. D., Jahn, R. G., Dobyns, Y. H., & Dunne, B. J. (2000). Contributions to variances in REG experiments: ANOVA models and specialized subsidiary analyses. *Journal of Scientific Exploration, 14*, 73–90.