JOURNAL OF SCIENTIFIC EXPLORATION
A Publication of the Society for Scientific Exploration

Volume 21, Number 2 2007

Page

Editorial
229 Editorial
Henry H. Bauer

Column
231 Inference and Scientific Exploration
Mikel Aickin

Dinsdale Award Lecture
241 The Role of Anomalies in Scientific Exploration
Peter A. Sturrock

Research Article
261 The Yantra Experiment
Y. H. Dobyns
J. C. Valentino
B. J. Dunne
R. G. Jahn
Suzel Fuezeau-Braesch
Jeant-Baptiste Denis

281 An Empirical Study of Some Astrological Factors in Relation to
Dog Behaviour Differences by Statistical Analysis and Com-
pared with Human Characteristics
Lynne I. Mason
Robert P. Patterson
Dean I. Radin

295 Exploratory Study: The Random Number Generator and
Group Meditation
Mikel Aickin

318 Commentary: Comments on Mason, Patterson and Radin
Mikel Aickin

325 Statistical Consequences of Data Selection
Y. H. Dobyns

354 Commentary: Comments from Mikel Aickin
Mikel Aickin

Letters to the Editor
353 The Wave Function Really Is a Wave
Robert D. Klauber

357 Hocking's Response to Klauber
M. G. Hocking

Obituary
361 In Memoriam: George Sassoon, 1936–2006
Ronald N. Bracewell

Review Essay
365 Stagnant Science: Why Are There No AIDS Vaccines?
Henry H. Bauer

373 Earthquake Prediction, Kooks, and Syzygy: A Review of The
Man Who Predicts Earthquakes
David Deming

383 Review of The Man Who Predicts Earthquakes: Jim Berkland,
Maverick Geologist: How His Quake Warnings Can Save Lives
by Cal Orey
Patrick McClellan

Book Reviews
397 Trusting the Subject (Vols. 1 and 2), by Anthony Jack and
Andreas Roepstorff
Etzel Cardeiia

402 Miracles: A Parascientific Inquiry into Wondrous Phenomena;
The Search for Yesterday: A Critical Examination of the
Evidence for Reincarnation; and Our Psychic Potentials, by D.
Scott Rogo
Analisa Ventola

405 Tectonic Consequences of the Earth's Rotation, by Robert C.
Bostrom
Karsten Storetvedt

413 Unstoppable Global Warming Every 1,500 Years, by S. Fred
Singer and Dennis T. Avery
Joel Kauffman

416 The End of the Certain World, by Nancy Thorndike
Karl Gustafson

422 Science Sold Out: Does HIV Really Cause AIDS?, by Rebecca
Culshaw
Henry H. Bauer
427 The Cult of Pharmacology, by Richard DeGrandpre
430 The Era of Choice: The Ability to Choose and Its Transformation of Contemporary Life, by Edward C. Rosenthal
431 Further Books of Note
433 Articles of Interest
434 Letters to the Book Review Editor

**SSE News**
435 Society for Scientific Exploration Officers and Council
EDITORIAL

Apologies once again to three sets of authors whose manuscripts are still being delayed in publication; the backlog will likely be exhausted in the next two issues.

Aickin’s column should be taken to heart not only by those who write but also by those who read material that dallies with statistics: the column can alert us to ways in which we may be misled by the manner in which results are reported or not reported. I also recommend a piece posted on Aickin’s website, "The Mystogram" (at www.ergologic.us), which sets out very nicely how misleading can be certain ways of presenting data graphically. Some classic books describing reasons why statistics is sometimes described as a way to lie are Best (2001, 2004) and Huff (195411993).

The written version of Peter Sturrock’s Dinsdale Award lecture certainly fulfils the promise of the earlier oral presentation. It underscores the authentic understanding of scientific activity that inspired Peter to found the Society for Scientific Exploration. I wish I had thought of his cogently descriptive and self-explanatory terms "OK anomalies", "not-OK anomalies", and "sleeping anomalies"; they will allow future authors to avoid some of the jargon that has often made discussions of these matters opaque to non-specialists. A substantive point that Peter has often emphasized, but that I have not seen put so plainly and forcefully elsewhere, is the need to draw up a complete set of hypotheses.

The four research articles in this issue feature three reports that directly illustrate why this Society and this Journal are needed: careful investigations are described on matters whose discussion is excluded from mainstream scholarly and scientific periodicals. The fourth article, by Dobyns with commentary by Aickin, goes to the heart of issues of validity in studies of these kinds.

There is some instructive to-and-fro in the Letters section, and the Book Reviews offer, as usual, informative descriptions of and comments on a rich variety of books.

I am already anticipating the enjoyment and stimulation to be experienced at our next annual meeting, less than a month away as I write. Those of you have not yet been to one of these meetings don’t know what you are missing, and you should make every effort to attend. It’s quite addictive.

References

A person closely associated with JSE posed the following question. When people report statistics in an article, what should they include? This is an area in which there is obviously enormous room for opinion, which immediately induced me to try to provide an answer. As you will see, there is no perfectly right answer, but there are several partial answers that I think frontier scientists might want to keep in mind. Everything said here represents my views, not necessarily those of anyone else associated with JSE, so prospective authors should not take any of this as being prescriptive.

Answer 1: Show All the Data

Perhaps due to their academic training, many scientists seem to believe that reduced data are more prestigious than raw data. It becomes important to them, therefore, to show their mastery of statistical methods by reducing their data reports to the near vanishing point, their work richly justified by obscure symbols decorated with statistical jargon. In extreme cases the report may consist of nothing but p-values. I will admit that when I see this, my immediate impression is that it comes from a lack of understanding of how data analysis works, or possibly that there is something to hide, but this may depend on my mood of the day.

In many cases, frontier research studies involve rather small datasets. In these cases, it can actually take up less journal space to publish the entire dataset than it does to roll out the statistical analysis tables. I recall reviewing an article for another journal in which the authors filled pages and pages with largely meaningless analysis of variance (ANOVA) tables, obviously copied literally from the computer output, which shed almost no light on their results. When I checked the sample sizes, I figured they could just publish their data in much less space than the tables took up. Since some analytic tables are certainly useful in such publications, the recommendation is probably better phrased as publishing the entire dataset in addition to a small number of analyses.

Another reason for full publication is that it is now easy to scan a journal page into a document and then copy and paste it into a text document, from which it
can be read into any decent statistical analysis program. This provides an invaluable opportunity for readers to try out their own analysis on the actual data, in order to see whether they agree with the analyses put forward by the authors. My experience has been that in any dataset of reasonable complexity, there are several stories about what the data mean, which are for all intents and purposes equally supported by the raw data. Sometimes these stories differ in major ways, leading to incompatible interpretations of the same underlying data. My feeling is that in these cases the scientific path is to give expression to these stories, even though they might vary. The problems are that (1) conventional science journals do a poor job of tolerating ambiguity in experimental results, and (2) in order to tell the different stories you have to actually be able to find them, and here is where having different people scrutinize the same data becomes useful.

Of course the third problem is that authors generally do not want their raw data re-analyzed by others. In some cases, I believe, it is because they are not fully confident in their own ability to analyze data, and they would prefer not to advertise any deficiencies in their approach and technique. Although it may seem harsh, I have little sympathy for this attitude. Impeachable analyses that cannot be checked constitute friction in the system of science, in the sense that they slow down real progress.

As an example of full disclosure, I have used an article from JSE (Bunnell, [1999]. Journal of Scientific Exploration, 13, 139–148) in my classes to illustrate linear regression, in part because the author published all the data. The experiment was designed to determine whether healing with intent had an effect on the rate of pepsin enzyme activity in vitro. When I plotted differences (healed minus paired unhealed) reaction rates against the time of the experimental replications, it was clear that there was a downward trend in the healing effect, a trend which all but vanished by the time of the last experiment. When I first saw this, I suspected that what was really going on is something that we see often in biomedical research: initially published results become successively weaker in subsequent publications. One prominent university in the eastern US has become famous for aggressively analyzing data, then skimming off the interpretable and statistically significant results for prestigious publication. A time-trend of subsequent work in the area shows the same kind of decline as in the healing intent experiment. In this case the eastern researchers are probably exploiting random variation for professional purposes, instead of trying to minimize chance influences. More generally, it is usually true that medical drug studies overestimate the beneficial effects of the drugs, in that later studies, in more representative patients, almost automatically show poorer results. In any case, when I presented the healer example to a class of alternative medicine practitioners, one of them pointed out that it was common knowledge that even very successful healers tend to get "tired" when they are called on to repeat results, thereby becoming less effective over time, and so my exploitation-of-chance interpretation was by no means the only explanation. I think this
Fig. 1. Histograms of two samples.

illustrates the value of full publication, because I was able to see something that
the author had missed, and by showing it to others I was able to see an
interpretation that I had missed.

There is another very important point in this example. No individual study
ever pronounces the final word on a scientific issue. It is cumulative experiments
within a program of research, and replication by others, that finally lead to
scientific acceptance. It then becomes critical to learn from prior experimen-
tation how to design future experiments, both to avoid problems and to correct
for difficulties that were previously found but unanticipated. The nearly
universal rule today seems to be that authors publish almost no information that
would help anyone else to design future research. Whether this is culture or
perversity is irrelevant, because simply making the raw data available would go
a long way toward solving the problem.

Answer 2: Graphics

Of course in many cases Answer 1 is of no use, due to the size of the dataset.
It has always seemed to me that graphical options are then rather attractive.
Perhaps the easiest graphic is the histogram. Figure 1 shows two histograms of
a measurement carried out in two groups. It clearly shows that Group 1 has
larger values than Group 2, and in particular, the extreme positive values in
Group 1 are not found in Group 2. This is an important fact, one that is hard to
see without a histogram. I will, however, point out that while histograms are
much used in elementary statistics texts, and although good data analysts employ
them frequently, they appear rather rarely in scientific publications. My opinion
is that histograms usually show how much natural variability there is in various
measurements, and scientists often regard this as an embarrassment, which they
should therefore avoid publishing. Because I think natural variability is
important, I encourage the publication of histograms.

Another way to portray the same data is in a dotplot (Figure 2). This simply
amounts to turning the histograms on their sides, but it facilitates visual
comparison, and often works well for multiple groups. It is perhaps worth
pointing out that one could reconstruct the original data (at least approximately) from the dotplot. When the dotplot is too messy, one can turn to the boxplot (Figure 3). Here the upper and lower edges of the box are at the 75th and 25th percentiles of the data, and the horizontal bar in the box denotes the median. The
extensions above and below the box are calculated as follows. Go out to the most extreme value that is within 1.5 box-heights from the upper (or lower) edge of the box and draw the horizontal line (then connect it to the box). Any values beyond this are denoted by circles, since they are conventionally interpreted to be outliers. The boxplots also show that Group 1 tends to have larger values than Group 2, although some of the detail is obscured. I have found many publications in which the interpretation of the boxplot is mangled (the upper and lower extensions are often said to be 95th and 5th percentiles, and at least one statistics package actually computes them this way).

Graphics can sometimes be very useful for portraying values of individual experimental units as they evolve over time. It is customary to plot the time-trajectories of the means in these cases, but what usually happens is that almost no individual time trajectory ever follows the pattern of the mean trajectory. Thus, stories about what happened to individuals come to be confounded with stories about how the group mean varied. In some of our studies we have segmented our sample into those who did well, mediocre, or poorly, and have shown the individual trajectories in three panels. By cutting down the number of individuals, the distracting variation is reduced. Again, I am not recommending graphics as a complete replacement for analysis and tables, but rather as a useful adjunct to give the reader a sense of what happened in a more immediate way.

**Answer 3: Small Number-Summaries**

If one wanted, the boxplot could actually be reduced to five numbers: median, 25th and 75th percentiles, and the upper and lower effective ranges (the horizontal lines on the extensions). In data presentations, however, one more frequently sees means and standard deviations (SDs), and occasionally ranges (either the largest and smallest values, or the difference between them).

The SD deserves some comment. Although any probability distribution has a SD, it is most interpretable when the distribution is at least approximately symmetric and sort of lump-shaped. Thus, giving the SD (as opposed to the 25th and 75th percentiles) for a skewed distribution is not enormously helpful. From the mathematical viewpoint, the SD is important because it plays a role in the large-sample distributions of many statistical estimators. The most familiar is the mean (or average) of a sample. The sample mean has a probability distribution (due to the fact that the sample was obtained by some chance mechanism), and under the usual assumptions its distribution has a standard deviation, which is $SD/\sqrt{n}$ (where $n$ is the size of the sample and SD is the standard deviation of the population from which the sample is selected). This leads to a terminological problem; there is a standard deviation both for the population distribution (SD) and for the probability distribution of the mean ($SD/\sqrt{n}$). Early on in the statistical literature, these were distinguished by calling $SD/\sqrt{n}$ the standard error, abbreviated SE. Most definitions you can find for the SE give it this way. Now the plot thickens. There are lots of estimators other than the mean that have
large-sample distributions that are approximately Normal and whose standard deviations are thus important for statistical inference. I use SDE (standard deviation of the estimate) for these standard deviations of estimates. The mean is only one example, where SE and SDE are the same. But in other cases (like regression coefficients, for example), the formula for the SDE is not the same as for the SE. Overwhelmingly, people still use SE for SDE, even though it is technically incorrect. On almost any output from a computer package what is labeled SE is actually SDE.

The practical problem is, which should be reported, SD or SDE? I think the answer depends on the purpose. If you want to characterize how variable the population is with respect to what you are measuring, then SD will do it (if the distribution of that measurement is roughly symmetric). If, on the other hand, you want to say how precise some estimate (like a mean or regression coefficient) is, for the purpose of statistical inference, then SDE will do it. This is because the large-sample distribution of these estimates is Normal, which is symmetric. In some cases both purposes are important, and so reporting both is useful, especially if they are labeled correctly.

In a different but related direction, one often sees experiments in which changes are the issue. We have two groups, we treat them differently, then we look to see how the changes in the two groups compare. This paradigm is extremely common in biomedicine. It is, however, almost never true, even in biomedicine, that researchers report a complete set of statistics. One complete set includes mean before intervention, mean after intervention, SD before intervention, SD after intervention, and correlation between pre- and post-intervention measurements. There are other sets of five estimates that are equivalent to these, but for simplicity it would seem that reporting at least these five should be a bare minimum. Another measure (which can be derived from the above five) is the SD of a change score. I think this is important because people tend to give too much attention to the mean change, as if everyone changed according to the mean, and the SD of change appropriately points out that changes did actually vary, and by how much.

If we go beyond means and SDs, in general statisticians use mathematical models to analyze more complex data. The idea is that the model incorporates features that we expect to be present, and so an analysis guided by the model will be appropriate for the actual data. The issue is not (as commonly thought) that the model must be a perfectly correct accounting for how the data were generated. It only needs to embody the most important features, which would result in biases if they were ignored. One example is when one has measurements on multiple people, but there is reason to suspect that the measures will be correlated (sometimes influences happen to groups, not just to individuals in an independent fashion). In cases such as these there are models that allow for the intercorrelation, and models that do not can produce biased estimates, or overstate the precision of estimates, or both. In all of these cases the model has one or more parameters (unknown constants) that reflect something of
importance in the experiment. Indeed, designing a good analysis depends on setting up parameters to capture the effects of interest. Then, the computer output estimates the parameters and provides their SDEs. Reporting the estimates and their SDEs then seems like the most sensible way to proceed. Examples of this type are multiple regression (for measured outcomes), Poisson regression (for counts of events), exponential regression (for times until a target event happens), and logistic regression (for yes/no outcomes). In each case, the parameter of interest intends to measure how much influence some explanatory variable has on the outcome, if all other explanatory variables could (magically) be held fixed.

Parameter estimates and the SDEs may seem remote from the actual underlying data, because they concentrate the information in the experiment so completely on the theoretically important issues. Therefore, it does not seem out of place to ask for both some representation of the actual data (as I have suggested above) and estimates of crucial parameters.

While there is no single definitive reference for statistical summaries and presentations, one that is well worth considering is *Statistical Rules of Thumb*, by Gerald van Belle (Wiley, 2002). I very strenuously disagree with some of his advice, but on balance the good ideas seem to me to much outweigh the not so good ideas.

**Methods to Avoid**

Because inference is an activity that is widely dispersed across multiple disciplines, various conventions tend to arise in different areas, take hold, and become dominant. Some of these are actually contrary to the spirit of inference, but since they never go through any formal review process, they establish themselves nonetheless.

One of these arcane practices is the over-reliance on p-values and "statistical significance". I have seen data presentations in which only p-values were presented, with no actual estimates of the parameters from the underlying models, which were designed to capture the effects of interest. Actually, I have seen a lot of such presentations. At one point in my career I began supervising a group of data analysts who would not report parameter estimates to the scientists who paid them unless the corresponding p-values were below 0.05. It is now a widely established practice in conventional journals to require the proliferate display of p-values, irrespective of whether they are appropriate or useful. This abuse naturally leads to others, such as reporting a parameter estimate and then decorating it with one or more asterisks to indicate whether its p-value is below 0.05, 0.01, or 0.001 (or some such scheme). One can only imagine that the next step will be to eliminate all estimates, and just report asterisks.

Connected to this is the reporting of numbers that are at most ancillary to the analysis. Examples are t-statistics and F-statistics, which are in fact merely
intermediate figures necessary to compute p-values. The fact that computer output displays them seems to promote the idea that they are important and that they belong in publications. I see them only as numbers that take up space in data tables, space that is more deserved by other numbers.

I have said above that whenever possible one should show all data, or graphical representations of all the data, along with something like the five-number summary. In this latter case, a method to be avoided is the reporting of the one-number summary, a mean or median with no indication of variability. Equally unrecommended is the three-number summary consisting of the median and the smallest and largest values. The problem is that the extreme values are influenced both by variability and the size of the sample, whereas the sample size measures the size of the sample and the SD or 25th and 75th percentiles measure variability.

Despite my preference for graphics, there is one extremely frequent graphic that has always seemed ludicrous to me. Instead of a boxplot, a solid bar (as in a histogram) is erected to connect zero with the mean of some measurement, and then a little TV antenna (like the upper part of a boxplot) is stuck onto the top. I have called this a "mystogram" because it does more to mystify than clarify. Readers who enjoy statistical satire may want to visit my website (www.ergologic.us) to read my entire (multiply rejected) article on the mystogram.

Another large area in which I think researchers have been misguided is the ANOVA. This encompasses a large number of techniques, many of which were invented by R. A. Fisher, the premier statistician of the 20th century. Social scientists are especially highly trained in this arena, and so they tend to want to look at every problem of statistical analysis through an ANOVA lens. From my standpoint, their chief failing is that they spend far too much time citing significant p-values for things such as "main effects", "interactions", and "time-by-group interactions", without giving proper attention to what actual effects have been estimated. Thus, a significant "time-by-group interaction" is very frequently taken to be a successful result, without consideration of whether the actual group effects over time are beneficial or uninterpretable and whether this procedure has reasonable statistical power (the conventional computation does not take into account the direction or pattern of the results). Even more damaging, in most approaches to ANOVA the numbers of individuals in experimental cells determine the definition of the parameters, so that experiments with different cell numbers are actually incomparable, because they apply to different parameters. In virtually every case, there is a (generalized) regression model that will do far better than the ANOVA model it replaces. See an earlier column of mine for suggestions on assessing change using a regression model (Aickin, M. [2004]. Journal of Scientific Exploration, 18, 361–367).

Along this line, it may be worth pointing out that John Tukey (inventor of the boxplot) once formed a "committee for the suppression of the correlation coefficient". This was because the correlation coefficient removes the scales on which the underlying variables are measured, whereas including them (using...
regression coefficients, for example) is the real purpose of science. Evidently Tukey’s committee did not succeed.

Yet another strange practice that has arisen lately is the reporting of "effect sizes". This is an estimate of an effect on some measurement divided by the SD of that measurement (for example, difference between post- and pre-means, divided by SD of the pre-measurement). The problem with this practice is immediately obvious; the pre-measurement SD will vary from population to population, so that one is actually measuring the effect with a ruler that changes its length, depending on the study. How this ever came to be considered a sensible strategy completely escapes me. (In fact, "effect size" is an invention to produce plausible sample-size calculations in grant applications when one does not know the relevant SD; in other words, it covers up ignorance, which is completely unnecessary after one knows the SD.) The situation is made considerably more murky by the fact that in some models (unlike ordinary regression) the SD is functionally related to the effect, rendering "effect size" all but meaningless for the purposes of inference.

**Why Is This Hard?**

One of the advantages of working in a narrow scientific area is that researchers can agree on their basic definitions, what paradigms they use to view experiments and reality, and how they communicate with each other. Statistical inference belongs to all empirical sciences in which there is any appreciable amount of unexplained variability. Thus there are many voices competing to have their views of statistical inference taken seriously, and perhaps even to dominate practice in other fields. It is, therefore, no wonder that there will be a certain amount of confusion about the question, "What statistical data should I present?"

I have not exhausted the problems or possibilities here, but the question is a good one, and probably I will have an opportunity in other columns to continue to give my thoughts on it. I re-emphasize that the opinions here are mine, and not necessarily shared by anyone else connected with JSE.

**Summary**

The problem of deciding what data or statistics to present in a research article is not easy to resolve. Above all, honesty is paramount, and whatever fairly reflects what was found is on the right track. Sometimes it will be better to present all the data, other times careful graphical displays will be best, and inevitably at other times only succinct, jejune summaries can be provided. In the end, good inference tells a story about what happened, and perhaps how or why it happened, and so long as the numbers selected for presentation are motivated by doing this well, they will be appropriate.
The Role of Anomalies in Scientific Research

PETER A. STURROCK

Center for Space Science and Astrophysics
Stanford University, MC 4060, Stanford, CA 94305
e-mail: sturrock@stanford.edu

Abstract — Anomalies play a key role in science, in calling into question some established belief: an anomaly is an anomaly only with respect to some hypothesis, theory, or belief system. Some anomalies (OK Anomalies) are greeted with interest and investigated vigorously, some (Not-OK Anomalies) are avoided or viewed with suspicion, and others (Sleeping Anomalies) may for some time go unnoticed. In this article, anomalies are viewed from the perspective of scientific inference. This requires that we compare the anomaly with a logically complete set of hypotheses, and that assessments of the evidence for the anomaly, and of its compatibility (or incompatibility) with various hypotheses, be expressed in terms of probabilities. Some anomalies may present a challenge to our "model of reality." (These are normally viewed as "Not-OK.") Identifying our "standard model of reality" makes it possible (and necessary) to identify alternative models so as to form a logically complete set of hypotheses.

Keywords: anomalies — inference — probabilities

The word "anomaly," according to Webster, is derived from the Greek "an" [="not"] and "homalos" [="even"] and signifies a "deviation from the common rule," or "something out of keeping, especially with accepted notions of fitness or order." In referring to an anomaly in science, we think first of the former, manifestly intellectual definition — a result in scientific research that does not conform to expectations based on the prevalent theory. However, members of this Society will be well aware that anomalies also have a sociological import — they may be "out of keeping with accepted notions of fitness or order." Some anomalies may be viewed primarily as intellectual challenges, but other anomalies may be in part a political challenge, in that the weight given to an anomaly depends on the status within the scientific community of the person proposing the anomaly. (Yes, dear reader, there are heresies and heretics in science, as well as in religion [Mellone, 1959].)

Anomalies should be the life-blood of science. Niels Bohr once said that "progress in science is impossible without a paradox," and Richard Feynman (1956) once remarked that "The thing that doesn't fit is the thing that is most interesting." More recently, Jahn and Dunne (1997) have written "... good science, of any topics, cannot turn away from anomalies; they are the most precious resource, however unrefined, for its future growth and refinement."
The first thing to note about "anomaly" is that it is a relative concept, not an absolute concept. A result is an anomaly only with respect to a given theory or hypothesis. In scientific research, it would be an experimental or observational result that is not in accord with current theory. Therein lies its importance. An anomaly provides a test of a theory. As Feynman’s remark implies, it is much more important to search for facts that do not agree with current theory than to find further facts that do agree with that theory. If a certain fact, which is incompatible with a given theory, can be firmly established, then that theory must be modified or abandoned.

We can consider one or two historical examples. In 1919, Ernest Rutherford gave one of his assistants, E. Marsden, the task of studying the scattering of alpha particles by a gold foil (Whittaker, 1953: 20). According to the prevalent "plum pudding" model, an atom was composed of electrons immersed in a blob of positively charged matter. According to this model, alpha particles should suffer only slight deflection in traveling through gold foil. Rutherford was astonished to find that some alpha particles were backscattered from the foil. He said "It was as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you." It took Rutherford over a year to digest the implications of that anomaly. He finally concluded, correctly, that the positive charge in an atom is concentrated in a very small space at the center of the atom.

But not everyone responds to an anomaly in such a direct and productive manner. Roentgen recognized an anomaly when a piece of paper painted with barium platino-cyanide fluoresced when current was passed through an adjacent Crookes tube (Whittaker, 1951: 357). However, that discovery was missed by several physicists. For instance, Frederick Smith, an Oxford physicist, when told by an assistant that photographic plates kept near a Crookes tube were fogged, told his assistant to keep them somewhere else. (Whittaker, 1951: 358). [We may all smile on reading this, but can every one of us be quite certain that he or she is not now failing to recognize an anomaly in his or her research?]

Different anomalies evoke very different responses from the scientific community. I suggest that there are at least three different categories of anomalies: "OK Anomalies," "Not-OK Anomalies," and "Sleeping Anomalies."

An "OK Anomaly" is one that has been discovered by an established scientist, preferably using expensive equipment, and which appears to be an anomaly that scientists can cope with.

A "Not-OK Anomaly" is one that is not obviously resolvable and presents an unwelcome challenge to established scientists, possibly (but not necessarily) because it has been discovered by a non-scientist.

A "Sleeping Anomaly" is one that has not yet been recognized as an anomaly.

As examples of OK anomalies, I cite two from astronomy: (1) Quasars are objects that, when first identified by Maarten Schmidt of the Mount Wilson and Palomar Observatories, were anomalous in that they appeared to be star-like but had redshifts similar to—or larger than—those of typical galaxies (see, for instance, [Shu 1982: 315 et seq.]). Quasars have subsequently been determined...
to be distant galaxies containing a massive black hole. (2) Pulsars are radio sources that pulse with periods of seconds or less (see, for instance, Shu [1982: p. 131 et seq.]). When discovered in 1967 they were an anomaly since all previously known radio sources were essentially constant or varied only erratically on much longer timescales. Pulsars have subsequently been determined to be rotating neutron stars with very strong magnetic fields.

It is worth noting that claims of both of these astronomical discoveries were made by established astronomers using powerful optical or radio telescopes. The discovery of pulsars led to Nobel Prizes for Professor Anthony Hewish, who was in charge of the research project, and for Professor Martin Ryle, director of the observatory (the Mullard Radio Observatory at Cambridge, England). However, the initial discovery was actually made by Miss Jocelyn Bell, then a research student. It is also interesting to note that the first records of pulsars were kept secret, due initially to the possibility that they may have been emissions from intelligent life forms in other “solar” systems, but later to some other motive—possibly noble, possibly “Nobel.”

Both anomalies were viewed as due more to limitations in our astronomical knowledge than to errors in astronomical or physical theory.

A classical example of a Not-OK Anomaly is that of meteorites. These objects fall from the sky and may be discovered by any citizen, educated or not. Moreover, no specialized equipment is necessary. They are now known to enter the atmosphere from outer space, originating in a vast cloud of such objects in the solar system. However, their nature was unknown until the 18th Century, when E. F. F. Chladni published a small book on them in 1794. Twenty-two years earlier, in 1772, French academicians had ruled that these objects could not have fallen from the sky, since there are no stones in the sky to fall. According to Sears (1978), "The scientific community ... made merry over the credulity of people who imagined the stones to have fallen from the heavens." [What are the topics that are genuine, over which present-day scientists make merry?] The authenticity of meteorite falls was established by the distinguished scientist Jean-Batiste Biot, who was sent by the President of the National Institute to investigate a particularly large meteorite fall (over 3,000 stony meteorites) that occurred at L’Aigle on April 25, 1803.

A list of current Not-OK Anomalies contains topics that are generally dismissed as bogus by the scientific community: precognition, telepathy, psychokinesis, reincarnation, "flying saucers," etc., etc. The distinguished English astrophysicist Malcolm Longair (1984) warns young scientists that "it is difficult to be taken very seriously as a scientist if you mix up real science with quasi-scientific pursuits such as spoon-bending, parapsychology, unidentified flying objects, extrasensory perception, etc." However, the list also contains topics studied by scientists with a good track record of scientific research, such as the proposal by Halton Arp that the redshift of quasars may contain a contribution other than the usual cosmological redshift (see, for instance, Arp and Sulentic [1985]), and the proposal by Martin Fleischman that nuclear
processes may be influenced by electrochemical processes (Fleischman et al., 1989; see also Storms, 1996).

The close geometrical match between the west coast of Africa and the east coast of South America may be regarded as a "Sleeping Anomaly." Although this fact had been noted by Francis Bacon, Antonio Snider-Pellegrini, Benjamin Franklin and others, it was not generally recognized as a challenge to understanding until Alfred Wegener pointed out, early in the 20th Century, that geologic features of the West African Coast would accurately line up with similar features on the East Coast of South America when the two continents were juxtaposed and proposed an interpretation. Wegener attributed the correspondence to the breakup of one large continent (referred to as "Pangaea") and the progressive separation of the parts by a process he called "continental drift." This proposal was ridiculed for many years. The distinguished geophysicist Sir Harold Jeffreys once remarked to me—with a smile—that there is no force inside the Earth that is strong enough to move continents. Members of one scientific community (in this case, geophysicists) seldom welcome with applause a proposal made by a scientist from another community. (Wegener was a distinguished scientist, but he was a meteorologist, not a geophysicist.) The tide turned when geophysicists found that the magnetic signatures were effectively mirror-imaged on the two sides of the Mid-Atlantic Ridge, showing that it was the spreading center and providing a mechanism for what became the new theory of plate tectonics.

We now know that the scientific community was in error in its response to the challenge of meteorites and to that of continental configurations. Can we be sure that scientists of the 21st century are not making similar errors in their responses to some current phenomena? To pursue this question, we need to give a little thought to the nature of science. Richard Feynman (1956) remarked succinctly that "The essence of science is doubt." Three and a half centuries earlier, Francis Bacon (1603) had written "If a man will begin with certainties, he shall end with doubts; but if he will be content to begin with doubts, he shall end with certainties." These precepts are fully in accord with the rules of scientific inference, as developed by Jeffreys (1973), Good (1950), Jaynes (2004), and others. According to this theory, it is advisable to proceed along the following lines:

We should

1. Think in terms of probabilities, not certainties.
2. Consider a complete set of hypotheses, not a single hypothesis.
3. Examine our initial beliefs, and represent them by "prior probabilities."
4. List the relevant items of evidence, and estimate the credibility of each item.
5. In this way, estimate the "weight of evidence" that each item gives for each hypothesis.
6. Combine the "weights of evidence" with the prior probabilities to arrive at our post-probabilities.
We adopt a logically complete set of hypotheses to be sure that the anomaly can be compatible with one of our considered hypotheses. The epigram "Anything that does happen, can happen" is attributed to the distinguished astronomer and physicist Robert Leighton. Jeffreys (1961) wrote "It is sometimes considered a paradox that the answer depends not only on the observations but on the question; it should be a platitude."

In order to clarify this procedure, it is helpful to consider a simple example. The above procedure need not be restricted to scientific questions. We consider, as an example, the authorship of the "Shakespeare" plays. It is surely an anomaly that the plays show a detailed knowledge of Italy, yet Shakespeare never left England. We may analyze this anomaly as follows, using a procedure based on the principles of scientific inference that have been described elsewhere (Sturrock, 1973, 1994). First, we need a complete and mutually exclusive set of hypotheses. We adopt the following:

H1: The author of the Shakespeare plays was William Shakespeare;
H2: The author of the Shakespeare plays was Edward de Vere, Earl of Oxford;
H3: The author of the Shakespeare plays was somebody else.

We begin by giving these three hypotheses equal prior probabilities:

\[ P(H1|Z) = P(H2|Z) = P(H3|Z) = \frac{1}{3}, \]  

(1)

where \( Z \) denotes "zero-order" or background information. One can usually ignore the term \( Z \), unless one runs into difficulties (such as finding that none of the specified hypotheses is compatible with the evidence), in which case one needs to consider what is really being implied by this term.

We now need one or more "items" that can be part of an interface between data and theory. For present purposes, we adopt just one item, which comprises two exclusive statements:

F1: The author had first-hand knowledge of Italy;
F2: The author did not have first-hand knowledge of Italy

We need to assign probabilities to these statements based on the hypotheses, and based on the relevant evidence (the plays).

We know that Shakespeare had no first-hand knowledge of Italy and that de Vere did. Whether a hypothetical "somebody else" had knowledge of Italy is problematical. Ordinary actors and theater managers would not have had that knowledge. On the other hand, some noblemen and perhaps some merchants may have had extensive stays in Italy. Let us suppose that there is a one percent chance that the unknown author might have had first-hand knowledge of Italy, then

\[ P(F1|H1,Z) = 0, \quad P(F2|H1,Z) = 1; \]
\[ P(F1|H2,Z) = 1, \quad P(F2|H2,Z) = 0; \]
\[ P(F1|H3,Z) = 0.01, \quad P(F2|H3,Z) = 0.99. \]  

(2)
Finally, we need to assess these options on the evidence of the plays. Personally, I find it hard to believe that a playwright who had no first-hand knowledge of Italy would have had the knowledge and motivation to write in such detail about Italy, but I will allow that possibility a chance of one percent:

\[ P(F1|\text{plays}, Z) = 0.99, \quad P(F2|\text{plays}, Z) = 0.01. \] (3)

Then some formal manipulations (Sturrock, 1973, 1994; the relevant equations are reproduced in Appendix A) lead to the following post-probabilities that combine our thoughts about the hypotheses and about the relevant evidence:

\[ P(H1|F, Z) = 0.005, \]
\[ P(H2|F, Z) = 0.980, \]
\[ P(H3|F, Z) = 0.015. \] (4)

We see that this item of evidence strongly favors de Vere, and even favors "somebody else" over Shakespeare.

Of course, this is just one piece of evidence, and most people will start out with the presumption that the "Shakespeare" plays were in fact written by Shakespeare. Let us suppose that we start out feeling 99 percent confident that the author was indeed Shakespeare, but allowing 0.5 percent chance that it may have been de Vere and 0.5% chance that it may have been somebody else. Then, if (using the procedure given in the articles just cited) we fold this initial assessment together with the above assessment that was based on the probable familiarity with Italy, we arrive at

\[ P(H1|F, Z) = 0.074, \]
\[ P(H2|F, Z) = 0.803, \]
\[ P(H3|F, Z) = 0.123. \] (5)

We see that de Vere still comes out ahead, and Shakespeare still comes in last.

We are really interested in the application of these procedures to anomalies in scientific research. Hopefully (but not necessarily) these assessments can be somewhat more objective (or, as Ziman [1978] would say, "consensible") in the realm of science than in the realm of historical literary speculation. However, the merit of this procedure is not so much that it leads to definite answers, as that it will typically lead to definite questions.

I now wish to describe briefly three anomalies that have turned up in my own scientific research in recent years. One of these comes from "mainstream" science, and the other two are from topics that Longair (1984) warns young scientists not to get involved in.

In considering an anomalous experimental result or observation which appears to contravene current theory, we need to be able to estimate the probability that the result could have occurred "by chance" on the basis of that theory. One way to do this is to consider a wide range of similar results so that we can say...
"If this particular result occurred by chance, then many other similar results should also have occurred by chance" — or — "very few similar results would have occurred by chance." That is to say, it is helpful to have some way of "scanning" a wide range of possibilities, of which the result in question is simply a special case.

Indeed, an anomaly in scientific research is typically an unexpected result or observation that follows, or is accompanied by, many results or observations that occur as expected. For instance, in the Rutherford experiment mentioned earlier, for every alpha particle that was backscattered by the gold foil, many more were only slightly deflected. Sometimes the anomalies are associated with particular values of some parameter. In this case, it is obviously helpful to "scan" the result as a function of that parameter.

One scanning procedure that is often helpful is known as "power spectrum analysis." (See, for instance, Jenkins & Watts [1968]). One searches for periodic modulations of a measurement as a function of frequency. If one were to record the sound of middle-C on a piano and then carry out a power-spectrum analysis, one would find a peak in the display corresponding to a frequency of 262 cycles per second, as in Figure 1. There would also be peaks corresponding to "harmonics" of this frequency, at 524 cycles per second, 786 cycles per second, etc. Of course, these patterns are not anomalies; they are expected. However,
one might find one piano in a thousand for which the power spectrum also shows a peak at 120 cycles per second, as in Figure 2. This would be regarded as an anomaly, until one found that the piano contained a piece of electrical equipment, when it would no longer be an anomaly. (The sound produced by 60 cycles per second electrical power is predominantly at 120 cycles per second.) Now suppose that the recording is not that of a musical instrument but the noise of a large room full of chattering people. Then one is likely to obtain a very ragged power spectrum, as in Figure 3. In this case, it would be an anomaly to find a sharp peak, as in Figure 4. On investigation, one might find that a security alarm had been triggered somewhere in the building. Once that was discovered, the peak would no longer be an anomaly.

This brings me to some of my recent research. I have been studying measurements of the solar neutrino flux. The process producing neutrinos (the thermonuclear conversion of hydrogen into helium, etc.) is generally believed to be constant, and the experimental teams analyze their data on the assumption that the flux is in fact constant. The only expected periodic modulation would be a small variation, with a period of 1 year, due to the fact that the Sun-Earth distance varies in the course of a year. This modulation has been detected.

However, I have been interested in the possibility that the physical pro-
cesses generating the solar neutrino flux may not be spherically symmetric. In this case, one might find fluctuations corresponding to the frequency of solar rotation as seen from Earth—about 13.5 cycles per year, corresponding to a period of 27 days. Many other forms of radiation from the Sun (X-rays, etc.) do vary in this way, since they are influenced by the Sun's magnetic field, which typically has a very complex structure. According to some theories (Chauhan & Pulido, 2005), neutrinos might be influenced by a magnetic field, in which case the measured neutrino flux might be found to vary with a period of about 27 days.

The Super-Kamiokande collaboration has made available an extensive compilation of solar neutrino measurements (Fukuda et al., 2003). My colleagues and I have carried out several power-spectrum analyses of this dataset, the most recent of which (Sturrock & Scargle, 2006) is shown in Figure 5. This does not show a peak at the solar rotation frequency, but neither does the power spectrum of the disk-center magnetic field. One of the main features in the magnetic-field power spectrum is a peak at the second-harmonic of the rotation frequency (three times the rotation frequency) at $39.60 \pm 0.42 \text{ yr}^{-1}$. The power spectrum of the solar neutrino data shows a peak in this frequency band, at $39.28 \text{ yr}^{-1}$. It shows a stronger peak at $9.43 \text{ yr}^{-1}$ which is due, we believe, to a mode of

Fig. 3. The power spectrum that might be found from a recording made in a noisy room.
internal oscillation of the Sun. These features in the power spectrum of solar neutrino data represent an anomaly since, on the basis of standard neutrino theory, the flux should be constant and the power spectrum featureless.

The next example is closer to the interests of this society. I have carried out an analysis of a catalog of 12,100 UFO reports taken from a catalog compiled by Larry Hatch (Available at: http://www.larryhatch.net; Sturrock, 2003). Figure 6 shows a power spectrum of the events. We see that there is a prominent peak at 1 yr\(^{-1}\), which is not unexpected, since we spend more time outdoors in summer than in winter. Hence this peak does not tell us anything new: it is certainly not "anomalous." However, we can carry out an analysis that is a little more complicated, which searches for evidence of a rotating pattern of modulation. A modulation associated with the location of the stars will show up as a peak with frequency 1 yr\(^{-1}\). A rotation with the same frequency, but in the opposite direction, would show up as a peak with frequency \(-1\) yr\(^{-1}\). The result of this "running-wave" analysis is shown in Figure 7. We see that there are exceedingly strong peaks for forward waves with frequencies 1 yr\(^{-1}\) and at 2 yr\(^{-1}\), and only weaker peaks for reverse waves at those frequencies. This result provides very strong evidence for what is called a "local sidereal time" effect: the probability of a UFO event is related to which stars are overhead at the time of the event.
Unless one can show that most UFO events are due to misperceptions of certain astronomical objects, in a restricted range of local sidereal time, this comprises an anomaly.

The third example is taken from current research recently carried out in collaboration with James Spottiswoode (Sturrock & Spottiswoode, in press). We have applied the two procedures used in the two previous examples to a catalog of 3,325 free-response anomalous cognition experiments. The results of the simple power-spectrum analysis are shown in Figure 8. The strongest feature in this power spectrum occurs at $v = 24.65 \text{ yr}^{-1}$, quite close to twice the synodic lunar frequency (24.74 yr$^{-1}$). When the data are analyzed in terms of rotating frames, as in our UFO analysis, we obtain the result shown in Figure 9. In this case, the reverse-wave peak is stronger than the forward-wave peak, but this is consistent with an association of the results of the experiments with the position of the moon. Hence this analysis provides quite strong evidence for an anomaly—a lunar effect on anomalous cognition experiments.

There is of course a vast literature of studies of PSI and UFO data, and of many other similarly curious "anomalous phenomena." (See, for instance, The Sourcebook Project, available at: http://www.science-frontiers.com/sourcebk.htm). Paul Kurtz (1983; a philosopher) refers to a wide range of such phenomena as
"paranormal." He writes "[The term] ‘paranormal’ . . . is applied to anomalous data that supposedly transcend the limits of existing science and are due to unknown or hidden causes. The paranormal world view . . . contravenes the model of the universe derived from the physical and behavioral sciences." Kurtz’s approach is proto-typical of the self-styled "skeptical" community, which I prefer to refer to as the "pseudo-skeptical" community.

However, practicing scientists do not regard our current scientific knowledge as absolute and immutable. Sagan (1973) wrote "I would like to return to the question of possible new or alternative laws of physics. [Maybe] there are new laws of nature to be found even under familiar circumstances. I think it is a kind of intellectual chauvinism to assume that all the laws of physics have been discovered by the year of our meeting." The Russian physicist Vladimir Ginzburg (1973) wrote "Science of course never ends. There will always be new laws and clarifications. When we say some law of physics is valid, we always bear in mind that it is true within certain limits of applicability." And Edgar Mitchell (1993) wrote "There are no unnatural or supernatural phenomena, only very large gaps in our knowledge of what is natural . . . We should strive to fill those gaps of ignorance."

Hence, a major challenge in the study of anomalous phenomena is to identify

![Power spectrum formed from a catalog of 12,100 UFO events.](image)

Fig. 6. Power spectrum formed from a catalog of 12,100 UFO events.
the basic assumptions of our current "weltanschauung," "world view," or "model of reality" with which these phenomena are incompatible. This important question could and should be the topic of a major research project. In the present discourse, I look only for a very simple model.

It is my impression that the following three hypotheses form the basis for the usual rejection of evidence for such phenomena:

- Any topic which is incompatible with physical theory, as it is now known, is impossible.
- Consciousness is simply a brain activity.
- No "superior beings" have any influence, or have had any influence, on events and developments on Earth.

I suggest that, if we wish to study such phenomena, we should consider not only these three assumptions, but also the possibility that one or more of these assumptions may be incorrect.

To formalize this procedure, we may introduce the following three pairs of hypotheses:

Ordinary Physics (OP). The world is governed by (and restricted by) laws of physics as they are now known.
Extraordinary Physics (EP). The world is also subject to laws of physics of which we now have no knowledge and which make possible phenomena that are now inconceivable.

Ordinary Consciousness (OC). Consciousness is a brain activity and is therefore localized in time and space.

Extraordinary Consciousness (EC). Consciousness has an existence independent of the brain and is not limited in either time or space.

No Intelligent Intervention (NII). There is not now, and never has been, any intervention by non-human intelligent beings in events and developments on Earth.

Intelligent Intervention (II). There is or has been intervention by non-human intelligent beings in events and developments on Earth.

In terms of this set of options, the "Standard Model of Reality" comprises OP, OC, and NII. In almost all scientific research, the standard model of reality is built into the zero-order information \( Z \). In the current study, it is likely that there are other assumptions built into \( Z \), which are unrecognized and therefore unquestioned. For instance, there may be phenomena which are real, but which cannot be verbalized, for which it would therefore be difficult to enunciate the underlying hypotheses.
Now that we have identified what we regard as the "standard model," we can immediately list seven non-standard models of reality: EP, OC, NII; OP, EC, NII; OP, OC, II; EP, EC, NII; EP, OC, II; OP, EC, II; and EP, EC, II. It is convenient to refer to these as "Model of Reality Version 000," etc., or, briefly, "MOR000," etc. Then the set of models becomes what is outlined in Table 1.

**TABLE 1**
Models of Reality

| MOR000 = {OP, OC, NII} |
| MOR100 = {EP, OC, NII} |
| MOR010 = {OP, EC, NII} |
| MOR001 = {OP, OC, II} |
| MOR110 = {EP, EC, NII} |
| MOR101 = {EP, OC, II} |
| MOR011 = {OP, EC, II} |
| MOR111 = {EP, EC, II} |

Note: MOR = Model of Reality; OP = Ordinary Physics; OC = Ordinary Consciousness; NII = No Intelligent Intervention; EP = Extraordinary Physics; EC = Extraordinary Consciousness; II = Intelligent Intervention.
If we wish to study anomalous phenomena according to the principles of scientific inference, we should consider all eight of these possible models of reality, not just the standard model.

An important question that now arises is whether we should regard these three choices as independent, or whether the probability of one choice is likely to depend on one or two of the other choices. My own view is that the choice \( \text{OP}/\text{EP} \) will have an important influence on the other two choices. If \"extraordinary consciousness\" is real, it probably cannot be understood in terms of ordinary physics, so the prior probability that we assign to \( \text{OC} \) or \( \text{EC} \) will depend on whether we are associating it with \( \text{OP} \) or \( \text{EP} \). Similarly, one possibility for intelligent intervention is that beings from another \"solar system\" are visiting or have visited Earth. Travel from other stars seems virtually impossible if we think in terms of ordinary physics, but—for all we know—it may be comparatively easy in terms of some form of extraordinary physics.

If we regard the \( \text{OC}/\text{EC} \) choice and the \( \text{NII}/\text{II} \) choice as independent of each other, then we can proceed to organize the eight prior probabilities as follows: We first assign prior probabilities to \( \text{OP} \) and \( \text{EP} \): \( P(\text{OP}|Z) \) and \( P(\text{EP}|Z) \).

Note that, in setting these prior probabilities, we should ignore all the experimental and observational results that support \( \text{OP} \); since \( \text{EP} \) must contain \( \text{OP} \) as a special case, it follows that any result that is consistent with \( \text{OP} \) will also be consistent with \( \text{EP} \).

We next consider the choice \( \text{OC}/\text{EC} \), but relate the prior probabilities to our choice of \( \text{OP} \) or \( \text{EP} \): \( P(\text{OC}|\text{OP}, Z), P(\text{OC}|\text{EP}, Z), P(\text{EC}|\text{OP}, Z), P(\text{EC}|\text{EP}, Z) \).

Based on \( \text{OP} \), the probability of \( \text{OC} \) will be high, and that of \( \text{EC} \) will be small. Based on \( \text{EP} \), the probabilities of \( \text{OC} \) and \( \text{EC} \) may be comparable. Similar considerations apply to the choice \( \text{NII}/\text{II} \).

Based on these assumptions, the prior probabilities for the eight possible models of reality may be listed as follows:

\[
P(\text{MOR000}) = P(\text{OC}, \text{NII}) = P(\text{OC}|\text{OP}, Z) \times P(\text{NII}|\text{OP}, Z) \times P(\text{OP}|Z),
\]

\[
P(\text{MOR100}) = P(\text{EC}, \text{NII}) = P(\text{EC}|\text{EP}, Z) \times P(\text{NII}|\text{EP}, Z) \times P(\text{EP}|Z),
\]

etc.

The prior probabilities of our eight possible models of reality are formed from combinations of the following ten probabilities: \( P(\text{OP}|Z) \) and \( P(\text{EP}|Z) \); \( P(\text{OC}|\text{OP}, Z), P(\text{OC}|\text{EP}, Z), P(\text{EC}|\text{OP}, Z), \) and \( P(\text{EC}|\text{EP}, Z) \); and \( P(\text{NII}|\text{OP}, Z), P(\text{NII}|\text{EP}, Z), P(\text{II}|\text{OP}, Z), \) and \( P(\text{II}|\text{EP}, Z) \).

However, we are most interested in estimates of \( P(\text{EP}|Z) \) and of the post probabilities for \( \text{EC} \) and \( \text{II} \), which are given by

\[
P(\text{EC}|\text{Post}, Z) = P(\text{EC}|\text{OP}, Z) \times P(\text{OP}|Z) + P(\text{EC}|\text{EP}, Z) \times P(\text{EP}|Z),
\]
and

\[ P(\text{II}|\text{Post}, \text{Z}) = P(\text{II}|\text{OP}, \text{Z}) \times P(\text{OP}|\text{Z}) + P(\text{II}|\text{EP}, \text{Z}) \times P(\text{EP}|\text{Z}). \]  

(9)

In considering our assessment of EP, we should bear in mind that our basic laws of motion and gravity are only 300 years old and that relativity and quantum mechanics are only 100 years old. What is the probability that we have already discovered virtually all of physics? What is the probability that, even if we continue research for the next million years, there will be no further developments as revolutionary as relativity and quantum mechanics? It would be hard to justify a very small value for \( P(\text{EP}|\text{Z}) \). Indeed, it would not be unreasonable to adopt a value larger than 0.5.

On the other hand, most scientists are probably of the opinion that \( P(\text{EC}|\text{OP}, \text{Z}) \) and \( P(\text{II}|\text{OP}, \text{Z}) \) are small. Hence, the above equations may probably be approximated as

\[ P(\text{EC}|\text{Post}, \text{Z}) = P(\text{EC}|\text{EP}, \text{Z}) \times P(\text{EP}|\text{Z}), \]

(10)

and

\[ P(\text{II}|\text{Post}, \text{Z}) = P(\text{II}|\text{EP}, \text{Z}) \times P(\text{EP}|\text{Z}). \]

(11)

Appendix B lists the prior probability and the two conditional probabilities that need to be assigned in order to arrive at estimates of the probability of the most interesting non-standard models of reality.

The first terms on the right-hand side of these equations represent assessments of very speculative possibilities on the basis of unknown physics. To give these quantities very small or very large values would be an act of faith. It appears that, on the basis of our present knowledge (and ignorance), we cannot assert that extraordinary consciousness and intelligent intervention are either very likely or very unlikely.

The key assessment is the prior probability for EP. If this is considered to be very small, then all the models of reality will be unlikely, except the standard model. However, if the prior probability for EP is thought to be non-negligible, then (since EP is beyond our present comprehension) assessments of the prior probabilities for the four models that involve EP are likely to be non-negligible.

In order to obtain an informed range of estimates of \( P(\text{OP}|\text{Z}) \) and \( P(\text{EP}|\text{Z}) \), it would perhaps be reasonable to consult a number of theoretical physicists, but it is not at all obvious to which intellectual communities one should turn for estimates of \( P(\text{OC}|\text{OP}, \text{Z}) \), \( P(\text{OC}|\text{EP}, \text{Z}) \), \( P(\text{EC}|\text{OP}, \text{Z}) \), and \( P(\text{EC}|\text{EP}, \text{Z}) \) or of \( P(\text{II}|\text{OP}, \text{Z}) \), \( P(\text{II}|\text{EP}, \text{Z}) \), \( P(\text{II}|\text{OP}, \text{Z}) \), and \( P(\text{II}|\text{EP}, \text{Z}) \). However, the assessment of the prior probabilities for the eight possible models of reality is not essential for progress to be made. More important is the change in our assessments of the probabilities of these models when we examine the relevant evidence, since scientists should be able to agree on the weight of evidence, even if they differ widely in their prior probabilities. The crucial point is that we will
no longer refuse to examine a phenomenon because it appears to contravene the standard model. We will simply estimate the "weight" which each piece of evidence contributes to each of the eight possible models (and any other models that might be proposed) and then keep track of the accumulated weight of evidence for each model of reality.

Note that this approach leads to a different interpretation of the terms "paranormal" and "super-natural." These terms can now simply be interpreted as indicating that certain phenomena are (or appear to be) incompatible with the standard model of reality. It is then an open question, to be investigated, whether these phenomena are compatible with a non-standard model of reality. This interpretation can therefore lead to productive scientific research, whereas the standard approach of the pseudo-skeptical community leads to very little research.

Acknowledgments

This research has profited from conversations with many colleagues, but I wish to acknowledge my special debts to Henry Bauer, John Derr, Federico Faggin, Bernie Haisch, Bob Jahn, Jeff Scargle, Jacques Vallee, and Ron Westrum, and to the late Ed Jaynes, the late Marcello Truzzi, and the late Ian Stevenson.

References

The Role of Anomalies in Scientific Research 259


APPENDIX A

This appendix reproduces equations derived elsewhere (Sturrock, 1973, 1994) and which are used in this essay.

We consider a complete set of hypotheses \( H_i \), \( i = 1, \ldots, I \) and assign them prior probabilities \( P(H_i|Z) \), \ldots, \( P(H_I|Z) \), where \( Z \) indicates "Zero-order" or background information. For each item of evidence \( E \) we introduce a set of statements \( S_n \), \( n = 1, \ldots, N \) and then estimate the probability that each statement follows from each hypothesis, \( P(S_n|H_i, Z) \), and from the evidence \( E \), \( P(S_n|E, Z) \). Then the posterior probabilities are given by

\[
P(H_i|E, Z) = \frac{\sum_n P(S_n|H_i, Z)P(S_n|E, Z)}{\sum_k P(S_n|H_k, Z)P(H_k|Z)} \cdot P(H_i|Z). \quad (A.1)
\]

If we need to combine results from more than one item of evidence, say \( E_1, \ldots, E_A \), the result is given by

\[
P(H_i|E_1, \ldots, E_A, Z) = \frac{P(H_i|E_1, Z) \ldots P(H_i|E_A, Z) [P(H_i|Z)]^{-(A-1)}}{\sum_i P(H_i|E_1, Z) \ldots P(H_i|E_A, Z) [P(H_i|Z)]^{-(A-1)}}. \quad (A.2)
\]
APPENDIX B

This appendix lists the principal prior probabilities that one needs to estimate in order to assign prior probabilities to the most interesting models of reality specified in this essay.

The first estimate one needs to specify is the probability that there is extraordinary physics still to be discovered:

\[ P(EP|Z) = \text{______} , \]

where \( EP \) indicates Extraordinary Physics and \( Z \) indicates "Zero-order" or background information. (Remember that each probability estimate must be larger than zero and less than unity.)

Then, if we ignore the probability that Extraordinary Consciousness (EC) and Intelligent Intervention (II) are compatible with Ordinary Physics (OP), the important estimates to make are

\[ P(EC|EP,Z) = \text{______} \]

and

\[ P(II|EP,Z) = \text{______} \]

Then the post-probabilities, related to the most interesting alternative models of reality, are given to good approximation by

\[ P(EP,EC,NII|Z) \approx P(EC|EP,Z)P(EP|Z) = \text{______} \] \hspace{1cm} (B.1)

\[ P(EP,OC,II|Z) \approx P(II|EP,Z)P(EP|Z) = \text{______} \] \hspace{1cm} (B.2)

\[ P(EP,EC,II|Z) \approx P(EC|EP,Z)P(II|EP,Z)P(EP|Z) = \text{______} \] \hspace{1cm} (B.3)

where \( NII \) indicates No Intelligent Intervention and \( OC \) indicates Ordinary Consciousness.
The Yantra Experiment

Y. H. DOBYSN, J. C. VALENTINO, B. J. DUNNE, AND R. G. JAHN

Princeton Engineering Anomalies Research Laboratory
School of Engineering and Applied Science
Princeton University, Princeton NJ 08544-5263
e-mail: rgjahn@princeton.edu

Abstract — Qualitative and analytical observations of consciousness-related anomalies in random event generator (REG)-based experiments suggest that direct conscious feedback regarding experimental performance may impede rather than facilitate anomalous effects. The Yantra experiment tests this hypothesis by providing no outcome-related feedback to the operator. Feedback is replaced by a visual and auditory environment expected to be conducive to anomalous performance. This environment allows a number of options which operators can adjust to suit their personal taste, or to explore alternative conditions. The lack of feedback intrinsic to the program is reinforced by an experimental policy that forbids an operator to receive feedback before completing 10 experimental sessions or declaring an inability to return for further data collection.

Data analysis assumes that individual operators perform idiosyncratically; that populations distinguished by gender and previous experimental experience may perform differently; and that operator performance may depend on the environmental parameters of the protocol. All of these dependencies are found to exist. The most general test for distinctive individual behavior, a $\chi^2$ constructed from the Z-scores for each segment in which intention, operator, and environment are held constant, produces $\chi^2 = 629.05$ on 558 degrees of freedom, $p = 0.020$. The effect appears to be asymmetric and driven by changes in the high intention data alone. Gender differences in differential success rates are comparable to those seen in earlier experiments and are statistically significant ($Z = 2.213$). Analysis of subgroups distinguished by both gender and previous experience shows that previously experienced female operators produce individually consistent performances regardless of the imposed environment (although variable between individuals), while all other operator subpopulations show strong sensitivity to environmental conditions. Overall, the effect size, as measured by local mean shifts, is approximately four to five times that seen in earlier REG experiments, suggesting that similar no-feedback, environmentally supportive protocols may be fruitful for future research.

Keywords: human-machine anomalies — consciousness-related anomalies — PEAR—REG — psychological correlates — subjectivity — individual variations — feedback

I. Introduction

The Princeton Engineering Anomalies Research (PEAR) program has studied the effect of human intention on microelectronic random event generators (REGs) in experiments dating back to 1979 (Jahn & Dunne, 2005; Jahn et al., 1987, 1997,
Various modes of performance-related feedback have been used over that time. In the original experiment, feedback was automatic unless the operator went to some effort to avoid it, since a large and conspicuous front panel on the REG device displayed both the current trial value and a running average for the current collection of trials. Moreover, since a final run mean was displayed for the operator to record in a logbook, the "no-feedback" condition was maintained only for the duration of the current trial sequence. Subsequent remote experiments with the same equipment were run in a genuine no-feedback condition. Alternative modes of graphical feedback were introduced in the late 1980s, and proved popular with operators.

The initial introduction of graphical feedback seemed not to have significant consequences for the effect size, except for some operators on an individual basis (Nelson et al., 2000). Later experiments, however, suggested that this might not be a universal generalization. An experiment designed specifically for its appealing feedback produced no significant results by overall outcome measures (Jahn et al., 2000b), while in the extensive replication effort of the IGPP consortium, graphical feedback (chosen as the default mode) actually seemed counterproductive, with two of the three participating laboratories reporting statistically significant differences of performance in which graphical feedback proved inferior to other feedback modes (Jahn et al., 2000a, table M.2). In addition, anecdotal reports indicated that at least some operators found outcome-related feedback, with its implications of evaluation and judgment, to be objectionable and preferred to work without feedback of any kind.

These considerations led to the design of an experiment that would provide no feedback regarding experimental outcomes. This design was facilitated by the availability of a new generation of REG sources without front panel displays. With the computer screen relieved of the necessity for a feedback display, it was decided to use the screen to present an image that it was hoped would be conducive to anomalous performance. The specific choice of image was motivated by the experience of the "ArtREG" experiment (Jahn et al., 2000b). In that experiment, operators were presented with two superimposed pictures, initially in a "double exposure," with half the pixels on the screen coming from each picture. The balance between the two images varied under the control of an REG input, and the operator's intentional task was to make the chosen target image dominate the screen. While the results of the experiment as a whole were non-significant, there seemed to be a substantial effect size associated with a subset of the images. These images were deemed "numinous," containing significant religious or spiritual imagery from a number of different traditions. After some deliberation it was decided to use a mandala design known as the "Sri Yantra" (see Figure 1) as a numinous visual display to accompany the new experiment.

2. The Yantra Environments

The environmental parameters presented by the Yantra experiment include options for both visual and auditory components intended to facilitate
a meditative state of mind and suppress analytical focus. The Sri Yantra is a pattern of interlocking triangles at the core of Figure 1, a symbol which is supposed to represent the interpenetration of spirit and the material world. The remainder of the design consists of a series of traditional framing elements commonly used to surround the Sri Yantra, which also appear frequently in other mandala designs.

Operators have three choices of visual environment. The Sri Yantra mandala can be presented as shown in Figure 1, as a static picture on the computer monitor (in white lines on a blue background screen). Alternatively, sectors defined by the various radial boundaries (the surrounding box, the internal circles bounding the "lotus blossom" patterns, and the Sri Yantra itself) can be presented in differing background colors, with the colormap changing in a steady rhythm driven by arrival of REG trials at the computer. (The values of the trials have no effect on this; only their reception by the computer is relevant.) The pattern of specific color changes is chosen by a pseudo-random process unconnected to the experimental data. As a third alternative the monitor can simply be left blank.

Similarly, operators are offered several options for audio environment. By means of a servomotor controlled from an output port and connected to a drumstick, the computer can beat a large Native American drum in the experiment room. The default audio operation is for the drum to beat once with
each data reception event (that is, in the same rhythm as the changes in the video if changing video is in use). An alternative rhythm beats the drum twice, quickly, with each trial, producing a pattern of quick double beats separated by slightly less than a second, strongly reminiscent of a heartbeat. A third option is silence, and a fourth allows operators to bring their own music CDs or other recording media to play any soundtrack that appeals to them while doing an experiment.

In addition to these various environmental options, another experimental parameter carried as a variable is the instructed versus volitional assignment of intention deployed in most of PEAR’s REG-based experiments. There are thus twenty-four possible combinations of intentional assignment, visual environment, and audio environment. These are chosen freely according to the operator’s preferences, although operators who explore more than one environment are encouraged to generate substantial databases in each.

3. Experimental Protocol

In Yantra, as in most PEAR REG experiments, the primary variable is operator intention: operators actively attempt to shift the REG output distribution in the high and low directions, in a balanced design. The basic unit of data collection is a trial of 200 random bits, summed to produce a random integer with theoretical mean 100 and standard deviation \( \sqrt{50} \approx 7.071 \). Trials are produced at a rate slightly faster than 1 per second. Sequences of 100 trials are generated automatically as runs. The basic unit of operator participation is a series in which an operator completes two runs in the high intention and two runs in the low intention. This requires approximately 10 minutes in a typical case. Unlike the standard REG protocol, Yantra is bipolar rather than tripolar, with no baseline intention. The assignment of intentions to runs may be made by the operator, or determined by the computer. In the latter case the determination is made by a pseudo-random process seeded by the time at which the program is started. For both volitional and instructed data, the program enforces the constraint that a series contains exactly two runs of each intention.

Since series are quite short, many operators chose to generate multiple series in one session. While operators could, in principle, generate as many or as few series as they cared to, the experimental protocol provides no feedback on their performance until they either (a) complete at least 10 series, or (b) declare that they will not generate any further Yantra data. This policy has the beneficial side effect of assuring that small databases from short-term operators could not be subject to optional stopping, since operators had no information about the outcome of their efforts. While operators could, if they wished, receive feedback after their 10th series, several of those who continued to larger databases chose not to be given feedback until they had completed their entire Yantra involvement.

When the Yantra experiment was launched it was decided that it would be closed after 1000 series had been generated. Practical considerations having to do with the availability and enthusiasm of operators, and the desirability of large
operator databases, led to a slight relaxation of this condition, to the stipulation that the experiment would run at least 1000 series and that after the 1000-series mark the experiment would be kept open only for the benefit of operators who were attempting to complete previously declared commitments regarding personal database size. Once these outstanding commitments were completed Yantra had generated a total of 1017 formal series. Space precludes the presentation of the raw data in the current article, but they can be found in the Appendix to the Technical Note on the Yantra experiment (Dobyns et al., 2006).

4. Data Analysis Methods

Yantra analysis was designed from the outset under the assumption that operators would produce individual and idiosyncratic results. Of course, individual Z-scores for operators have always been computed in PEAR experiments; individual variability becomes relevant only when constructing an overall "bottom-line" evaluation for the population of operators. The standard pooled, weighted Z-score test used in earlier experiments is clearly not acceptable under this hypothesis. It is tantamount to assuming that all operators are interchangeable. While it is the most sensitive possible test for detecting a consistent universal effect, individual variations are averaged out and become invisible.

Given the hypothesized situation of effect sizes that will vary among individuals in an unpredictable manner, there is no one statistical test that is optimally sensitive for all conditions; sensitivity depends on the model of variation. A test that is very broadly useful, however, is a \( \chi^2 \) test based on the 2-scores of components. This is computed by simply squaring the Z-scores of all component databases and summing the squares; the number of degrees of freedom (d.f.) of the \( \chi^2 \) is equal to the number of components. Two features of this test make it particularly useful and versatile. First, \( \chi^2 \) values follow an addition rule: the sum of two \( \chi^2 \) values is another \( \chi^2 \) value with a number of d.f. equal to the sum of the d.f. in the two contributions. Second, if the composite Z mentioned above is squared and subtracted from the overall \( \chi^2 \), the result is again \( \chi^2 \) distributed* with one fewer d.f. This secondary \( \chi^2 \) is driven solely by the variation between subsets, the mean effect having been removed by the Z-score subtraction. To express these three quantities mathematically, if there are a total of \( N \) subsets, with the ith subset comprising \( n_i \) data units and having an aggregate Z-score of \( Z_i \), the composite \( Z \), raw \( \chi^2 \), and variability \( \chi^2 \) can be written:

\[
\begin{align*}
Z_c &= \frac{\sum_{i=1}^{N} Z_i \sqrt{n_i}}{\sqrt{\sum_{i=1}^{N} n_i}}; \\
\chi_r^2 &= \sum_{i=1}^{N} Z_i^2 (N \text{ d.f.}); \\
\chi_v^2 &= \chi_r^2 - Z_c^2 (N - 1 \text{ d.f.}).
\end{align*}
\]  

* This is not a general subtraction property for \( \chi^2 \); the difference of two \( \chi^2 \) is not in general \( \chi^2 \) distributed. It can be shown, however, that in this specific case, the residual, after subtracting the mean \( Z^2 \) from a \( \chi^2 \), is in fact \( \chi^2 \) distributed.
These equations provide the basic tools for most of the Yantra analytical treatments. In addition to inter-operator variability, previous experiments led to an expectation that operators might either individually or collectively vary in their responses to the 24 operating environments, and display distinct effects in high and low intentions. Moreover, it is expected from previous REG observations that if the operator pool is divided into subtypes by gender and previous experience, different patterns of performance appear in the subtypes. Analyses for all of these factors are obviously necessary for the interpretation of the experiment.

5. Results

The total database of 1017 series was contributed by 61 different operators. At least some exploration of each of the 24 possible environments was conducted. The extreme form of the idiosyncratic-effects hypothesis is that a different effect may be seen in any data subset generated by a different operator, in a different environment, in a different intentional effort. The set of all data generated by a single operator in a single intention and environment will be referred to hereafter as a segment. There are 558 such segments in the formal database, 279 in each intention. These segments have a raw \( \chi^2 = 629.04, p = 0.020 \). Figure 2 illustrates the outcome in the closest possible analog of PEAR's traditional cumulative deviation, with the excess of \( \chi^2 \) over its theoretical expectation (i.e., the number of d.f.) plotted against the number of segments accumulated.
Figure 3 illustrates the result of separating the segments according to operator intention. The high segments have $\chi^2 = 357.36$ on 279 d.f., $p = 0.0010$. The low segments, in contrast, have $\chi^2 = 271.68$ ($p = 0.612$). The effect is thus driven by the large mean shifts observed in the high intention alone, a result similar to that seen in other non-feedback experiments (Dunne & Jahn, 1992). One may also construct the population of 2-scores for the intentional difference: $Z_{\Delta} = (Z_H - Z_L) / \sqrt{2}$, for each matched pair of segments (i.e., the segments run in the high and low intentions by a given operator in a given environment). Not surprisingly, this produces an intermediate result: $\chi^2_{\Delta} = 321.59$ on 279 d.f., $p = 0.040$.

These results are almost purely driven by inter-segment variation. The overall pooled $Z_c$ results are 0.0307 and -0.2070 in the high and low intentions, respectively: the pooled $Z_{\Delta} = 0.1681$. Subtracting out this average effect yields variability-driven $\chi^2_v$ (all with 278 d.f.) of 357.36 ($p = 0.00091$) in the high intention, 271.64 ($p = 0.596$) in the low, and 321.56 ($p = 0.037$) in the high – low difference.

5.1. Individual Operators and Operator Subtypes

The effects are less impressive when operators are considered singly, without regard to environmental differences. Figure 4 shows a scatterplot of the 61 operator performances in the two intentional conditions. These contributions produce overall $\chi^2$ values of 76.165 ($p = 0.091$) in the high, 68.150 ($p = 0.247$) in the low, and 56.060 ($p = 0.655$) in the delta condition.

There is, nevertheless, evidence of anomalous performance in the operator-
by-operator database as well. The largest 2-score attained by any operator (marked by a square in Figure 4) is $Z = 3.833 \ (p = 1.26 \times 10^{-4}$, two-tailed). After Bonferroni correction for having 122 such scores to examine, this remains a conventionally significant value of $p = 0.015$. Nor is this performance alone; datasets by three different operators show $|Z| > 3$, an overpopulation that is a $p = 0.0046$ event.

This would seem, at face value, to indicate that some operators produce consistent, individual effects, though the population as a whole does not. This can be clarified by examining Figure 5, which shows the operator-based $\chi^2$ values for each of the four subsets resulting when the operators are segregated according to their gender and previous experience with REG-type experiments. More specifically, for reader visual comparison this figure shows the ratio of $\chi^2$ to d.f., so that the horizontal line at 1 shows the chance expectation for each test. The plotted letters show the $\chi^2$/d.f. value for the high and low intentions. The dotted lines show the 95% confidence limits for the $\chi^2$; they are at different heights in the different subsets because $\chi^2$/d.f. has different quantiles for different d.f., even though its expectation is always 1. It is clear from Figure 5 that the experienced female operators have highly significant individual effects in both high and low intentions; all of the other operator subtypes show no such effects. It is worth noting that all three of the $|Z| > 3$ databases were produced by such previously experienced female operators. In contrast, Figure 6 shows the segment-based (or operator X environment) $\chi^2$ for these same operator subpopulations. Here we see the interesting outcome that the females with previous experimental experience have a non-significant result, while each of

---

Fig. 4. Scatterplot of all individual operator performances, not divided into environment segments.
Fig. 5. Operator-based $\chi^2$ values separated by operator gender and experience. The plotted letters show the ratio $\chi^2$/d.f. for each of the two intentions: the solid line at 1 is thus the theoretical expectation. The dotted lines show the $p = 0.05$ confidence limit; they are at different heights in different sections due to different numbers of d.f.

The other operator populations produces a $\chi^2$ in the high intention that exceeds the 95% confidence limit for chance variation. The implications may be clearer if the numeric data are presented in tabular form, as in Table 1.

The first row of each section of Table 1 gives the number of d.f. in the $\chi^2$ for that column. It should be noted that for the operator $\chi^2$, the d.f. do not add up to 61 because five of the "operators" in the full dataset are actually male-female co-operator pairs and cannot be assigned to a specific gender. Below the d.f. entry is the cutoff value for $p < 0.05$ significance in a $\chi^2$ with that number of d.f.

Of particular interest in Table 1 is the comparison between operator-only and operator $\times$ environment $\chi^2$ for the experienced female operators. These 20 operators produce an excess $\chi^2$ of 15.48 above expectation ($p = 0.018$) in the high intention, and 18.06 ($p = 0.0087$) in the low, when an operator-based $\chi^2$ is computed for each operator's total performance. When the data are further subdivided by environment, the number of d.f. increases from 20 to 89, while the $\chi^2$ values increase from 35.483 to 96.687 and from 38.064 to 106.526 in the high and low intentions, respectively. Put another way, the further subdivision of the data adds 69 d.f., while adding 61.204 and 68.462 to the two $\chi^2$ values. We thus see that the previously experienced female operators show strong evidence for an effect when their total databases are examined, but the subdivision by environments increases the $\chi^2$ only by amounts such as would be expected from the increase in d.f., that is, consistent with these operators displaying only
random variation between environments. We may thus conclude that they show characteristic personal effects which are unaffected by the operating environment.

In contrast, the other operator subgroups (females without prior experience, and both experienced and inexperienced males) show only the expected level of random variation in their overall personal performances, but they show variation far beyond chance levels when their data are subdivided by environment. (This is evident in the high intention, as is obvious from Figure 6; the pooled high data from Table 1 for these operators has $\chi^2 = 250.63$ on 181 d.f., $p = 0.00047$. The

---

### Table 1

Analysis by Operator Subtype

<table>
<thead>
<tr>
<th>Subtype</th>
<th>New Female</th>
<th>New Male</th>
<th>Exp Female</th>
<th>Exp Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operator d.f.</td>
<td>5</td>
<td>12</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>$p &lt; 0.05$ cutoff for this d.f.</td>
<td>11.07</td>
<td>21.03</td>
<td>31.41</td>
<td>30.14</td>
</tr>
<tr>
<td>Op $\chi^2$, HI</td>
<td>2.700</td>
<td>10.506</td>
<td>35.483</td>
<td>23.257</td>
</tr>
<tr>
<td>Op $\chi^2$, LO</td>
<td>2.964</td>
<td>6.874</td>
<td>38.064</td>
<td>14.279</td>
</tr>
<tr>
<td>Op $\chi^2$, $\Delta$</td>
<td>0.981</td>
<td>8.835</td>
<td>28.840</td>
<td>12.490</td>
</tr>
<tr>
<td>Op $\times$ env d.f.</td>
<td>20</td>
<td>57</td>
<td>89</td>
<td>104</td>
</tr>
<tr>
<td>$p &lt; 0.05$ cutoff</td>
<td>31.41</td>
<td>75.62</td>
<td>112.02</td>
<td>128.80</td>
</tr>
<tr>
<td>Op $\times$ env $\chi^2$, HI</td>
<td>33.146</td>
<td>76.912</td>
<td>96.687</td>
<td>140.572</td>
</tr>
<tr>
<td>Op $\times$ env $\chi^2$, LO</td>
<td>18.429</td>
<td>43.718</td>
<td>106.526</td>
<td>94.904</td>
</tr>
<tr>
<td>Op $\times$ env $\chi^2$, $\Delta$</td>
<td>28.698</td>
<td>60.665</td>
<td>98.281</td>
<td>125.239</td>
</tr>
</tbody>
</table>

Note: Exp = experienced; Op = operator; HI = high intention; LO = low intention; env = environment.
low data are at chance levels, but even pooled across both intentions the high results drive a marginally significant outcome: \( \chi^2 = 407.681 \) on 362 d.f., \( p = 0.049 \), for high and low intentions combined.) We may conclude from this that all operators, except experienced females, produce anomalous effects that are not only personally idiosyncratic, but also strongly influenced by the operating environment.

This analysis by gender does not include the co-operator subset, for which a meaningful assignment of gender cannot be made. While previous analyses have suggested that co-operators display interesting gender-like effects according to their status as same-sex or opposite-sex pairings (Dunne, 1991), the co-operator database in Yantra is too small and homogeneous to extract meaningful results from such a breakdown. There are five co-operators, all opposite-sex pairs, contributing operator-based \( \chi^2 \) values of \( \chi_h^2 = 4.217 \), \( \chi_f^2 = 5.974 \), and \( \chi_{\Delta}^2 = 4.914 \), none of which are significant. They contribute 9 of the 279 segments in each intention, for segment-based \( \chi^2 \) of \( \chi_h^2 = 10.045 \), \( \chi_f^2 = 8.104 \), and \( \chi_{\Delta}^2 = 8.703 \), all likewise nonsignificant.

5.2. Gender Analysis

While the above subdivision into types has been instructive, it differs from the gender analysis performed by Dunne (1998), which found striking gender-based differences in a much simpler statistic, namely, the rate of differential success by operator gender. That is, if one simply counts, for each gender, how many operators "succeed" in their intentional effort (have a higher mean in the high intention than in the low), one finds different success rates for male and female operators.

This effect has been exactly replicated in Yantra, as shown in Figure 7, where 20 of the 31 male operators succeed in the direction of intention, while only 9 of the 25 females do so. (For completeness, we may note that 4 of the 5 co-operators do so.) The difference is equally present in the data of experienced and new operators; only the smallness of the database prevents it from achieving statistical significance among the new operators. It would appear that despite the numerous distinctions between Yantra and other REG-type experiments, a basic gender-based difference in response remains pervasive.

5.3. Operating Conditions

The strongest effects in the Yantra database are the excess of variation seen in the high intention, when the data are split into segments according to both operator and environment, and the consistent personal performances of experienced female operators. All of this has been established from a viewpoint that individual operator performance is primary, and that operating conditions provide extra sources of variation within a particular operator's database. This is not, however, the only way the Yantra data segments can be organized. We may ask equally well whether there are characteristic patterns of operator performance
in particular operating environments, and how much inter-operator variation occurs within a fixed environment.

These questions can be answered directly by computing an overall composite $Z$ for all of the segments produced in a given operating environment. The sum of the squared 2-scores of all segments in the environment is the basic $\chi^2$ for that environment. As discussed in section 4, when the squared composite $Z_c$ is subtracted from this we are left with a $\chi^2$ showing the degree of inter-operator variability. It has, of course, one less d.f. than the number of segments in that environment.

Adding up the inter-operator variability $\chi^2$ for each of the environments yields a $\chi^2$ with $279 - 24 = 255$ d.f., driven by the amount of inter-operator variability that exists when environmental conditions are held constant. Similarly, the sum of all of the $Z_c^2$ for the 24 environments is a $\chi^2$ derived from any effects that are consistent within environments, although they may vary between environments. From the construction of these two values it is obvious that they must add up to the same total segment-based $\chi^2$ presented in earlier analyses, with the same total d.f. This is why the calculation is referred to as an alternative way of organizing the Yantra data. Instead of partitioning the list of segments by operators and then examining within-operator variability from environmental conditions, here we are partitioning the segments by environments and then examining within-environment variability from operators.

Figure 8 shows the results of this partitioning. Somewhat surprisingly, it indicates that both inter-operator variation and consistent mean shifts contribute
significantly. In addition to the expected inter-operator variability component, there is also a significant contribution from consistent performance across operators within each environment. Indeed, considered in terms of effect size (the proportional increase in the $\chi^2$ over its expectation), the latter is more than twice as large as the more highly significant effect of inter-operator variation.

Figure 9 plots the 24 operating environments individually against these two measures of anomalous effect. The three-letter codes indicate the three features of the environment: assignment of intention (volitional [V] or instructed [I]), type of visual display (changing [C], static [S], or none [N]), and type of audio environment (single beat [S], heartbeat [H], none [N], or other [O]). This plot shows only the high intention, since the low intention data are indistinguishable from chance in this representation. The vertical axis is $Z_e$ for that environment, the pooled $Z$-score for all data run under those environmental conditions. The horizontal axis is constructed by converting the inter-operator variability $\chi^2$ for that condition to its equivalent $Z$-score (specifically by applying the inverse normal distribution to the $p$-value calculated for the $\chi^2$). The dotted circle shows the 95% confidence bounds for the null hypothesis in such a plot; if the points are distributed according to two independent, normally distributed variables, 95% of them should fall within the circle. Five of the 24 points are clearly well outside this circle (the three-letter labels are centered over the exact points); in fact, a sixth (the VCH condition at upper left) also falls just outside the boundary. Thus, 6 of the 24 environments exceed the $p < 0.05$ criterion for their distribution along these two parameters of consistent internal effect and inter-operator variation; this overpopulation is itself a $p = 0.00096$ event by exact binomial calculation.
Fig. 9. The 24 conditions, plotted against their consistent effect (vertical axis) and inter-operator variation (horizontal axis). Dotted circle shows 0.95 confidence limits of null hypothesis; 95% of plotted points expected to fall within this circle. High intention data only.

This figure also provides potential insights into which environments actually are more conducive to producing anomalous yields, either in a global or an operator-specific mode. Of the six individually significant outliers, three used instructed assignment and three volitional, indicating no preference. All involved some form of audio stimulation—the no-audio environments are all well within the circle. Moreover, three of the six involve specifically the single-drumbeat audio. Since each of the four audio options appears in 6 of the 24 possible environments, this means that fully half of the single-drumbeat environments show individually significant anomalous performance.

To determine the effects of the particular environmental parameters individually on the anomalous yield, a segment-wise $\chi^2$ may be computed on those segments containing those parameters. Different parameter values can then be compared by an F-ratio test. Table 2 summarizes these results for the high intention only, again because only the high intention results display an overall anomalous effect.

While it appears that instructed assignment is driving the effect and the volitional condition contributes little, this assessment must be made with caution. The F-ratio test between these two $\chi^2$ values is 1.267 on 198 and 81 d.f., $p = 0.111$. There is thus a reasonable likelihood that the instructed and volitional databases are samples from the same underlying distribution, and the
lack of significance in the volitional segments is a combination of happenstance and smaller database size.

For the video environment, the face-value conclusion is that the changing video offers no anomalous yield, while the static video contains a strong effect. In contrast to the previous case, this is confirmed by an F-test between the two: F = 1.633 on 77 and 157 d.f., p = 0.0051. Even after a factor-of-three Bonferroni correction to allow for the fact that there are three ways to pick two comparison sets out of a group of three, this remains clearly significant at p = 0.015. The no-feedback condition is intermediate between the two in both effect size and significance, and F-tests confirm that it cannot be distinguished reliably from either.

For the audio environment, both of the drum-based environments show robustly significant effects. The "Other" environment, indicating an operator-provided audio background, appears to contain comparably strong effects, although its small size precludes statistical significance. In contrast, the no-audio condition is clearly null. Unfortunately, this distinction, while highly suggestive, may also be subject to overinterpretation, since the comparison of the no-audio condition with the pooled results of the active audio conditions still only achieves a marginal F-ratio of 1.358 on 199 and 80 d.f., p = 0.0585. The factor-of-four Bonferroni correction required reduces this almost-significant result to nonsignificance, indicating that although the anomalous yields appear to be present only when audio feedback is used, we cannot claim statistical confidence that this correlation is not coincidental.

As a final note on environmental effects, it is worth recalling that the environment of every experimental series is chosen by the operator to suit his or her current mood and preferences. Despite this, many of the environments seem to produce no anomalous yield. Statistical scrutiny of the environmental components confirms that the most popular choice of video display is associated

<table>
<thead>
<tr>
<th>Parameter Value</th>
<th>d.f.</th>
<th>$\chi^2$</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructed (I)</td>
<td>198</td>
<td>270.16</td>
<td>0.000495</td>
</tr>
<tr>
<td>Volitional (V)</td>
<td>81</td>
<td>87.20</td>
<td>0.299</td>
</tr>
<tr>
<td>Changing (C)</td>
<td>157</td>
<td>164.60</td>
<td>0.323</td>
</tr>
<tr>
<td>Static (S)</td>
<td>77</td>
<td>131.81</td>
<td>0.000102</td>
</tr>
<tr>
<td>None (N)</td>
<td>45</td>
<td>60.96</td>
<td>0.0565</td>
</tr>
<tr>
<td>Single beat (S)</td>
<td>100</td>
<td>140.95</td>
<td>0.00440</td>
</tr>
<tr>
<td>Heartbeat (H)</td>
<td>84</td>
<td>111.09</td>
<td>0.0256</td>
</tr>
<tr>
<td>Other (O)</td>
<td>15</td>
<td>23.69</td>
<td>0.0705</td>
</tr>
<tr>
<td>None (N)</td>
<td>80</td>
<td>81.63</td>
<td>0.428</td>
</tr>
</tbody>
</table>
with a null result that can validly be distinguished from that of the rest of the experiment. It thus would seem that the aesthetic preference for a particular environment is no guarantee of its facilitation of anomalous performance, even for the particular operator expressing the preference. We may observe that this is consistent with the outcome of the ArtREG experiment (Jahn et al., 2000b), wherein most operators reported that they found the experience enjoyable, but which nevertheless produced no overall anomalous yield.

5.4. Miscellaneous Observations

Three operators, all males lacking previous experience, performed a sub-experiment within the main Yantra experiment. These operators were all practitioners of a Japanese healing discipline known as Johrei. They intentionally employed Johrei techniques in exactly half their Yantra data, and refrained from using Johrei in the other half. These data have been reported in more detail elsewhere (Jahn et al., 2006). The results may be summarized by noting that these operators produced strong segment-based responses in their Johrei data and null results in their non-Johrei data. Since the Johrei condition was not part of the formal definition of Yantra segments, this distinction has been diluted in the current analysis. Taking Johrei use into account as a fourth "environmental" condition would slightly increase the statistical significance of the segment-based analysis for the overall data and for the inexperienced male operators, but it would produce no qualitative change in the conclusions drawn thus far.

The overall effect size in the Yantra experiment appears to be larger than that seen in the original REG studies. Since Yantra expects, and uses statistical tests for, idiosyncratic effects that vary in both size and direction between operators, direct comparisons with the overall average effect size seen in the original REG experiment are somewhat problematic. In terms of the size of the mean shift driving the anomalous effects, however, we may note that the $\chi^2$ for the pooled high and low data is 629.0425 on 558 segments, indicating a mean $Z^2$ on the segments of 1.1273. Since the mean length of segments is 729 trials, if the excess in $Z^2$ is (as hypothesized) driven by consistent mean shifts within segments, this mean effect amounts to a $Z$ of 0.01321 per trial, or a mean shift of 0.0932. Although this is an estimate resting on several assumptions about the nature of the effect, it may be compared to the observed mean shift of 0.0208 in the original REG experiment; the Yantra figure is approximately 4.5 times larger. It is notable that the only other fully non-feedback experiment in the REG repertoire, the remote database, shows an effect size that is indistinguishable from the original REG, although it displays the same high/low asymmetry as Yantra (Dunne & Jahn, 1992).

If anomalous performance in the high and low intentions were independent, or were present only in one intention, we would expect the $\chi^2$ in the $\Delta$ condition (the high minus low difference for a given segment) to be
intermediate between the $\chi^2$ for high and low. This is in fact observed for the operator $X$ environment tests. In contrast, if effects were symmetric (that is, an operator attained the same mean shift in the direction of intention regardless of the sign of the intention), the $\chi^2$ on the $A$ condition would be substantially larger than that in either intention. This is not seen in any of the Yantra analyses. Instead, the operator-based $\chi^2$ (without regard to environment) consistently shows a $\chi^2_A$ that is smaller than $\chi^2_h$ or $\chi^2_l$. This is especially pronounced for the experienced female operators, where both intentions have significant $\chi^2$ while $A$ does not. This odd behavior suggests that there may actually be a correlation between the two intentions at the level of operators' complete databases, or a tendency for operators to produce mean shifts in the same direction in both high and low intention, regardless of environment. Indeed, of the 61 operators, 37 have the same sign in their overall high and low databases, vs. 24 who produce opposite signs in high and low. The correlation coefficient between the high and low intentional results, operator-by-operator, is $\rho = 0.2201$, $p = 0.044$ (one-tailed).

6. Conclusions

The analysis of the primary intentional data in the Yantra experiment leads to the following conclusions:

1. As a whole, the operators have anomalously shifted the means of the intentional data, although the mean shift is asymmetrical between intentions and its direction varies unpredictably among operators and among environmental conditions.
2. Female operators who have previously participated in REG experiments show consistent individual anomalous performance in both high and low intentions, regardless of environment, although the performance still varies unpredictably among operators.
3. Female operators new to REG experimentation, and male operators in general, show strong sensitivity to environmental conditions, and collectively produce effects only in the high intention.
4. Despite the fact that operators choose environments that appeal to them, certain environments are apparently conducive to anomalous yield while others are not. This suggests that an environment's ability to foster anomalous effects may not correlate with its aesthetic appeal, as was noted in the ArtREG experiment.
5. The gender-based patterns of differential success seen in earlier experiments are replicated in Yantra, on very similar scales.
6. Examination of the individual components of the environments suggests that instructed assignment of intention is more conducive to anomalies than is volitional assignment, and that drumbeat accompaniment is more conducive than is silence. However, the statistical confidence of these conclusions is modest.
7. Examination of the video component of the environment, in contrast, shows that the static Sri Yantra mandala produces strong anomalous yields, while the changing mandala does not, a distinction that is statistically robust. The state involving no visual stimulus at all is intermediate between the two and cannot be resolved statistically from either.

8. The overall Yantra effect size can be estimated to be between four and five times the effect size seen in the original REG experiments. Given that the effect seems to be concentrated in certain conducive subsets, the actual increase in effect size in those cases may be even larger.

These observations provide valuable hypotheses for future research. For example, would experiments focusing on the conditions found to be conducive in Yantra in fact produce larger yields? Despite the fact that the experiment as a whole produced unpredictable anomalous mean shifts with considerable inter-operator variation, some environments showed consistent mean shifts in the direction of intention, while others showed consistent mean shifts contrary to that direction. Can such tendencies be used to foster more consistent intentional performance among operators? What are the implications, in this context, of the gender-related differences in differential intentional success? We may conclude that while it shows a resounding confirmation of the basic hypothesis that anomalous human-machine interactions may take place in the complete absence of feedback, and while it displays numerous intriguing structural features which hint at the nature of the anomalous effect, the Yantra experiment actually raises more questions than it answers.

Acknowledgments

The PEAR laboratory gratefully acknowledges the support of Sekai Kyusei Kyo, the Hygiea Foundation, the Institut für Grenzgebiete der Psychologie und Psychohygiene, and numerous private philanthropists. PEAR also expresses its gratitude to the many uncompensated volunteer operators without whom these data could not have been collected.

References


Jahn, R., Dunne, B., Bradish, G., Dobyns, Y., Lettieri, A., Nelson, R., Mischo, J., Boller, E., Bosch,


Suzel Fuzeau-Braesch
Université de Paris, Orsay, France
fuz.bra@wanadoo.fr

Jean-Baptiste Denis
INRA, Jouy-en-Josas, France

Abstract — A survey of 500 pedigree dogs was carried out in the Paris region. For each dog, six behavioural traits were determined and ten of their astrological traits were retained. A statistical interpretation of the possible relationships between the two sets of traits was performed based on permutation tests. Two strong associations were detected between the angular positions of Jupiter and the Sun, and the extraversion dominant trait. There were indications of other associations. These associations have a remarkable resemblance to the standard associations usually proposed in “human” astrology.

Keywords: behaviour—dogs and humans—permutation test—astrology—survey

Introduction

For an empirical study, the dog is an appropriate subject for the investigation of possible relationships between birth time and the position of sky elements of the solar system. The precise aim of this study is to see if behavioural differences, attributable to these, appear in two-month-old dogs. There are of course differences between animals and humans but it seems reasonable to describe a dog’s behaviour with the usual descriptions employed by breeders even if these seem anthropomorphic. There is also a recognizable proximity of psychological relationships between dogs and humans compared with other animals (e.g. the cat, the rabbit or the snake).

The first position of the Sun in its ecliptic course, and at the same time, positions of the Moon and planets, the rising (Ascendant) and setting (Descendant) points, the highest (Mid Heaven) and lowest (Nadir) points within the 24 hours of a day were defined. This applied one of the classical tools of astrology, according to which a sky element situated in one of the four described points (= "Angular", ± 10°) is particularly important in determining behaviour. It must be emphasized here that, so far, almost no scientific confirmation has been sought for this. Other classical tools of astrology, such as signs of the zodiac related to the seasons, are impossible to investigate due to the
irregular fertility of females during the year (most births take place in spring and autumn).

The results obtained for dogs are then compared with those classically described in human astrology.

Methods

Organisation of the survey

A population of 500 pedigree dogs was identified by one of the researchers (S.F.B.). Pedigree dogs were used because breeders are always particularly attentive to the conditions of birth, given the potential value of the pups. Thus, when a female begins to give birth, a breeder will stay patiently by the mother day and night, ready to take the pups, note the time, individual colours and so on. When they sell the young dogs they need very precise information to answer the buyer's questions. Purchasers frequently want to know the time of birth, the order of births in the litter (was my dog first, second, or last? and so on . . .), how the pup behaved in its first few days and weeks of life. As the pups must live with their mother and cannot be sold until they are two months old, their behaviour is very well documented over this period. Every breeder of pedigree dogs keeps a very precise diary, where all this information is carefully entered for each animal, the individuals being identified either by colour differences (zones, patches, spots and so on) or in the case of uniform coloration, by means of a cropped area of the coat. (The official book, called "LOF" in France, records pedigrees and births.)

It was decided to use different breeds of pedigree dogs to prevent any bias linked to a given breed. They were: Bearded Collie, Belgian Shepherd, King Charles Spaniel, Chihuahua, Coton of Tulear, French Bulldog, German Shepherd, Labrador, Lhassa Apso, Malinois, Poodle, Sharpei, Shitzu, Tibetan Spaniel, and Yorkshire Terrier. Geographically, the kennels were all in the Paris area to ensure easy contact with the breeders.

The breeders who agreed to participate have no special knowledge of, or interest in, astrology. Over a period of five years, a total of 100 litters were investigated, from two to eight pups in each, for a total of 500 pups. Twelve breeders participated (see acknowledgments).

Recorded traits

For behavioural traits, data from the breeders were used. They noted all behavioural characteristics in detail during the two first months of the pups' lives. The breeders' notes were freely written in ordinary language. Information collected for the experiment was summarized according to Pr. Eysenck’s method (1975) by expressing behaviour under "Extraversion" and "Neuroticism", giving six well defined items. They are detailed in Table 1 and the transcription
from the free description is given in the Appendix. The different items are scattered over the entire range of births in the litters. There are many personality theories and various systems of behavioural description; we have chosen Eysenck's method as most appropriate to classify the very detailed observations of breeders because of its simplicity and non-subjectivity. For example, a dominant dog and a dominant human demonstrate the same characteristics – except, of course, for the absence of those involving speech. Whereas one can describe a dominant human as being a "powerful speaker" or being able to "capture his audience's ear" etc., these qualities would hardly be adaptable to a canine subject (without changing the meaning).

Numerous methods exist for the study of personality. Eysenck's was chosen for this study largely because of the arguments of its creator, summarised as follows: "To find out the laws according to which this may happen, and to isolate the major dimensions along which we can classify people, seems to me a fundamental and critically important part of psychology [...] These three major dimensions (P–E–N = psychoticism, extraversion, neuroticism) emerge from practically any large-scale analysis of traits published in the literature" (Eysenck, 1990).

The most important element is the group of behaviours attributed to each major trait derived from the PEN and these are easily recognised in the descriptions given by the breeders. The six major traits retained for this study are thus not just abstract characteristics but the result of pragmatic observations (see Appendix).

Insofar as the question of ascribing human traits to animals is concerned, the issue was addressed by Eysenck himself for whom this transposition was not only valid but an objective criterion: "... another criterion for the acceptability of major dimensions of personality, namely that they should be apparent not only in humans, but also in animals ..." (Eysenck, 1990). McFarland (1990) makes the same point.

Finally, the age of the dogs, two months at final evaluation, was considered satisfactory on the one hand because of the difference between the lifespan of dogs and of humans, and on the other because all breeders agree that behavioural structures of pups are formed very early in the context of the social group consisting of the bitch and her litter.
For astrological traits, the following ten sky elements were considered: Sun, Moon and eight planets of the solar system (Mercury, Venus, Mars, Jupiter, Saturn, Uranus, Neptune, and Pluto). All are usually defined as "Planets" in traditional astrology.

As described above in the introduction, the unique astrological criterion applied is the "angular" position of the sky elements, that is, rising, setting and highest and lowest points at the place and time of the birth. As the earth rotates, each element rises, sets and reaches its highest and lowest point every 24 hours either in the visible sky or, in the case of the lowest point, sometimes in that part of the sky which is invisible. The whelping of a bitch is always a slow process and the intervals between the birth of successive pups can vary between 15 minutes and two hours. This factor makes dogs particularly appropriate for this study.

The distribution of the 500 dogs is given in Table 2 (program "Astropc" from Aureas, 30, rue Cardinal Lemoine, 75005 PARIS France).

### Statistical Analysis

The objective was to explore possible links between behavioural traits and astrological traits. Rather than use sophisticated multivariate approaches such as correspondence factorial analysis, which are not always easily interpreted and from which it is not appropriate to draw inferences, it was decided to practice simple and well-known non-parametric tests for each of the 60 behaviour traits in planet traits combinations.

As an example, let us consider the $2 \times 2$ frequency table associated with Jupiter (Ju) and extraversion dominant (ED) which is a sub-table of Table 3. 44 pups are (Ju+,ED+), 65 are (Ju+,ED−), 76 are (Ju−,ED+) and the majority of them, 315, are (Ju−,ED−). To assess the degree of association between the two traits, we used the proportion of the ED+ dogs positive for the planet. That is $44/120 = 0.367$. It is worth mentioning that given the total margins of the table (109,
TABLE 3
Joint Distributions of the Dogs for Each Combination of Astrological Traits (in Rows) and Behaviour Traits (in Columns)*

<table>
<thead>
<tr>
<th></th>
<th>EA+</th>
<th>EA-</th>
<th>ED+</th>
<th>ED-</th>
<th>ER+</th>
<th>ER-</th>
<th>NA+</th>
<th>NA-</th>
<th>NN+</th>
<th>NN-</th>
<th>NS+</th>
<th>NS-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Su+</td>
<td>56</td>
<td>51</td>
<td>42</td>
<td>65</td>
<td>24</td>
<td>83</td>
<td>43</td>
<td>64</td>
<td>7</td>
<td>100</td>
<td>42</td>
<td>65</td>
</tr>
<tr>
<td>Su-</td>
<td>181</td>
<td>212</td>
<td>78</td>
<td>315</td>
<td>113</td>
<td>280</td>
<td>151</td>
<td>242</td>
<td>36</td>
<td>357</td>
<td>140</td>
<td>253</td>
</tr>
<tr>
<td>Mo+</td>
<td>48</td>
<td>58</td>
<td>21</td>
<td>85</td>
<td>35</td>
<td>71</td>
<td>33</td>
<td>73</td>
<td>10</td>
<td>96</td>
<td>39</td>
<td>67</td>
</tr>
<tr>
<td>Mo-</td>
<td>189</td>
<td>205</td>
<td>99</td>
<td>295</td>
<td>102</td>
<td>292</td>
<td>161</td>
<td>233</td>
<td>33</td>
<td>361</td>
<td>143</td>
<td>251</td>
</tr>
<tr>
<td>Me+</td>
<td>50</td>
<td>60</td>
<td>35</td>
<td>75</td>
<td>25</td>
<td>85</td>
<td>47</td>
<td>63</td>
<td>9</td>
<td>101</td>
<td>37</td>
<td>73</td>
</tr>
<tr>
<td>Me-</td>
<td>187</td>
<td>203</td>
<td>85</td>
<td>305</td>
<td>112</td>
<td>278</td>
<td>147</td>
<td>243</td>
<td>34</td>
<td>356</td>
<td>145</td>
<td>245</td>
</tr>
<tr>
<td>Vu+</td>
<td>43</td>
<td>45</td>
<td>18</td>
<td>70</td>
<td>27</td>
<td>61</td>
<td>39</td>
<td>49</td>
<td>8</td>
<td>80</td>
<td>29</td>
<td>59</td>
</tr>
<tr>
<td>Vu-</td>
<td>194</td>
<td>218</td>
<td>102</td>
<td>310</td>
<td>110</td>
<td>302</td>
<td>155</td>
<td>257</td>
<td>35</td>
<td>377</td>
<td>153</td>
<td>259</td>
</tr>
<tr>
<td>Ma+</td>
<td>52</td>
<td>61</td>
<td>31</td>
<td>82</td>
<td>29</td>
<td>84</td>
<td>48</td>
<td>65</td>
<td>5</td>
<td>108</td>
<td>44</td>
<td>69</td>
</tr>
<tr>
<td>Ma-</td>
<td>185</td>
<td>202</td>
<td>89</td>
<td>298</td>
<td>108</td>
<td>279</td>
<td>146</td>
<td>241</td>
<td>38</td>
<td>349</td>
<td>138</td>
<td>249</td>
</tr>
<tr>
<td>Ju+</td>
<td>58</td>
<td>51</td>
<td>44</td>
<td>65</td>
<td>20</td>
<td>89</td>
<td>48</td>
<td>61</td>
<td>8</td>
<td>101</td>
<td>38</td>
<td>71</td>
</tr>
<tr>
<td>Ju-</td>
<td>179</td>
<td>212</td>
<td>76</td>
<td>315</td>
<td>117</td>
<td>274</td>
<td>146</td>
<td>245</td>
<td>35</td>
<td>356</td>
<td>144</td>
<td>247</td>
</tr>
<tr>
<td>Sa+</td>
<td>38</td>
<td>55</td>
<td>26</td>
<td>67</td>
<td>24</td>
<td>69</td>
<td>29</td>
<td>64</td>
<td>12</td>
<td>81</td>
<td>38</td>
<td>55</td>
</tr>
<tr>
<td>Sa-</td>
<td>199</td>
<td>208</td>
<td>94</td>
<td>313</td>
<td>113</td>
<td>294</td>
<td>165</td>
<td>242</td>
<td>31</td>
<td>376</td>
<td>142</td>
<td>263</td>
</tr>
<tr>
<td>Ur+</td>
<td>53</td>
<td>56</td>
<td>23</td>
<td>86</td>
<td>30</td>
<td>79</td>
<td>46</td>
<td>63</td>
<td>8</td>
<td>101</td>
<td>37</td>
<td>72</td>
</tr>
<tr>
<td>Ur-</td>
<td>184</td>
<td>207</td>
<td>97</td>
<td>294</td>
<td>107</td>
<td>284</td>
<td>148</td>
<td>243</td>
<td>35</td>
<td>356</td>
<td>145</td>
<td>246</td>
</tr>
<tr>
<td>Ne+</td>
<td>47</td>
<td>52</td>
<td>26</td>
<td>73</td>
<td>25</td>
<td>74</td>
<td>47</td>
<td>52</td>
<td>7</td>
<td>92</td>
<td>37</td>
<td>62</td>
</tr>
<tr>
<td>Ne-</td>
<td>190</td>
<td>211</td>
<td>94</td>
<td>307</td>
<td>112</td>
<td>289</td>
<td>147</td>
<td>254</td>
<td>36</td>
<td>365</td>
<td>145</td>
<td>256</td>
</tr>
<tr>
<td>Pi+</td>
<td>55</td>
<td>50</td>
<td>20</td>
<td>85</td>
<td>33</td>
<td>72</td>
<td>41</td>
<td>64</td>
<td>12</td>
<td>93</td>
<td>32</td>
<td>73</td>
</tr>
<tr>
<td>Pi-</td>
<td>182</td>
<td>213</td>
<td>100</td>
<td>295</td>
<td>104</td>
<td>291</td>
<td>153</td>
<td>242</td>
<td>31</td>
<td>364</td>
<td>150</td>
<td>245</td>
</tr>
</tbody>
</table>

Note: Each 2 X 2 sub-table comprises all 500 dogs. The two most significant sub-tables are in bold.

391, 120, 380) this statistic is equivalent to all scores one can imagine to measure the link between the two traits (one degree of freedom is involved). For instance, the odds of the behavioural trait among the JU- dogs (761315 = 0.241) can be expressed as (120(1 - 0.367))/(391 - 120)(1 - 0.367).

For the same reason, the Chi-square statistics of independence can be expressed as a function of this proportion, so our permutation test is equivalent to the Chi-square test. The advantage of the proportion is that the type of association (positive = high proportion; negative = low proportion) is preserved.

Once a proportion has been computed, the existence of a significant association between the two traits must be tested. To this end the classical procedure of permutation tests (Good, 2004) was used. The principle is simple: under the null hypothesis of no effect a large number of similar samples of data (having the same margins) are simulated. For each of them the proportion is computed, providing an empirical distribution where no effect is present. This must be done for a sufficient number of simulations, say N, with respect to the level of the test, say a, one wants to perform. Finally the observed proportion is compared to this distribution, and if it is outside the (a/2 quantile, [1 - a/2] quantile) interval, then the effect is declared significant.

To perform the random permutations, the elementary data set can be seen as a matrix of 500 rows by two columns, where rows correspond to dogs and columns to the two traits. A (1,1) row means that the corresponding dog is
positive for both traits; a (0,1) row means that the corresponding dog is negative for Ju trait but positive for ED trait; and so on. The number of (1,1) rows is 44, the number of (0,1) rows is 76, and so on. If there is no link between the two columns, we can permute without consequence the first column giving rise to different numbers of (1,1), (0,1), (1,0), (0,0) dogs but keeping 120 ED+ dogs, 380 ED− dogs, 109 Ju+ dogs and 391 Ju− dogs. A new proportion can be calculated and stored. This is done N times.

Another point deserves some consideration: the level \( a \) at which the tests were performed. The traditional level is 5\%: \( a = 0.05 \). However in the present case, 60 tests were carried out on the same set of data. If this level were used and no links existed between any of the pairs of traits, we would nevertheless expect to see about three (=0.05 × 60) significant tests. To avoid this inconvenience, the 5\% level was used globally, dividing it by 60 (using \( \alpha/2 = 0.0004 \)) according to a majoration known as Bonferroni inequality. This resulted in a substantial decrease in the probability of stating significant effects, that is, producing a very conservative procedure. The less stringent correction proposed by Benjamin & Hochberg (1995) was also used. To obtain sufficient precision for such extreme quantiles, \( N = 1,000,001 \) permutations was chosen. In this case the number of values greater or less than the target quantile of \( \alpha/2 = 0.0004 \) is 400 simulated values. This is the traditional statistical theory. P-values were computed for each test, giving the significance for every level. If the P-value is 0.02, then the corresponding test is significant for greater levels (e.g. 5\%), and not significant for lower levels (e.g. 1\%).

It is worth noticing that a possible litter effect is not taken into account, and could bias the planet effect under study. But due to the fact that whelping takes several hours, pups belonging to the same litter have different planetary positions, and the consequence of neglecting such an effect is attenuation of the planet effect.

Using this approach, two planets were found to have an effect on the same trait of behaviour. It was therefore decided to examine the possible interactive effect of the planets. To this end, the planet1 \( \times \) planet2 \( \times \) behaviour trait table (\( 2 \times 2 \times 2 \)) was considered as a 4 \( \times \) 2 table, with four rows associating the combination of planets and two columns for the behaviour trait.

This provided a Chi-square of independence with three degrees of freedom that was further broken down, nesting the two planets' effects according to the two possibilities.

**Results**

The distribution of the dogs over all combinations of behaviour trait and astrological trait is given in Table 3. The main results of the statistical tests are proposed in Table 4 and for trait ED in Figure 1. For the global level of \( a = 0.05 \), Hochberg’s correction and Bonferroni correction gave identical results: two significant tests out of the 60. These are the associations between ED and Jupiter
TABLE 4
Indications for Each Behaviour Trait Detected by Statistical Analysis and Classical Signification
Attributed in Traditional Astrology for Humans

<table>
<thead>
<tr>
<th>Behaviour trait</th>
<th>Associated planet (with P-values of significance)</th>
<th>Traditional interpretation for humans</th>
</tr>
</thead>
<tbody>
<tr>
<td>EA (active)</td>
<td>Jupiter in excess (−, 0.069)</td>
<td>Active, extravert, sociable, charismatic</td>
</tr>
<tr>
<td></td>
<td>Saturn in deficit (−, 0.099)</td>
<td>Not reserved, not introvert</td>
</tr>
<tr>
<td>ED (dominant)</td>
<td>Jupiter in excess (***, 0.000)</td>
<td>Active, extravert, sociable, charismatic</td>
</tr>
<tr>
<td></td>
<td>Sun in excess (***, 0.00002)</td>
<td>Strong personality</td>
</tr>
<tr>
<td></td>
<td>Mercury in excess (*, 0.012)</td>
<td>Communicative</td>
</tr>
<tr>
<td></td>
<td>Pluto in deficit (−, 0.112)</td>
<td>?, various interpretations</td>
</tr>
<tr>
<td>ER (reserved)</td>
<td>Jupiter in deficit (*, 0.009)</td>
<td>Non-dominant, non-charismatic</td>
</tr>
<tr>
<td></td>
<td>Sun in deficit (−, 0.12)</td>
<td>Non-sociable, weak personality</td>
</tr>
<tr>
<td></td>
<td>Moon in excess (−, 0.059)</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>NA (affective)</td>
<td>Moon in deficit (*, 0.042)</td>
<td>Insensitive</td>
</tr>
<tr>
<td></td>
<td>Neptune in excess (*, 0.019)</td>
<td>Dreamy</td>
</tr>
<tr>
<td></td>
<td>Saturn in deficit (−, 0.059)</td>
<td>Unthinking</td>
</tr>
<tr>
<td>NN (nervous)</td>
<td>Mars in deficit (*, 0.047)</td>
<td>Lacking in force</td>
</tr>
<tr>
<td></td>
<td>Saturn in excess (*, 0.038)</td>
<td>Introvert</td>
</tr>
<tr>
<td>NS (stable)</td>
<td>P1 in deficit (−, 0.095)</td>
<td>?, various interpretations</td>
</tr>
</tbody>
</table>

Note: Effects are indicated as follows: planet effects detected (***); strongly suggested (*); and suggested (−). By detected, we mean that it is considered significant at 0.0002; by strongly suggested, that it is considered significant at 10%; and by suggested, that it is considered the strongest effect among the ten planets, or almost 10% significant.

and between ED and the Sun and they are amazingly strong. The drastic level we used for the Bonferroni test was far from exceeded. It is striking that not one of the 1000001 proportions computed for Jupiter was greater than the observed value. Some other much less impressive associations are suggested and these are shown in Table 4.

Concerning the effect of Jupiter and the Sun on the same behaviour trait (ED), possible interaction was analysed in the $(2 \times 2) \times 2$ table (Table 5). No additional effect was found among dogs positive for both Jupiter and the Sun. Both planets have a strong effect but it does not appear to be cumulative.

Discussion and Conclusions

This empirical study demonstrates that some relationships exist between the moment of birth of dogs characterized by the "angular" positions (e.g. rising, setting and upper/inferior culminations) of astrological planets, and independently assessed behaviour traits. They appear particularly strong in the case of dominant dogs influenced by the Sun, Jupiter and, to a lesser extent, Mercury.

The effects must be compared with one of the tools of classical human astrology concerning the relationship described (Fuzeau-Braesch 2004; Lewis 2003) for births with the Sun and Jupiter in these "angular" positions. Humans in this category are generally described as charismatic, dominant, strong, sociable and influential in a group. This is obviously comparable with the canine
Fig. 1. For the ED behaviour trait, the proportion of positive dogs for each of the ten planets is displayed (dots). The lines show the empirical distribution computed by the permutation tests: the (heavy) solid line is the median, the dashed lines are respectively, from bottom to top, the quantiles 0.0001, 0.0004 (solid), 0.001, 0.01, 0.05, 0.95, 0.99, 0.999, 0.9996 (solid), 0.9999. Planets have been ordered according to their P-values.

equivalent where the corresponding pup holds a dominant position among its peers during its two first months of life. It is always the first to eat and this is accepted by the entire group, and it will push the others away with impunity to get the attention of human attendants or just to move around, breeders report. This parallel is remarkable and can not be due to chance.

Other effects are no more than suggestions; probably a larger sample of dogs would be necessary to detect them statistically with greater confidence. Nevertheless, there are striking similarities with traditional human astrology indicated in Table 4. Notable among them are those concerning the Sun, the Moon, Mercury, Mars, Jupiter, Saturn and Neptune. A "nervous" (NN) dog is often born with Saturn in an "angular" position, which may result in a tendency to

<table>
<thead>
<tr>
<th>Table 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution of the 500 Dogs According to Jupiter, the Sun and the Extraversion Dominant Trait</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Ju+ and Su+</td>
</tr>
<tr>
<td>Ju+ and Su-</td>
</tr>
<tr>
<td>Ju- and Su+</td>
</tr>
<tr>
<td>Ju- and Su-</td>
</tr>
</tbody>
</table>
introversion. A lack of Mars can be a weakening influence and this, too, can result in a sensitive and timid animal.

The results for the "reserved" (ER) animals must also be considered here: they show Jupiter and the Sun in deficit: they are non-dominant, non-sociable, sensitive with the Moon in excess, which is also remarkably similar to classical interpretations for humans. An ambiguity must also be noted in the "affectionate" (NA) case. This term is always used by breeders for dogs which like being picked up and are happy to be handled: this is difficult to interpret. No convincing results have been obtained for "stable" (NS).

It may be underlined that the results are all the more convincing in that the tools we applied (description of behaviour, classical astrology) are not commonly in use for dogs.

The similarity between observations of dogs and human astrological descriptions can only be explained by the existence of a physical causal effect, so far unknown. Dogs seem to react in a very similar way to that which would be predicted by one of the classical astrological rules for humans, the "angular" sky elements. This eliminates the argument frequently advanced to "explain" this astrological tool; the fact that the human mother, knowing the birth chart of her children, influences her child in the "right" direction. Clearly no such cultural factor can occur in dogs. It is also difficult to evoke a factor of hereditary nature. For such a factor to be effective, all pups of a given letter should be borne under the same planet position, which is not the case due to the duration of whelping. Indeed, pups coming from the same litter have different behaviours and different sky positions.

Thus it must be supposed that a causal physical influence exists. It is worth recalling here various studies on the reception of waves emanating from sky elements, particularly the Sun and Jupiter. It is well known that in short wave radio, for example, receivers must be retuned at the rising, the culmination and the setting of the Sun, this being a result of the ionosphere acting as a plasma (Soloviev, 1998). Jupiter has also been much studied for its own waves which reach the Earth in spite of its magnetic environment (Rogers, 1995; Rosolen et al., 2002). Planetary magnetospheres of the various elements of the solar system are now a subject of new and vigorous research with spacecraft observation. They are very dynamic objects (Blanc et al., 2005) and it is not inconceivable that the time may be ripe to consider interdisciplinary work between astrophysics and astrology.

These observations in dogs must be followed up by much further similar research, in the search for more insight into the veracity and the limits of astrology. This is all the more necessary as so very few studies of the subject, anywhere in the world, have been so far recognized as scientific (Dean & Mather, 1977), with the exception of those of Gauquelin (1973, 1982) on angular planets and professions.

In future, studies may also concern the cognitive sciences linked to the organization of behavioural differentiation of individuals.
Acknowledgments

Thanks are due to the following breeders of pedigree dogs who kindly agreed to participate in this study: Mesdames and Messieurs Calais, Cattelain, Corillon, Falchi, Gora, Jenny, Ladiray, Lalliot, Le Borgne, Lepoudère, Morisset and Reinard. We would also like to thank the reviewers for their interesting comments, especially the indication concerning Hochberg’s correction for simultaneous testing.

References


APPENDIX

List of words used by the breeders (translated terms with original French terms) to describe the behaviour of pups, and how the words were associated with the six behaviour traits in the study.

ACTIVE (actif)
Active – actif
Bold – audacieux
Rascally – coquin
Daring – culotté
Curious – curieux
Clever – débrouillard
Bright – dégourdi
Impudent – effronté
Wide-awake – éveillé
Expressive – expressif
Frisky – exubérant
Go-getter – fonçeur
Cheerful – gai
Noisy – gueulard
Playful – joueur
Crafty – malin
Responsive – réactif
Animated – remuant
Spontaneous – spontané
Lively – vivant
Roguish – voyou
Vivacious – vif

**DOMINANT (dominant)**
Aggressive – agressif
Belligerent – bagarreur
Strong character – caractère fort
Boss of the litter – chef de la portée
Determined – décidé
Dominant – dominant
Shameless – effronté
Strong – fort
Greedy – gourmand
Eats well – mange bien
Snappy – mordant
Doesn’t give in – ne céde pas
Gets what he wants – obtient ce qu’il veut
Afraid of nothing – peur de rien
Knows what he wants – sait ce qu’il veut
Happy everywhere – se plaît partout
Beguiling – séducteur
Sociable – sociable

**RESERVED (réserve)**
Aloof – à l’écart
A little silly – bêta
Always give in – céde toujours
Timorous – craintif
Discreet – discret
Distant – distant
Dominated – dominé
Sleepy – dormeur
Not dominant – non dominant
Unaggressive – pas agressif
Unplayful – pas joueur
Timid – réservé
Self-effacing – s’écrase devant les autres
Solitary – solitaire
Touchy – susceptible
Shy – timide
The reviewers raised several points that readers should note.

First, are there any independent data to justify applying to dogs a scheme developed for humans? Any precedents? Moreover this scheme is only one among many that have been proposed—there is no mainstream consensus among astrologers.

Second: the classification of descriptors would have benefited from input from some disinterested outsiders, as a way of avoiding subjectivity. As it
stands, one wonders whether "dominant" really should subsume all of shameless, greedy, sociable, and happy.

Third: In trials of human astrology, some have suggested that one should not use young subjects because traits have not had enough time to be clearly expressed. Might not the same concern apply here? The subjects were 2-month-old puppies.
Exploratory Study: The Random Number Generator and Group Meditation

LYNNE I. MASON\textsuperscript{a} AND ROBERT P. PATTERSON

Department of Physical Medicine and Rehabilitation, Biomedical Engineering Institute, MMC 297, and Bakken Medical Instrumentation and Device Lab, University of Minnesota, Minneapolis, MN 55455
\textsuperscript{a}e-mail: Lynnemason08@yahoo.com

DEAN I. RADIN

Institute of Noetic Sciences, 101 San Antonio Road, Petaluma, CA 94952

Abstract — Experiments using truly random number generators (RNGs) have reportedly demonstrated anomalous deviations in various group settings. To explore these claims, group meditation (average 261 females, 398 males) was tested as a venue for possibly inducing these deviations using a true RNG located in a large meditation hall. A total of 94 hours and 33,927 trials, each trial consisting of 1,000 random bits collected in 10-second periods, were recorded during meditation (Transcendental Meditation and advanced techniques). Cumulative deviation results were in accordance with chance expectation for baseline data, but showed significant non-randomness for the first \( p < 0.00001 \) and second set of meditation data \( p < 0.000001 \). A sub-section of the meditations, known as "yogic flying," showed significant deviations for both the first \( p < 0.000001 \) and the second data sets \( p < 0.000001 \). Results at a second test location known as the Vedic Observatory were significant for the first \( p < 0.01 \) and second data collections \( p < 0.05 \). All results were analyzed for any possible mean drift by subtracting differences in the pre- and post-test baseline slopes. After the adjustment for any drift, the direction and the experimental results were still significantly atypical, with a greater number of zeros being generated than ones. The use of non-exclusive-or-ed methods to eliminate drifts of the mean of the random data is discussed as well as the use of RNGs for measuring changes in collective consciousness associated with standardized meditation.

Keywords: random number generator — random event generator — group consciousness — global consciousness — meditation — Transcendental Meditation—human/machine interactions

Introduction

The putative anomalous influence of groups of humans on truly Random Number Generators (RNGs) have been used to measure the effect of global and group consciousness in a variety of settings including meditations, meetings, ceremonies, sports events, and tragedies (Bierman, 1996; Jahn et al., 2000; Nelson 1997; Nelson, 2001; Nelson et al., 2002; Nelson et al., 1998; Radin, 1997, 2002, 2006; Radin et al., 1996). RNGs have also been used with
individuals and pairs to study the effect of human intention and human/machine interactions (Dunne, 1998; Jahn et al., 1997; Nelson et al., 1998; Radin & Nelson, 1989). Nelson has reported that RNGs or random event generators in group situations were found to act non-randomly with significant deviations of the means (or in some cases, variance) in situations involving "calm but unfocused subjective resonance" and those "that foster relatively intense or profound subjective resonance" (Nelson et al., 1998: p. 425).

Of the various contexts tested thus far, perhaps group meditations are closest to Nelson's (Nelson et al., 2002) prescription for the optimal environment to produce deviations in the RNG outputs. Previous research has suggested that time-synchronized as opposed to non-synchronized meditation appears to influence the RNG to a greater extent. That is, a meditation involving a large number of people worldwide practicing an assortment of types of envisioning, prayer and meditations at the same time reached significance (p = 0.047) (Nelson et al., 1998), as did another group meditation with a coordinated time (p = 0.012) (Nelson, 2002a). However, a third group meditation with a non-synchronized time yielded a non-significant result (Nelson, 2002a). It should be noted that the meditations in these tests included a wide variety of mental activities, from casual and celebratory to formal meditation techniques. The present study explored whether the group consciousness effect might be enhanced by using a single, standardized form of meditation practiced by hundreds of people at the same time and place.

Radin (2001, 2002), reported the effects of the violent events on 9/11/01 in the U.S.A. on a collection of international RNGs from the Global Consciousness Project (GCP) (Nelson, 2002a) that became significantly non-random with increasing variance. May and Spottiswoode, (2001) have presented a reanalysis of that data and contest the original interpretation of the results. In contrast to May and Spottiswoode (2001), four researchers independently report significant anomalies in the data (Nelson et al., 2002).

In response to 9/11/01 over 1700 practitioners of Transcendental Meditation gathered together from 9/23/01 to 9/27/01 at Maharishi University of Management (MUM) in Iowa. Additional meditations and extended group meditations with varying numbers of participants were organized in addition to their normal meditation schedule. The normal daily schedule called for group meditations to begin at 7:05 AM CST (except on Sundays, which were to begin at 7:35 AM) and at 5:20 PM CST.

RNG data from the GCP were analyzed from 37 RNGs located at different locations around the earth, but not including Iowa (Nelson 2002c). Significant deviations from chance were not achieved when evaluating all 735 minutes of data collected over the five days of meditations. On the day of the peak number of meditators (over 1800), there was an exploratory significant result (p = 0.0012). A trend was also reached for a specific section of the meditation period known as "yogic flying" when cumulated over the 5-day period. Nelson reported that the relatively small number of days, five, ruled out further analysis of the yogic flying deviations, underscoring the need for a longer multi-day study to allow for more
extensive investigation. Nelson (2002c) noted that during the yogic flying portions of the meditations the significant deviation was in the direction opposite to that observed in the majority of data from the Global Consciousness Project and from Princeton University's Princeton Engineering Anomalies Research (Jahn, 2002). This atypical direction result was also reported by Nelson during a Silent Prayer on 9/14/01, Full Moon ceremonies, sacred sites in Egypt, and a prayer vigil (Nelson, 2002a; Nelson et al., 1998). Because of these directional effects, Nelson (2000c) discussed including directional predictions in future meditation studies. Nelson (2006) stated "that a little more than half the events for the GCP that are somewhat like meditation show the downward trend."

There is an independent body of experimental evidence supporting the idea that large groups of meditators practicing a single type of meditation (Transcendental Meditation and advanced meditation practices) at a synchronized time have been found to decrease violence, crimes, car accidents, hospital admissions, alcohol consumption (Dillbeck, Landrith, & Orme-Johnson, 1981; Hagelin et al., 1999), war casualties (Orme-Johnson, et al., 1988), and improve the stock market performance (Cavanaugh, Orme-Johnson, & Gelderloos, 1989). A time lag or carryover effect that diminishes over months has been measured in studies evaluating the effect of group meditation on societal indexes (Dillbeck, 1990; Dillbeck et al., 1987; Hagelin et al., 1999; Orme-Johnson et al., 1988). The effects of these group meditations appear to involve a distance factor, with the effect being greater in the vicinity of the meditation groups (Hagelin et al., 1999). Similarly, RNG research has shown potential distance effects with peak effects closer to the source of large global events as measured by hemispheres, continents, country and region (Radin, 2001), but further research is necessary because distance effects in research involving intention typically have not been found when experimenting with individual subjects (Radin, 1997; Jahn and Dunne, 1987). The use of both local RNGs as well as distant RNGs with meditation groups would be necessary to conduct a systematic study of the role of distance.

The objective of the present study was to extend the previous research on RNGs and meditation by 1) expanding the number of meditation sessions, 2) incorporating a local RNG at the site of interest, 3) measuring a standardized type of meditation practiced at coordinated times, with a precise count of participants, and 4) taking note of the direction of the non-random nature of the response. Three predictions were made:

a) groups of people practicing the same meditation simultaneously in one location would result in a significant departure from chance expectation (50% ones and 50% zeros, an 0.5 expectation), specifically the cumulative deviation of the percent zeros would be greater than chance expectation obtained on a local RNG as measured over the whole meditation, b) particularly for a specific subsection of the meditation known as yogic flying, and c) the direction of the non-randomness would show a decrease in ones and thus an increase in zeros.
Equipment and Methods

A laptop computer and a truly Random Number Generator (Orion V1.2) was employed. The Orion RNG uses noise-based analog signals that are converted into random bit streams. These bits are transmitted in the form of random bytes to a standard RS-232 serial port. According to the manufacturer's manual (Orion 2006) the baud rate is 9600 characters per second and the device is capable of supplying about 960 random bytes or 7600 random bits per second. Co-author Radin notes that transmission of a byte in the context of serial communications takes 10 bits, not 8, so the Orion provides about 9600 random bits per second.

The RS-232 port was tested for accurate minimum voltage (> 5V) with the actual voltage at 8.9 volts. The field recordings used a battery source for the laptop and a time-stamped marker for recording sections of interest.

Additionally, a second type of RNG (Mindsong, Inc. Research, microREG) was used for a limited time. The Mindsong is described by the manufacturer as incorporating, "Brownian movement of electrons using a Junction Field Effect Transistor (JFET) in a high gain circuit that generates the Noise signal" (Haarland, 2003). Non-deterministic randomicity is assured by the electron noise in this JFET circuit. According to the Mindsong's manufacturer, (Haarland, 2003) the bits are transmitted to a standard RS-232 serial port with a baud rate of 9600 characters per second and a 2600 bits per second sampling rate. The majority of research involving RNGs and consciousness has been done with additional software to apply exclusive-or (xor) logic to the data. In the xor technique, the raw data from the random number generator is "masked" or "exclusive ored" (xored) either against a pseudo random byte or a regular 0/1 sequence. According to the RNG manufacturers, the advantage of xoring is to ensure randomness with less chance of a bias; specifically, it eliminates systematic drifts in the mean. A disadvantage of xoring the data against a fixed mask is that the output is no longer raw binary data, and it may constrain long-term changes in the mean numbers of ones and zeros. Scargle (2002) has proposed using non-xored data. Scargle (2006) explains that using a logical xor operation and reversing of the data in the bit stream may totally eliminate according to the design philosophy anomalous effects and all physical effects in consciousness research. Nelson (Nelson et al., 2002) emphasizes the existing large database of RNG studies that use xoring and have significant experimental effects contradicts Scargle’s viewpoint.

The RNGs manufactured by Mindsong have additional hardware xoring (Bradish et al., 1998; Haaland, 2003). As stated in the patent, "the analog output of this random signal is converted to a random binary stream ... and further treated with a selective inverter that inverts some but not all of the series of data values according to a pseudo-random sequence mask. The selective inverter coupled to a sampler that inverts some, but not all, of the series of digital data values to produce a selectively inverted series of digital data values is an essential feature of the patent and our device. One of the benefits of this is the prevention of baseline drift."
In conclusion, a comparison of software xored and non-xored data initially at baseline was conducted. This was followed by the use of non-xored data (no additional software for xoring) for the rest of the experiments based on the design specification of the RNG: to avoid altering the original data with a mask (Scargle, 2002).

One trial with the Orion and the Mindsong RNGs consisted of 1000 bits collected every 10 seconds with 1000 trials per run (approximately 10,000 seconds/run or 1,000,000 bits/run) for 2.78 hours. This configuration was chosen for its capacity to capture one complete meditation period (approximately 120 minutes which is well within the 2.78 hours limit), in one run, and within the capacity of the laptop batteries. Acquisition and analysis software (Watson, version MREG00s1, 2001) provided the total number of bits counted, and the calculation of the deviation of ones (likewise zeros) from the RNG. Count mode was set for ones to indicate a positive, increasing direction, and conversely increased zeros to indicate a decreasing, negative direction. RNGs with increasing ones means there is less randomness due to generating more ones than zeros. In the acquisition software used this was designated by an increase in the positive upward direction as represented on the analysis graphs. Increasing zeros means there is less randomness due to generating more zeros than ones. This is represented by an increase in the negative downward direction of the graphs. This was a mean-shift analysis not an analysis of variance. The information in cumulated deviation (in reference to 50% ones, an 0.5 expectation) of the ones counted from the trials consisting of 1000 bits every 10 seconds was used for the statistical analysis.

Data analysis followed procedures previously described (Jahn and Dunne, 1987) for single RNG use and with Z-scores (Radin, 2002). Specifically sequential samples of 25 bytes were collected from the RNG since each byte consists of 8 bits and each sample yields 200 bits of ones and zeros. The number of ones beyond 100 (100 is the theoretically expected mean) were counted in each sample and this number was added to the previous accumulated number. The total number of bits counted was calculated, the percent deviation of ones was calculated, and the sums of the deviation were calculated. Z scores based on 95% confidence levels were calculated. The Z scores $Z = x - \mu/\sigma$, where $x$ is the sample value, $\mu$ = mean, and $\sigma$ = standard deviation. All tests are reported as one-tailed.

Please note the term "xored" as used here refers to additional software xoring. Likewise "non-xored" refers to not using additional software for xoring. No changes were made internally to any of the RNGs as described above which employ internal xoring techniques.

Four main tests were run (See Chart 1):

- a pre-test baseline "control" period in our laboratory comparing xored and with no additional software xoring (non-xored) data for the Orion RNG as well as a pretest baseline "control" period in our laboratory for the Mindsong RNG with no additional software xoring (non-xored).
- experiment A, consisting of recordings of a meditation group, a subsection
of the group meditation known as yogic flying and also a site known as the Vedic Observatory for the Orion and Mindsong RNGs.

- experiment B, consisting of a replication of experiment A but for a longer period using the Orion RNG.
- a post-experimental "control" period in our laboratory/offices using the Orion RNG.

No formal predictions were made for the post-experimental "control" results. The experimenters in this exploratory research decided they did not have enough information to precisely predict a possible lag effect, but a lag effect was considered as previous studies had reported "carryover" effects after the meditations had ended (Dillbeck, 1990; Dillbeck et al., 1987; Hagelin et al., 1999; Orme-Johnson et al., 1988.) The effects do not appear to end immediately when the group meditation is over, just as the music does not immediately end when a coherent orchestra stops playing. We hear the music for a few moments after the musicians have stopped playing, the sound lags or carries over. Likewise the effects of the meditation as recorded by the RNG may not end when the meditation recordings are over but also carry over or lag.

The control condition in this experiment was defined in purely operational terms, i.e., data collected while not "exposed" to meditation. Nelson et al. (1998, p. 452) notes "that even in laboratory experiments there is evidence

---

**Chart 1:** Shows a summary of the procedures of the experiments and the resulting slopes. Slope is the number of excess ones beyond expectation when random, per one thousand bits.
that traditional control data may not be immune to anomalous effects of consciousness.

_The Laboratory Pretest Baseline Control Comparison of Xored and Non-xored Data_

The Orion RNG was run in our University device lab in a small office, 5’ X 7’, used occasionally as a library and 3’ from the first author’s desk. Data collection consisted of 30 hours xored, 30 hours non-xored and an additional 89 hours non-xored with the Orion RNG. As noted previously, non-xored in this paper refers to no additional software xoring masking of data. These were considered the pretest “control” baseline samples for an inactive period. A Mindsong RNG non-xored was also run in our lab.

Experiment A in the Meditation Hall for Group Meditations Including Yogic Flying

Location of the RNG was at Maharishi University of Management, (MUM) in Fairfield Iowa, a rural university town with a population of approximately 10,000, with two meditation halls 0.25 km apart and designated by gender. RNG recordings took place in the first hall with an average of 261 female meditators (range 178–356, sd = 42). Attendance numbers are methodically tallied before each meditation for the purposes of future research at MUM. Collection during the summer was practical due to our laboratory’s overall research schedule, even though fewer meditators were present because of their summer vacation schedules. It should be noted that in addition to the meditators in the first hall where the RNG was located, there were male meditators in the adjacent meditation hall (average 398) during the recording periods on the same time schedule practicing the same meditation techniques. Only the first two authors and a research officer (from MUM) were aware of the recordings taking place. The meditation group was not aware of the experiment and the RNG located in the woman’s meditation hall was not visible to the participants. Since the participants were not aware of the RNG experiment the participants did not use intention to attempt to influence the results.

Meditation recording sessions lasted approximately 2 hours each and began at 7:05 CST (Sundays at 7:35 AM) and 5:20 PM CST. Our schedule and budget estimates allowed us to collect in two trips for a total of 94 hours of meditation data. This would expand the previous research done at MUM of 58.75 hours over 5 days (Nelson, 2002~). It would also exceed the RNG meditation research conducted elsewhere that used multiple meditation techniques including a single 3-minute period, a single 10-minute period and a single 1-hour period (Nelson, 2002a).

Extended and additional smaller group, meditations scheduled before and after the main group meditations (7:05 AM CST and 5:20 PM CST), in the meditation halls made planned comparison before or after the meditation periods
impractical. Furthermore, others have discussed a residual effect or lag effect of the group meditation predicted to last even after the daily meditation time is finished (Oates, 2002). The meditation hall has been used for meditation for over 20 years, 365 days a year and accordingly may not qualify as a neutral non-active control site even during non-meditation periods of the day.

Analysis was planned for the entire meditation period as a whole and then following Nelson (2000c) to analyze a specific section of the meditation known as yogic flying. Phenomenological reports of yogic flying include descriptions of waves of bliss (Alexander and Langer, 1990). Yogic flying is based on the ancient Yoga sutras of Patanjali (1978) and is predicted to create peace in the collective consciousness (Hagelin et al., 1999).

Experiment A in the Vedic Observatory. Within 5 kilometers of the meditation hall is an open-air site known as the Maharishi Vedic Observatory. It consists of ten precisely designed and positioned astronomical instruments based on ancient designs of sundials or "yantras," each about 2 meters high (Global Vedic Observatories Corporation, 1996). Observing the instruments is predicted to create psycho-physiological balance (Global Vedic Observatories Corporation, 1996) as well as development of "peak experiences" (Maslow, 1962) and stabilized "higher states of consciousness" (Alexander & Langer, 1990; Mason et al., 1997; Travis et al., 2002).

Each recording session using the RNG located in the center of the Vedic Observatory was approximately 90 minutes long. The majority of sessions involved only the first author at the Vedic Observatory site, although there were short periods of unscheduled visitors in a minority of sessions. The majority of sessions were recorded between 1:00 PM CST and 3:00 PM CST with a few exceptions due to weather and schedule conflicts. None of the recordings at the Observatory were made during the same time as the meditation recordings.

Experiment B

Replication of experiment A with increased data collection for 18 days.

In the Laboratory Post-experimental "Control"

Non-xored recordings in our laboratory repeated the pre-test baseline "control" recordings.

Results

Laboratory Pre-test Baseline "Control" Comparison of Xored and Non-xored Data

The baseline for both xored and non-xored data in the laboratory setting were random for the RNG (Orion). Results show no significant terminal (end of interval) non-randomness in the 30 hours, 10,883 trials, or 10,883,000 bits of xored data ($Z = 0.834, p = 0.798$) (Figure 1A). Likewise, there was no significant
A pre-test control period showing non-significant deviations from a RNG. Pre-test RNG xored shows 30 hours of data, 10,883 trials, 10,883,000 bits of xored data. Pre-test RNG non-xored shows 30 hours, 10,883 trials, 10,883,000 bits of non-xored data using the same equipment, RNG and location in our lab. As expected the RNGs did not reach terminal significance at the $p = 0.05$ level in the control period.

An additional pre-test control RNG non-xored period of 89 hours, 32,000 trials, 32,000,000 bits of non-xored data from a RNG in our lab. As expected there is no significance for the control period at the $p = 0.05$ level. The cumulative deviation plots show parabolic lines for one standard deviation and $p = 0.05$ for a chance criteria as a function of increasing trials. The jagged solid lines show the cumulative deviations over all the trials. SD = standard deviation. RNG = Random Number Generator.
terminal non-randomness for 30 hours, 10,883 trials, 10,883,000 bits of non-xored data (terminal $Z = -1.33, \ p = 0.091$) (Figure 1A) and for 89 hours, 32,000 trials, 32,000,000 bits of non-xored data (terminal $Z = -1.138, \ p = 0.127$) (Figure 1B).

**Experiment A in the Meditation Hall for Group Meditations**

Results included significant anomalies (terminal $Z = -8.434, \ p = 1.697 \times 10^{-17}$) for nineteen group meditation sessions from experiment A totaling 32 hours, 11,360 trials, 11,360,000 bits of data. As described below, data were reanalyzed to take into account a possible mean drift and still maintained significance (terminal $Z = -4.726, \ p = 1.1449 \times 10^{-6}$) (Figure 2A). The results were in a decreasing direction indicating increasing cumulative zeros.

The reanalysis for a possible mean drift involved finding the linear regression slope of the cumulative pre-test data (using the Orion) and subtracting it from the slope of the cumulative post-test data. This reanalysis was performed because a difference was found between the pre- and post-data terminal $Z$ scores. Slope was calculated on the cumulative deviation scores for both the pre-test data and post-test data non-xored. The comparison involved the non-xored pre-test data of 89 hours and the non-xored post-test data of 89 hours. The slope was determined by using regression data analysis in Excel 2000 with a zero intercept. The input ranges were the cumulative deviation in bits and the number of trials. The output variable represents the slope. The difference pre-post in the slopes were subtracted from each cumulative bits deviation score for each trial during the experimental phase. Specifically in this case the difference in the slopes pre- and post-test was subtracted from the meditation cumulative bits per trial scores.

**Experiment B in the Meditation Hall for Group Meditation**

Data collection consisting of 32 group meditation sessions totaling 63 hours, 22,567 trials and 22,567,000 bits of data from experiment B is also significant (terminal $Z = -9.068, \ p = 6.126 \times 10^{-20}$). The data was reanalyzed for a possible cumulative drift and remained significant (terminal $Z = -3.872, \ p = 5.397 \times 10^{-5}$), (Figure 2A.) The results were in a decreasing direction, indicating more zeros than ones.

**Experiment A in the Meditation Hall for Yogic Flying**

The yogic flying portions of the meditations are also highly significantly non-random (terminal $Z = -14.046, \ p = 4.12 \times 10^{-46}$) for the first set of data from experiment A consisting of 5 hours of data, 1,728 trials, 1,728,000 bits. It maintains significance (terminal $Z = -12.600, \ p = 1.061 \times 10^{-36}$) after reanalysis for a possible cumulative drift (Figure 2B).

**Experiment B in the Meditation Hall for Yogic Flying**

The data for the yogic flying portions of the meditations from experiment B are also significant (terminal $Z = -14.774, \ p = 1.087 \times 10^{-49}$) for 8 hours, 2,971
trials, 2,971,000 bits and maintains significance (terminal $Z = -12.639$, $p = 6.471 \times 10^{-37}$) after reanalysis for a possible cumulative drift (Figure 2B). The direction for the yogic flying data for experiments A and B is an atypical decreasing direction indicating more zeros than ones. The yogic flying slopes ($-9.52$ and $-9.03$) are higher (see Chart 1) as compared to the slopes of the meditations as a whole ($-2.18$ and $-2.03$).

**Mindsong RNG**

Figure 2C represents trials with a second type of RNG by Mindsong. The pre-test RNG Mindsong data used no additional software to xor the data. The pre-test data was taken in our laboratory's office at pre-test and was non-significant ($Z = 0.222$, $p = 0.5878$) for 23.5 hours of data, 8470 trials, and 8,470,000 bits. Meditation RNG Mindsong in Figure 2C was acquired without using software xoring and recorded in the meditation hall and is significant (terminal $Z = -5.248$, $p = 7.6951 \times 10^{-8}$) for 23.5 hours of data, 8470 trials, and 8,470,000 bits. The Mindsong RNG stopped functioning after 8000 trials and was unable to output data. There was no graphical display or numerical data display from the RNG.

**Experiment A in the Vedic Observatory**

The Vedic Observatory recordings for experiment A consisting of 24 hours of data, 8,918 trials, 8,918,000 bits are significantly non-random (terminal $Z = -5.950$, $p = 1.378 \times 10^{-9}$) and was significant after reanalysis for a possible cumulative drift (terminal $Z = -2.64$, $p = 0.004$) (Figure 3).

**Experiment B in the Vedic Observatory**

Experiment B consisted of significant Vedic Observatory recordings for 31 hours of data, 11,271 trials, and 11,271,000 bits, and were significantly random before (terminal $Z = -5.440$, $p = 2.664 \times 10^{-8}$) and after reanalysis for a possible cumulative drift (terminal $Z = -1.75$, $p = 0.040$) (Figure 3).

**The Laboratory Post-experimental "Control"**

Possible cumulative drift. The RNG (Orion) after approximately 480 hours of data collection consisting of 80–90 hours per month over 6 months did not appear to behave as it did before recording in the meditation hall and the Vedic Observatory. Specifically, when the RNG was rerun non-xored in our laboratory it showed a possible cumulative linear downward drift as compared to the pre-test non-xored control baseline in our laboratory. The slope was calculated as excess ones greater than expected if random, per 1000 bits. A negative number indicates an increased proportion of zeros. At post-test, 89 hours of post-test control data collection for 32,000 trials, 32,000,000 bites was significant (terminal $Z = -7.28$, $p = 1.70 \times 10^{-13}$) with a slope equivalent to $-1.3$ excess ones/1000 bits (Figure 4). At baseline before the experiment there was not a significant terminal Z score.
A

B

C

number of trials

cumulative deviation in bits

cumulative deviation in bits

cumulative deviation in bits

number of trials

SD

terminal probabilities

SD

terminal probabilities

SD

terminal probabilities

0.05

0.05

0.05

RNGA

RNGB

10^{-5}

10^{-36}

RNGA

RNGB

10^{-5}

10^{-37}

RNG M

RNG MP

10^{-8}

number of trials
downward trend (Figure 1B). At the end of the post-test period the RNG had not returned to pre-test baseline behavior. However, the terminal Z scores were not significantly different pre and post when run xored.

Electronic devices are more likely to fail or develop drifts in the early part of their operating life, then level off and then again increase in failures when they become older (US government inspector's technical guide, 1987). The new RNG (Orion) could have developed a cumulative drift with use. However, the manufacturer prior to shipping reported they tested the device for randomness. Co-author Radin reports continued randomness after 5 years of use with the same type of RNG. It is conceivable but not likely that both internally built-in xored independent random data streams in the Orion RNG developed similar biases resulting in more zeros than ones.

No formal predictions were made for the post-experimental results. The experimenters in this exploratory research decided they did not have enough information to precisely predict a possible lag effect. A lag effect was considered as previous studies with this type of meditation had reported a carryover or lag effect on the experimental measurements even after the experimental period of meditation had ended (Dillbeck, 1990; Dillbeck et al., 1987; Hagelin et al., 1999; Orme-Johnson et al., 1988).

Regardless of the reason for the possible mean drift the difference of the slope of the post-test data (−1.3 excess ones/1000 bits) from the pre-test control data

---

Fig. 2A. Shows significant deviations from randomness ($p = 1.1449 \times 10^{-6}$ adjusted), for Meditation RNG Experiment A during 19 group meditation sessions for a total of 32 hours, 11,360 trials, 11,360,000 bits of data. The 2-hour sessions of group practice of Transcendental Meditation and advanced practices involved an average of 261 females and 398 males. The results were replicated ($p = 5.397 \times 10^{-5}$, adjusted) as shown in Meditation RNG Experiment B with 32 group sessions for a total of 63 hours of meditation, 22,567 trials, and 22,567,000 bits of data. RNGA = Random Number Generator Experiment A, RGNB = Random Number Generator Experiment B.

Fig. 2B. Shows the concatenated accumulated deviations from a more advanced section of the meditations known as yogic flying. Yogic Flying RNG Experiment A shows nineteen, 15-minute sections, 4.7 hours, 1728 trials, and 17,280 bits. Yogic Flying RNG Experiment B shows 32 sections, 8 hours, 2,971 trials, and 29,710 bits. The yogic flying portions of the meditations are highly significantly non-random for both experiments ($p = 1.061 \times 10^{-30}$ and $p = 6.471 \times 10^{-37}$, adjusted) respectively and the slopes (Chart 1) are eight times more significant than the meditation data (Fig. 2A). The direction of the data is an atypical decreasing direction indicating increasing zeros. RNGA = Random Number Generator Experiment A, RGNB = Random Number Generator Experiment B.

Fig. 2C. Shows a second type of RNG by Mindsong. Pre-test RNG Mindsong shows a control period of 23.5 hours of non-xored data collection, 8470 trials, and 8,470,000 bits recorded in our laboratory and as expected, it is not significant. Meditation RNG Mindsong shows 23.5 hours of non-xored data collection, 8470 trials, and 8,470,000 bits taken during meditation that are significant ($p = 7.69 \times 10^{-6}$, adjusted). The cumulative deviation plots show parabolic lines for one standard deviation, and $p = 0.05$ plot of chance criteria as a function of increasing trials. The jagged solid lines show the cumulative deviations over all the trials. SD = standard deviation, RNGMP = Random Number Generator Mindsong Pretest, RNGM = Random Number Generator Mindsong.
Fig. 3. Data collected at the Vedic Observatory for RNG experiment A shows 24 hours of 8,918 trials and 8,918,000 bits and for RNG experiment B of 31 hours, 11,271 trials, and 11,271,000 bits. Both experimental results are significantly non-random ($p < .01$ adjusted), ($p < .05$ adjusted) respectfully.

The cumulative deviation plots show parabolic lines for one standard deviation, and $p = 0.05$ plot of chance criteria as a function of increasing trials. The jagged solid lines show the cumulative deviations over all the trials. SD = standard deviation, RNGA = Random Number Generator Experiment A, RNGB = Random Number Generator Experiment B.

(-0.2 excess ones/1000 bits) was calculated and this slope was subtracted from the original data for the meditation, yogic flying, and Vedic observatory (Chart 1). The difference in the slopes was subtracted from each cumulative bits deviation score for each trial during the experimental phase. The reanalyzed data takes into consideration a possible mean drift and is presented in Figures 2A, 2B and 3.

Malfunctioning and equipment failure was checked by substituting an alternate laptop and connector post-test. This did not change the post-test results which continued to have a significant terminal Z score. There was no indication of computer or connector failure. Bierman (2002) offered the opinion that if the RNG is terminally significant for some runs and non-significant for others in the upward direction and terminally significant and non-significant in the downward direction for others, this would not indicate a malfunctioning. In his viewpoint, if the RNG is malfunctioning, all the individual runs would be expected to be similar as opposed to a variety of results across runs. However, non-symmetrical distributions could affect outcomes.

Additional testing was performed for electrical and magnetic interference by running the RNG without additional software for xoring in an electrically
Fig. 4. Shows post-test RNG xored, 89 hours of xored data collection, 32,000 trials, 32,000,000 bits in our lab. The results are not significant. Post-test RNG non-xored, shows 89 hours of non-xored data collection, 32,000 trials, and 32,000,000 bits in our lab. The non-xored results are significant and could indicate a cumulative or mean drift or a lag effect of the meditation. The cumulative deviation plots show parabolic lines for one standard deviation, and \( p = 0.05 \) plot of chance criteria as a function of increasing trials. The jagged solid lines show the cumulative deviations over all the trials for data set 1 and data set 2. SD = standard deviation, RNG = Random Number Generator.

shielded isolated room for 48 hours post-test. The RNG (Orion) was run for 48 hours in a Faraday cage for electrical shielding with mu foil for magnetic shielding and then placed for 48 hours in a Faraday cylindrical cage with a height of 25 centimeters and diameter of 12 centimeters for electrical shielding with a 37 centimeters connector cord to distance the RNG from the laptop. The results using the Faraday cage (without additional software xoring) had a slope of \(-1.10\) similar to the experimental post-test slope of \(-1.34\) without the Faraday. The concern of electrical and magnetic interference was not supported.

**Z Score Analysis**

Following Radin (2002) percentages of significant Z scores at the 0.05 level for the pre-test, meditation, and yogic flying for experiments A and B were calculated to test for outliers. The purpose was to examine if many of the group meditation sessions and yogic flying sessions were significantly contributing to the outcome, and not just a few highly significant meditation or yogic flying sessions skewing the results. A window of 115 minutes was selected because it is the length of
Fig. 5. Z scores using a window of 15 minutes for all the yogic flying sessions (experiments A and B combined). This indicates that a majority of the yogic flying sessions were significant at the 0.05 level.

Discussion

As predicted, the meditation and yogic flying data are significantly anomalous (meaning more zeros than ones in the random binary stream) even after statistically controlling for a possible cumulative drift. The meditation data consisted of a total of 94 hours of standardized group meditation (average 261 females, average 398 males in adjacent meditation hall) recorded in two experiments, at uniform times, collected over a total of 30 days with a RNG on-site in the females' meditation hall. The two experiments are both significant and therefore the second experiment offers a replication of the anomalous results. Our results extend and support previous work (Nelson, 2002c, 2006) involving the same type of group meditation. The Vedic Observatory data was also significant after reanalysis for a possible cumulative drift, for the two experiments but
less so than the meditation. Future work could include subjects with the Vedic Observatory recordings as a more appropriate test of the putative influence.

**Direction of Results**

The direction of the anomalies in our work is similar to RNG research involving prayer, full moons and sacred sites in Egypt, but unlike that typically observed in the majority of work involving tragedies as seen in the Global Consciousness Project data (Nelson et al., 1998) and Princeton University's Princeton Engineering Anomalies Research labs. Nelson et al. (1998) note that often in past RNG research the focus has been on the variance and mean shift direction is ignored, so the methods in certain studies may make it inappropriate to infer any meaning from the direction. Our work adds supports to the premise that activities with "calm but unfocused subjective resonance" (Nelson et al., 1998, p. 425) or those that foster transcendental experiences (Alexander & Langer, 1990; Mason et al., 1997; Orme-Johnson et al., 1988; Travis et al., 2002), or "flow experiences" (Csikszentmihalyi, 1990) may reflect a more decreasing directional trend for the RNG. Specifically there is less randomness due to generating more zeros than ones. This can be represented by an increase in the negative downward direction of the graphs. By contrast, events that "foster relatively intense or profound subjective resonance" (Nelson et al. 1998, p. 425) involving emotionally laden environments, such as tragedies, may result in less randomness with deviations in the increasing direction.

Nelson (2006) notes that "Despite the accumulations of more than 200 events over the past 8 years we can not definitively interpret the negative versus positive slopes in either case there is a change toward less randomness." Many of the 200 events used variance measures so caution is advised in generalizing to studies involving mean shifts. If the present preliminary findings are supported with further formal confirmatory research in the future, this could lead to the use of RNGs as a potential means of measuring the intentional "direction" for collective consciousness.

**Alternative Explanations of Results**

What possible alternative explanations of the anomalous data could be responsible for the results? The following is an examination of other potential explanations including an experimenter effect, temperature bias, a trial density bias, insufficient number of sessions, non-xor influence, equipment failure, electro-magnetic interference, statistical bias, related factors, and lag effects.

*Experimenter effect.* In regards to the experimenter effect, various RNG experiments (Jahn & Dunne, 1987; Jahn et al., 1997; Nelson et al., 1998; Radin & Nelson, 1989) have shown a significant effect of individual intention on the RNG. It is possible that the conscious or unconscious intention of the experimenters influenced the results (Wisemen & Schlitz, 1997). This study could be replicated and designed specifically to test for experimenter effect including using other
experimenters with pre-registered intentions. However, an exploration with RNGs (Nelson, 2002b) found no definitive evidence for experimenter effects in a situation where the experimenter had a personal involvement in the subject matter and expectations about the outcome. Significant outcomes were not reached and the author concluded in this single study that there was no clear evidence of an experimenter effect for this deeply important personal event (Nelson, 2002b). A previous pilot study (Nelson, 2000c) conducted by a non-mediator testing the same meditation technique occurring in the same meditation halls, as the present study, did attempt to control for experimenter bias. Those data were collected first without any predictions, then before data analysis predictions were made by a meditator blind to the data. Those results were significant and do not lend support to an experimenter bias explanation for the present results. Likewise, the participants of the present study were not aware of the experiment and therefore had no specific intentions or subject bias for the results. The experimenter effect cannot be completely ruled out as the first two authors were aware of the time of the recordings but previous studies (Nelson, 2002a, 2000c) do not support this alternative explanation.

Temperature. Temperature biases do not appear to be a likely alternative explanation as all recordings were within the 4° to 32° Celsius range prescribed by the manufacturer's specifications. Furthermore, the recording temperatures for the pre-test control, meditation, yogic flying and post-test were all similar but the results vary for these different venues and cannot be explained by temperature effects. The alternative explanation of temperature being responsible for the results does not appear to be supported.

Trial density. A trial density of 1000 bits per trial was selected in order to capture the whole meditation period in 1000 trials. A trial density of 1000 has been previously used for RNG research without any reported concerns for trial density influencing the results (Nelson et al., 1998). The research on trial number bias density issues is still limited (Ibison, 1998), and future experiments could directly compare 1000 bits to other density levels to test the influence of particular random processes on statistical outcomes. The alternative explanation of trial density being responsible for the results does not appear to be supported.

Number of sessions. Were there sufficient meditations to accurately measure an effect? Other meditation research has involved multiple RNGs but the length of meditation ranges from a single 3-minute period (Nelson, 2002a), to 58.75 hours over 5 days of recording (Nelson, 2002c). In comparison to other meditation recordings the present study is longer, with a total of 94 hours of meditation sessions (51 meditation sessions), and appears sufficient within the context of the literature. The alternative explanation of the number of sessions being responsible for the results does not appear to be supported.

Non-xoring. Non-xor (no additional software for xoring) was used as there was no evidence in our baseline control tests to support using additional software xor data. In the baseline control tests (Figures 1A and 1B) there was no significance for the xor (xor refers to additional xoring software) and the non-xor
data. Further exploration of the advantages and disadvantages of using non-xored data appear to be warranted.

Assuming a cumulative drift exists, if non-xoring is responsible for the drift it would be expected to equally affect the non-xored control baseline data, and non-xored meditation test data. This is not what was found in the results. The pre-test baseline control data terminal Z score is not significant while the meditation data is significant, even though both are not xored. The validity of the baseline recordings can also be examined. However, the pre-test baseline control recordings (Figures 1A and 1B) are typically random as expected for RNGs and appear valid. Hence, the alternative explanation that not including additional software xoring being responsible for the results does not appear to be supported. Nonetheless we have reanalyzed the data for any possible cumulative drift and it remains significantly anomalous.

**Equipment Failure.** The pre-test non-significant baseline data is more random than the significantly non-random test data recorded during meditation and the post-test results. It is conceivable that our relatively new RNG was experiencing an electronic "burning-in period" that resulted in a difference in the pre-test control baseline with the post-test (US government inspector's technical guide, 1987). If the results were completely due to a linear burn-in there would not be significance after controlling for pre and post differences, but there is. Also if equipment burn-in was responsible for the results we would not expect the different results for the meditation, yogic flying and Vedic Observatory that were taken on the same days with the same RNG.

While the post-test software xored data is similar to the pre-test xored baseline data, the post-test non-xored data is clearly different from the pre-test baseline non-xored. At post-test the equipment was tested, and no evidence of equipment failure was found for the RNG or associated computer and connector. The alternative explanation of equipment age or equipment failure being responsible for the results does not appear to be supported.

**Electro-magnetic interference.** To determine if electro-magnetic interference was the source of the results the RNG were run electrically and magnetically shielded as well as unshielded. No evidence for electro-magnetic interference as an alternative explanation of the results was found, especially since the experimental data was taken with batteries as the power source. The alternative explanation of electro-magnetic interference being responsible for the results does not appear to be supported.

**Statistical bias.** For RNG research in general, a Bayesian statistical analysis as opposed to the null hypothesis with independent running means (not cumulative deviations) (Scargle, 2002) has been suggested as a more stringent approach to the results (Sturrock, 1997). Sturrock (1997) emphasized the limitations of Z scores and analysis using p values. In this study, it was thought prudent to use the accepted to date statistical methods (Radin, 2002). Future research will have to clarify this line of Bayesian inquiry. All data windows reflected the length of real-time events, not arbitrary times, and no data was
excluded from the analysis, therefore the results are not related to data manipulation or "data fiddling" (Scargle, 2002).

**Related factors.** Though the results are supportive of our exploratory predictions, at this point, it cannot be definitively concluded that the results are due to an influence of group meditation and/or the Vedic Observatory. Other factors besides meditation or related auxiliary factors to meditation could be involved. Further research could include ruling out the simple effect of large numbers of people in silence, or numbers of people sitting non-actively. However, no significance has been found for relatively silent non-mobile audiences at conferences (Nelson et al., 1998).

**Lag effect.** The post-test results of this study could be interpreted as a candidate for a carryover residual effect, lag effect or entrainment effect. Could using the RNG, during the meditations or at the Vedic Observatory create a lag effect or alter the results of RNG? A new RNG (Orion) developed possible cumulative drifts after exposure to the group meditation, but not before. Intentional time delay effects have been previously reported in the literature involving single subject studies (Dunne and Jahn, 1992). A time lag or carryover effect that diminishes over months has also been measured in studies evaluating the effect of group meditation on societal indexes (Dillbeck, 1990; Dillbeck et al., 1987; Hagelin et al., 1999; Orme-Johnson et al., 1988). Extensive longitudinal research would be needed to support or dismiss the lag effect or entrainment effect as an alternative explanation of the results.

Co-author Radin, reports continued randomness after up to 5 years of use with multiple Orion RNGs when run non-xored. However two of these RNGs were used for experiments involving meditation. Co-author Radin notes these two RNGs then developed in the post-test a downward drift similar to that reported in the present experiment. Preliminary reviews found no downward drift in subsequent non-meditation related experiments with these RNGs. Radin’s investigation and reanalysis of this previous data (Radin 2006; Radin & Atwater, 2006) is underway in order to discover if there is a mediation and RNG interaction responsible for the drift in the mean/variance or a lag effect or some other possible mundane answer.

It is not clear there is a cumulative drift involved or what the source of the drift is or if there is a lag effect. At this point, the difference in post-test data from pre-test baseline "control" data is not clearly accounted for, therefore a statistical control for any cumulative drift regardless of the source was used. The experimental data was still significant after reanalysis for a possible drift for the meditation, yogic flying and Vedic Observatory for both the initial experiment and its replication.

**Conclusion**

Our predictions for the meditation data, yogic flying and Vedic observatory data were significantly supported and were in the predicted direction. Our work adds to the premise that certain activities that foster transcendental experiences
(Alexander & Langer, 1990; Mason et al., 1997; Orme-Johnson et al., 1988; Travis et al., 2002) may reflect a more decreasing directional trend (increased proportion of zeros) in RNG outputs. Alternative explanations do not clearly account for the observed results. The results were still significant even after controlling for a possible cumulative drift of the mean from an unknown source.

To our knowledge this is the first experiment with specific predictions for the direction of a mean shift, and it involves the largest number of synchronized meditations recorded with a local RNG on site. Having a population doing a standardized mental technique on a regular basis is advantageous in studying various aspects of the phenomenon. Further research appears warranted to explore group meditation as a venue for anomalous results with the RNG. Future research could test the direction of the results, distance effects from the group, possible lag or entrainment effects, experimenter effect, non-xoring data techniques, group size effects, number of RNGs and possible auxiliary factors. Theoretical questions could include a continued inquiry (Hagelin, 1987; Nader, 2000; Nelson, 2002d; Radin, 2002; Routt, 2005) as to whether or not consciousness is a causal factor.

What is the possible practical contributions and application of this research? It is conceivable that RNGs could be used to indicate directional changes in a proposed global collective consciousness. Just as changes in seismic meters are used to detect high and low indications of impending earthquakes, RNG outputs could warn us of changes in collective consciousness while considering any anticipatory effects. RNGs could also be employed to evaluate preventive and ameliorative measures that utilize collective consciousness. For example, the RNG could evaluate the efficacy of various technologies from many traditions, including group meditations to reduce collective stress in global consciousness in order to prevent and reduce local and global tragedies.

Acknowledgments

The authors gratefully acknowledge the financial support of the Bakken MIND Lab from Earl Bakken. The researchers would like to extend their appreciation to Maharishi University of Management, the participants of the group meditations, and T. Fitz-Randolph, Director of the Vedic Observatory. We would like to thank D. Orme-Johnson, R. Nelson, A. Belalcazar, Y. Pu, and J. Zhang for editorial assistance; L. Stradal and D. Watson for technical expertise; and Charles N. Alexander and Otto Schmidt for continued inspiration.

References

Bierman, D. J. (2002). Personal communication.


Mindsong Inc. Mindsong MicroREG, Minneapolis.


RNG and Meditation


COMMENTARY

Comments on Mason, Patterson & Radin

Experiments on physical random number generators are fascinating for a very specific reason. In a general sense, we use the term "random" to refer to events that happen over time and/or space, for which we have no causal explanation. When we observe regularities between physical random number generators (RNGs) and events in the world, especially if they happen in experimental settings, we have to take notice. The reason is obvious; whatever influence the world events have on the physical RNGs must operate along causal pathways that we have not yet discovered. Surely, finding new causal pathways must be a central issue in scientific exploration.

Although I am enthusiastic about research on the effects of world events on physical RNGs, I am not as impressed as I would like to be with the results in the RNG literature, as are the authors of the various papers that make up this literature, many of which have appeared in JSE. The reason is based on my feeling that there is much room for improvement in the experimental designs and methods of analysis that RNG researchers use, and so I would like to use the Editor's generous offer to comment on the Mason et al. article in order to make points in general about RNG research, some of which are illustrated in the article.

Computations Related to RNG Data

One of my general criticisms of the RNG literature is that the computational procedures are frequently described in ordinary language, which does not always translate unambiguously into actual computation or statistical analysis. In my opinion this tradition is continued in the Mason et al. article. For this reason, I think it is worthwhile to make some of the computational issues more precise.

The raw data in an RNG experiment consist of a binary sequence; that is, a sequence $x_i$ for $i = 1 \ldots n$, in which each component $x_i$ is either 0 or 1. It turns out to be far easier to analyze binary data if we apply the "sign" transformation, $s(x) = 2x - 1$. This leaves 1's alone, but transforms 0's to $-1$. Thus, $s(x_i)$ for $i = 1 \ldots n$ is a sequence of 1's and $-1$'s. Note that summing $s(x_i)$ gives the excess of 1's over 0's in the underlying x-sequence (where a negative excess is interpreted as an excess of 0's over 1's), and that sums like this are routinely portrayed in RNG articles.

To reverse the 1's and 0's in the underlying x-sequence, we simply replace $x$ by $1 - x$. Since $s(1 - x) = -s(x)$, the reversal process for the signed sequence is
just accomplished by multiplying by \(-1\). There are other interesting algebraic properties of the sign function that are related to whether applying the exclusive-or operation is a good idea or not, one of the issues raised by the Mason et al. article. Define the eq operation on two binary numbers \(x\) and \(y\) so that \(x\) eq \(y\) is 1 when \(x\) and \(y\) are equal, and 0 when they are unequal. Then \(s(x\) eq \(y) = s(x)s(y)\), as can be easily checked. It would have been nice if the RNG scientists had combined sequences with eq, but instead they chose the xor operation (exclusive-or), defined by \(x\) xor \(y\) = 1 if \(x\) and \(y\) are unequal, and 0 if they are equal. Obviously \(x\) xor \(y\) = 1 – (\(x\) eq \(y)\), and so \(s(x\) xor \(y) = -s(x)s(y)\). The take-away point from this is that it is easier to study the effects of xor-ing two binary sequences using the sign transformation, although we do have to put up with an annoying sign change (which has implications, as we will see).

Combining binary sequences with xor happens in two places in RNG research. Evidently all putative physical RNGs actually generate two binary sequences internally, and then xor them for their output. This is done in hardware, so there is nothing anyone can do about it. I believe that the reason RNG manufacturers do this is because their goal is to offer a genuine random number source, which is not influenced by world events. To see why this makes sense, take expected values to show \(E[s(x)] = s(E[x])\), and note that \(E[x]\) is the probability of a 1 for the binary \(x\)-sequence. If \(x\) and \(y\) are two independent binary sequences, then \(E[s(x)s(y)] = s(E[x])s(E[y])\). Now \(E[x] = 1/2\) corresponds to "pure randomness", and \(s(1/2) = 0\). Therefore, the closer the expected value of the sign-transformed sequence is to 0, the closer it is to pure randomness. Since \(E[x\) xor \(y\] = \(-s(E[x])s(E[y])\), it follows that the \(x\) xor \(y\) sequence will always be closer to pure randomness than either \(x\) or \(y\) are (an order of magnitude closer). In fact, even if only one of the sequences is purely random, then the xor-ed sequence will also be purely random. Therefore, the RNG manufacturers can claim that by xor-ing they are delivering on their claim to produce a purely random number sequence. There are two aspects of this we need to keep in mind. (1) This is, of course, the opposite of the aims of RNG scientists, who want to be able to detect departures from pure randomness, so it is strange that they have chosen to use RNGs with hardware xor-ing, and this substantiates Scargle’s criticism, cited in the Mason et al. article. (2) All of the above assertions depend on the assumption that the \(x\) and \(y\) sequences are independent, which is perhaps somewhat less than obviously true.

The second place that the xor operation appears is in software "masking" of the sequence generated by the RNG. Evidently the most common scheme is to xor the signal \(x\) from the RNG with an alternating sequence of 0’s and 1’s. If we let \(a_i\) for \(i = 1 \ldots n\) denote this sequence, then \(s(a_i) = (-1)^i\). Thus, \(s(x\) xor \(a) = s(x_i)(-1)^{i+1}\).

We now have all the machinery we need to analyze RNG signals. First, the RNG internally generates binary sequences \(x\) and \(y\), and puts out \(x\) xor \(y\). The RNG scientist can either use this signal or xor it with the alternating binary sequence, to obtain \(x\) xor \(y\) xor \(a\). The sign-transforms of these signals are
Commentary

\[ s(x_i \ xor \ y_i) = -s(x_i)s(y_i) \]
\[ s(x_i \ xor \ y_i \ xor \ a_i) = s(x_i)s(y_i)(-1)^i \]

It has become quite conventional in RNG research to sum the sign-transformed version of a binary sequence for use in assessing non-randomness. Plots of cumulative sums have been much used in the RNG literature, and although Mason et al. repeat this, they base their conclusions on the “terminal” values of the sums. While this analysis seems to have served Mason et al. well, in general it is simplistic. One of the important ways that an RNG can fail to produce a random number sequence \( x \) is that the sequence \( p_i = E[x_i] \) of probabilities of 1 may depart from \( \frac{1}{2} \). Mason et al. refer to the situation \( p_i = p \neq \frac{1}{2} \) as "drift". (The reason this is a misnomer is that the sum of sign-transformed values can drift for other reasons.) The expected value and variance of the sum of the sign-transformed sequence are \( ns(p) \) and \( 4np(1-p) \), assuming \( p_i = p \) for all \( i \) and independence.

Because it will turn out to be important below, let us just consider the case \( p_i = p \) for the moment. Large-sample theory says (assuming the components of the binary sequences are independent) that approximately

\[ \frac{S - ns(p)}{2\sqrt{np(1-p)}} = Z \]

where \( S \) is the sum of the sign-transformed sequence, and \( Z \) represents a Normal chance variable with mean 0 and variance 1. Rewriting,

\[ S = 2\sqrt{np(1-p)}Z + ns(p) \]

The reason this is important is that RNG scientists regularly plot \( S \) vs. \( n \). If \( p = \frac{1}{12} \), then \( S = Z\sqrt{n} \), which explains why the curved lines in the plots shown by Mason et al. are proportional to \( \sqrt{n} \). The curves in the plots represent something about what we expect when \( p = 12 \). In order to see what would happen when \( p \neq 12 \), note that \( 2\sqrt{p(1-p)} \) is actually very close to 1 for values near \( p = 1/2 \). Thus, for small departures of \( p \) from \( \frac{1}{2} \) we have nearly

\[ S = \sqrt{n}Z + ns(p) \]

To summarize, the sum of a sign-transformed binary sequence should behave like the square root of the number of components times a standard Normal chance variable, but if the probability (\( p \)) of a 1 in the underlying binary sequence deviates from \( \frac{1}{2} \), then \( S \) should in addition have a component linear in \( n \) and proportional to \( s(p) \).

This sheds a bit of light on the issue of software \( xor-ing \). No matter what the value of \( p \) in the original sequence, \( xor-ing \) with the alternating sequence changes \( p \) to \( \frac{1}{2} \) (without changing the variance). Consequently, any software-xor-ed signal that gives a statistically significant result is rather hard to interpret, because (as Scargle argues) exactly what we might want to see has been completely removed. Needless to say, this makes previous positive research with software-xor-ed
sequences difficult to understand. I believe it is one of the major strengths of the Mason et al. article to have departed from the previous, convention-driven practice.

Before leaving this section, I want to point out that there is another very important point that is not addressed by this analysis. It is the possibility that the twin binary sequences $x$ and $y$, generated internally by the RNG, are not temporally independent. That is, pairs $(x_i, y_i)$ generated at one time might be correlated with pairs generated at other times, and of course each $x_i$ could be correlated with its paired $y_i$. This is, in fact, an entirely plausible way in which RNGs might produce non-random numbers, but the conventional analysis, based on partial sums of sign-transformed sequences, will never sort it out, because the method is too simple.

**Issues Raised by the Paper**

1. All of the experimental results reported by Mason et al. show plots of $S$ vs. $n$ with what appears to be very nearly a linear trend. In the light of the above analysis, this is consistent with the original binary sequence (from the RNG) having a departure from $p = \frac{112}{2}$, in the direction of $p < \frac{112}{2}$. The slope of each line is the value of $s(p)$ for that experiment. I estimated these from the figures in the article (in the version I had), and came up with Table 1.

<table>
<thead>
<tr>
<th>Figure</th>
<th>$s(p)$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2A</td>
<td>$-20000/(22500 \times 1000)$</td>
<td>0.49996</td>
</tr>
<tr>
<td>2B</td>
<td>$-24000/(3000 \times 1000)$</td>
<td>0.49600</td>
</tr>
<tr>
<td>2C</td>
<td>$-17500/(8500 \times 1000)$</td>
<td>0.49897</td>
</tr>
<tr>
<td>4</td>
<td>$-44000/(32500 \times 1000)$</td>
<td>0.49932</td>
</tr>
</tbody>
</table>

   If this is accurate, then a typical effect on the probability of a 1 in the underlying binary sequence is to shift it from 0.50 to something like 0.4986, a deviation of 0.0014.

   In my opinion, one of the weaknesses of the RNG research program is that it focusses on being able to collect sufficiently large numbers of bits to show that such tiny effects are real. It cites the "odds against chance" for its findings, and shows impressive-looking figures like those in the Mason et al. article, while underplaying just how miniscule the findings really are. If we want to assert that there are causal pathways that are undiscovered by conventional science, but which can be detected by RNG experiments, and that these causal pathways might actually have effects worth paying attention to in the world we live in, then the current path of RNG research does not seem to be taking us where we would like to go.

2. The authors say that a surfeit of 1's in a bit-stream indicates more randomness, while a surfeit of 0's indicates less randomness. No reason is
offered for this assertion, and indeed it is hard to imagine that simply reversing the sense of the data (multiplying the sign transformation by $-1$) would interchange things on some "randomness" scale. Given the analysis I set out above, I would offer a different view of the dominant negative trend in the Mason et al. results. A negative trend in the hardware-xor-ed binary sequence output by the RNG implies that both of its internal binary sequences have shifted in the same direction (both of their $p$'s above $\frac{1}{2}$, or both below $\frac{1}{2}$). If this is true, then the negative trend is more plausibly interpreted as a trend in the same direction by both internal sequences. What this might mean depends on how those internal sequences are physically generated, a fact not revealed in the Mason et al. article.

3. I find the authors' interpretation of the post-experiment results a bit strange. It seems disingenuous to say that there were no predictions about this phase of the experiment. An obvious reason for doing a post-experiment is to see that the RNGs returned to normal after the circumstances in which they showed an influence. When this fails, then introducing "entrainment" as a supportive explanation for the results in effect means that there is no way that the post-experimental results could ever falsify an RNG influence, raising a question about their scientific standing. Moreover, if "entrainment" were a serious explanation, we would have expected Mason et al. to report carefully on the prior history of the RNGs before their periods of data gathering, and we would have also expected to see cumulative "entrainment" accounted for somehow in the results during the meditation intervals. My conclusion is that while an "entrainment" hypothesis is appealing ab initio, it loses some of its luster post hoc.

4. In this paper (and virtually every other one I have seen on RNGs) the experimenters select a segment of bits from the underlying bit-stream, in some way that is not entirely clear. In other words, not all of the data generated by the physical random number generator are used. This amounts to applying a data "mask" that literally removes large amounts of data from the experiment. I am not suggesting that this was done in some sinister fashion, but I am claiming that this process is poorly described, and its effect on the statistical results has evidently never been tested. This is related to the next point.

5. I found the method of statistical analysis to be more obscure than I would have liked. At one point a "trial" is defined as 1000 bits collected over 10 sec, and a "run" is 1000 trials. But then in the description of the statistical method, 200 bits are sampled in an undescribed way over an undefined time period. If Fig. 1A (in the version I have) is to represent about 30 hours of sampling, then the sampling rate might be 0.1 Hz but with 200 rather than 1000 bits, or maybe with 1000 bits, or maybe at some other sampling rate—we cannot tell. Given that there is no deficiency of critics
of RNG research, it would be a good idea for RNG scientists to be more careful in describing what they have done.

6. A peculiarity of the pre-experimental data is this. If the original bit-stream is purely random, then after an alternating mask xor-ing it is still purely random. This is a result of probability theory, not something that needs to be tested empirically. It is not clear why one would test things known to be true, except as a negative control (which in this case would be a test that the software did the xor-ing correctly, something that could probably be more reliably checked directly).

7. It would seem to be useful in RNG research to have more controls than are usually employed. In this paper, for example, although there was suspicion that the continuous usage of the meditation hall for this purpose might have suggested an RNG influence from the site itself, no test of this hypothesis was made. Having simultaneous control RNGs in different locations would also, in general, seem to be a worthwhile enhancement.

8. Over the years I have been struck by how primitive and ritualized the analysis of RNG data has become. RNG experiments produce a very large amount of data, and many scientists (especially certain kinds of engineers) have an impressive armamentarium of tools for assessing and interpreting influences on signals. The use of terminal standardized statistics, as in the current paper, seems to waste a valuable opportunity for more informative analyses. I fully concede, however, that in the Mason et al. article the main results are so simple and compelling that probably nothing more elaborate is needed to make the points that they emphasize in their article. Nonetheless, the linear departures from randomness that they have found, while supportive of a constant $p_i$ model, never test that model. Thus there seems to remain, at this late stage, room for fundamental analyses of how $p_i$ values fluctuate over time.

9. Mason et al. very properly examine some potential influences on RNGs from known sources, such as electromagnetic fluctuations. This is a definite step forward in RNG research, and one that should be investigated more systematically.

10. As a tiny terminological quibble, I would suggest that RNG researchers stop referring to "unconscious" mental effects on RNGs. The subjects of these studies are always fully conscious, but the phenomenon referred to is supposed to be below their level of conscious perception; that is, it is "subconscious".

Suggestions

Finally, in the spirit of supporting research into world event influences on physical RNGs, I would offer some personal recommendations for future research.

- Abandon the path of trying to mount ever more statistically significant results to prove that the RNG phenomenon exists. Although the "odds
against chance" may be steadily rising over time, they are already high enough, and we are not learning anything more in the process.

- Efforts need to be made to understand how physical random numbers are generated and to develop some hypotheses about what kinds of influences might cause them to shift. Up to now RNGs have mostly been black boxes, which doesn't push the science very far forward. Do temperature, pressure, ambient light, variation in the electromagnetic field, or the force of gravity, or any of a host of other physical characteristics produce an influence? Do RNGs "age" in some way (as Mason et al. suggest), and if so, how do we take that into account? The evidence so far seems fragmentary; we need it to become systematic.

- The experimental designs need to be much stronger. Simultaneous controls, duplicate RNGs, and balancing the use of different RNGs over the experimental design (to take out the effects of an idiosyncratic RNG, for example) are all strongly indicated. There seems to be a rather large amount of ad-hoc-ness to many of the RNG experiments.

- Search for something that does a better job of capturing whatever influence we are seeing. Ultimately one would like to measure it over shorter periods of time, to relate it to changing conditions, and to therefore study its properties. If 3 hours are necessary to even see whether an influence is present, it is going to continue to be difficult to do interesting experiments.

- What kinds of human events influence RNGs? Do they have to be spiritual or "alternative" in some sense, or would one see a bigger effect, say, at a political rally, or a marriage counseling session? Some systematic research along this line might be useful.

- Research physical RNGs should simply put out the physical signal, and not somehow pre-process it in hardware. Hardware xor-ing may have done considerable damage to RNG research, and one can only wish that RNG researchers had been more critical of it before the Mason et al. article.

- It would seem to be useful to have more people involved in RNG research. By standards in other areas, RNG research is very inexpensive, and reasonably easy to do. Seeing results from more groups, in a variety of settings and with a variety of approaches, might be a very good thing.

MIKEL AICKIN
maickin@comcast.net
Statistical Consequences of Data Selection

Y. H. DOBYNS
Princeton Engineering Anomalies Research School of Engineering and Applied Science, Princeton University, Princeton NJ 08544-5263

Abstract — Data selection can result from unconscious biases or preferences on the part of experimenters, or from deliberate efforts to skew the apparent character of an experimental database. In either case the same formalism can be applied to compute the statistical signature of the selection process. Since the result of a suitably chosen selection process can be arbitrarily close to any desired distribution of experimental outcomes, it also is necessary to take into account the fraction of data that would have had to be discarded. When the selection formalism is applied to the Princeton Engineering Anomalies Research (PEAR) benchmark random event generator (REG) database, it is found that no selection model examined is consistent with the data. An unusual subset of these data, produced by a single operator, which has in the past been the target of suspicion, is likewise inconsistent with any selection hypothesis, even under a worst-case scenario of deliberate fraud.

Keywords: Statistical Methods — Meta-Analysis — REG — Human-Machine Interaction — Consciousness-Related Anomalies — PEAR

Introduction

"Data selection” is a common term for the biased selective reporting of data in scientific research. It is frequently invoked as a dismissive explanation for peculiar or anomalous results. While it usually implies a deliberate attempt to deceive, data selection also can result from unconscious biases of experimenters or inadequate controls in the recording and reporting of data (Gould, 1996). The work presented here explores the possibility that the anomalous effects seen in the benchmark random event generator (REG) database generated at the Princeton Engineering Anomalies Research (PEAR) laboratory (Jahn et al., 1997) might be due to such a selection process. None of the selection models examined are consistent with the data, and some general properties of selection models suggest that no possible model of this class can be constructed to be consistent with the data.

Premises and Definitions

The formalism developed here assumes that the data under consideration accrete in the form of samples from a standard normal distribution, each
individual sample being the final outcome of a single human action to generate data (e.g., pressing a button to start the apparatus). While the only experimental data considered here come from the PEAR program's REG experiments, the normal distribution is ubiquitous enough that it may be hoped the formalism and general arguments have a broader application. In the case of the PEAR REG data, the human action is the initiation of data generation for a single "run," and the resulting standard normal deviate is simply the mean score for that run, as normalized by the expected mean and standard deviation. That is, $x = (m - \mu)/\sigma$, where $x$ is the normalized outcome, $m$ the observed mean, $\mu$ the theoretical mean, and $\sigma$ the theoretical standard deviation.

REG run scores are in fact binomially rather than normally distributed, but since a single REG run involves a minimum of 10,000 bits (and may involve as many as 200,000), the deviations from normality are inconsequential. For analytical purposes, all REG data will be treated here at this level of normalized run deviations. The observed anomalous effect in the PEAR REG experiments, as reported elsewhere (Jahn et al., 1997), is a shift in these mean run scores, correlated with the operator's pre-stated intention.

These experiments involved a tripolar protocol, in which approximately equal amounts of data were generated under three intentional conditions, high, low, and baseline. In the high intention, the operator's goal was to increase the mean value of the data; in the low intention, to decrease it. (The baseline was a passive intentional condition in which the operator was not directed to make any effort.) The only distinction between the two intentional conditions is the direction of effort; any formal analysis whatsoever applies equally to both intentions, up to a sign change. The following discussion, therefore, is written as applying only to the high intention, with "positive" outcomes or shifts being in the direction of intention, and "negative" outcomes or results contrary to the intention. The reader should bear in mind that exactly the same formalism, up to a sign reversal, applies to the low intention.

Because of this theoretical symmetry, all subsequent comparisons between theoretical predictions and data will pool the results from both intentional conditions. A single distribution of intentional outcomes is computed by inverting the sign of all low-intention deviations and combining these inverted results with the high-intention deviations to construct a single population of deviations in the direction of intention.

We will consider two genres of explanatory model for the observed anomalies. The hypothesis that these data reflect an actual change in the machine's operation is referred to as a "mean shift model." In contrast, a "selection model" presumes that certain runs are selectively discarded, depending on their values, so as to bias the distribution statistics of the retained runs and create the spurious appearance of a mean shift. This altered distribution of runs will be called the selected distribution. The undistorted distribution existing before selection will be referred to as the source distribution. It is assumed throughout that, since the selection model is an alternative to an
anomalous change of the distribution, the source distribution is the undisturbed, standard-normal-distributed output of the apparatus. In general, the selection process is probabilistic (reject a fraction of runs of a given value); it can be made deterministic by setting the selection probability for a given value at 0 or 1.

The symbol $p(x)$ will be used to refer to the selected distribution, where $x$ is the run value. The functional notation $f(x)$ will be used to denote the standard normal probability distribution, i.e. $f(x) \equiv \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. $F(x)$ refers to the antiderivative of $f$, the cumulative normal probability function: $F(x) = \int_{-\infty}^{x} f(t)dt$.

By hypothesis, the selection process produces $p(x)$ from a source distribution $f(x)$. The probability that a run of value $x$ will be retained is given by a selection function $s(x)$. Since it is a probability, $0 < s(x) < 1$ for any $x$. In general the total integral $S = \int_{-\infty}^{\infty} s(x)f(x)dx$ will be less than 1 (the only exception being the trivial "selection" function $s(x) = 1$). Since $p(x)$ should be a properly normalized probability distribution, its value must be $p(x) = s(x)f(x)/S$.

The description of a selection process entails that a certain fraction of the original data has been discarded. For purposes of treating this "filedrawer" quantitatively, we will define the filedrawer quotient $Q$ as the ratio of the amount of discarded data to the amount of data retained. Since the quantity $S$, defined above, gives the total integral of the selected distribution relative to the original, $Q = (1 - S)/S$.

Accusations of data selection frequently invoke an intuitive but generally incorrect rule of thumb: since the mean shift is a very small fraction of the observed mean value, one supposes that it could be attained by discarding a comparably small fraction of negative data, with the implication that bias or carelessness easily could cause a few critical runs quietly to disappear. The commonest form of this criticism asserts that, since PEAR’s effect amounts to $1$ part in $10^4$, discarding $1$ part in $10^4$ of contrary data would be adequate to produce it. Nevertheless, the explicit calculation of filedrawer values will show that, despite the small absolute scale of the effect, the amount of discarded data required to produce it usually would be quite substantial. In some extreme cases, the discarded data would need to be comparable in quantity to the entire reported database. Even in cases where the fraction of discarded data is modest, it is orders of magnitude larger than the naive criticism would suggest, and entails a remarkably large aggregate of actual instances of discarded data due to the large sizes of the observed databases.

**General and Limitative Results**

While an arbitrarily chosen selection function $s(x)$ is likely to produce a distorted distribution, it is natural to speculate whether this is necessarily so for all such functions. The answer is no: the output of a selection process on a normal source distribution can produce a shifted distribution that is also normal.

Let $f(x, \sigma) = (1/\sqrt{2\pi\sigma})e^{-x^2/2\sigma^2}$ be the generalization of $f(x)$ to non-unit variance. $f(x, a)$ is a proper probability distribution with total integral 1, for any
Suppose that a selection process applied to standard normal input with distribution \( f(x) \) produces output distributed according to \( f(x - \mu, \sigma) \) for some positive \( \mu \) (since the topic of interest is the production of spurious positive mean shifts via selection) and some \( \sigma \) not necessarily equal to 1. This implies that

\[
Sf(x - \mu, \sigma) = s(x)f(x),
\]

since the total integral of \( s(x)f(x) \) is \( S \) by definition and the integral of \( f(x - \mu, \sigma) \) is 1 for any \( \mu, \sigma \). It then follows that \( s(x) = Sf(x - \mu, \sigma)/f(x) \), which by substitution of the explicit forms for \( f \) can be seen to be

\[
s(x) = \frac{S}{\sigma} \exp\left(\frac{1}{2\sigma^2}[(\sigma^2 - 1)x^2 + 2\mu x - \mu^2]\right).
\]

Three distinct cases can be distinguished in this formula according to the value of \( \sigma \).

- If \( \sigma^2 > 1 \), the leading dependence on \( x \) is as \( e^{cx^2} \) with positive coefficient \( c \). This grows without limit for large \( x \). However, \( s(x) \) cannot exceed 1 for any \( x \). Therefore this is not a possible case. Selection cannot produce a normal distribution with greater variance than its source; any selected distribution with increased variance must show some departures from normality.

- If \( \sigma^2 = 1 \), the leading dependence on \( x \) is \( e^{\mu x} \). This also grows without limit for increasing positive \( x \), and is therefore again an impossible case. The selected distribution, if normality is preserved, cannot have the same variance as its source.

- If \( \sigma^2 < 1 \), the polynomial in the exponent, and hence the ratio itself, has a maximum at \( x = \mu/(1 - \sigma^2) \). This is a possible situation; thus, a selection process can produce an undistorted normal function as its output, provided the variance of the output normal distribution is less than that of its source.

This result can also be used to derive a relation among the variables \( S, \mu, \) and \( \sigma \), on the assumption that \( s(x) = 1 \) at its maximum. This gives the largest possible value of \( S \), and hence the smallest possible value of \( Q = (1 - S)/S \), the filedrawer quotient. Solving the equation \( Sf(x - \mu, \sigma) = f(x) \) at \( x = \mu/(1 - \sigma^2) \) gives, after some algebra, the relation

\[
S = \sigma \exp\left[-\frac{1}{2} \left(\frac{\mu^2}{1 - \sigma^2}\right)\right].
\]

It is obvious that \( S \) vanishes at \( \sigma = 1 \), as the argument of the exponential goes to \(-\infty\). It is equally obvious that \( S \) vanishes at \( \sigma = 0 \). The derivatives of \( S \) in the region between are

\[
\frac{\partial S}{\partial \mu} = -\frac{\mu}{1 - \sigma^2}S; \quad \frac{\partial S}{\partial \sigma} = S\left[\frac{1}{\sigma} - \frac{\sigma \mu^2}{(1 - \sigma^2)^2}\right].
\]
Since $\mu$ is positive by assumption, and $S$ is nonnegative, it is obvious that $\partial S/\partial \mu < 0$ anywhere $S$ is nonzero. This accords with intuition: the larger the shift in the distribution mean, all else being equal, the more stringent the selection must be. Setting $\partial S/\partial \sigma = 0$ and applying the quadratic formula shows that $S$ has exactly one maximum in $a$ in the allowed range $0 < a < 1$, at

$$\sigma_{\text{max}}^2 = 1 + \frac{\mu}{2} \left( \mu - \sqrt{4 + \mu^2} \right).$$

This again accords with intuition: if $a$ is too small, much of the source distribution must be discarded in order to make the selected distribution narrow enough. If $a$ is too large, much of the source distribution must be discarded to match the limit imposed by available data in the upper tail.

For a general consideration of what is possible with selection, the important points to note from these formulae are:

1. For any given value of $a$ in the selected distribution, increasing $\mu$ requires a decrease in $S$.
2. For any particular value of $\mu$ in the selected distribution, there is an optimal value of $a$ that maximizes $S$. As $\mu$ increases, the optimal value of $a$ decreases, as does $S$ at that value of $a$.
3. Any attempt to increase $a$ beyond the value that maximizes $S$ leads to a rapid decrease in $S$, since $S = 0$ at $a = 1$ is required.

Point 3 in particular means that there is an unavoidable tradeoff between the degree of distortion from the original distribution, and the amount of data discarded to achieve it; the more nearly the selected distribution approaches an otherwise undistorted, mean-shifted version of the source, the smaller the fraction of the source is being retained.

The discussion above proves that, while a selection process can produce a shifted but still normal selected distribution from a normal source distribution, the output distribution necessarily has smaller variance than the source. Moreover, the larger the mean shift, and/or the more closely the selected distribution approaches the source distribution's variance, the more data must be discarded in the selection process.

A more general selection process that is not constrained by producing normal output need not obey this variance rule: for example, if all data smaller than a certain absolute value are discarded, the selected distribution will have greatly increased variance. It will also be bimodal and hence grossly non-normal. Given the above result for normal output, in which all properties of the selected distribution except the variance are fixed (since the target mean shift is chosen in advance), it seems reasonable that more general selection processes must be subject to a similar tradeoff between the intensity of selection and the degree of departure from normality in the output. This conjecture will be revisited after some development and examination of specific selection models in the following sections.
Some Plausible Selection Models

The selection function $s(x)$ can vary arbitrarily at every $x$, subject only to the constraint $0 \leq s(x) \leq 1$ for all $x$. However, a real selection process is not likely to employ an arbitrarily complicated selection function. The selection models discussed below attempt to sample a reasonably broad range of the space of practical and plausible models.

Each heading below describes a one-parameter family of models, where the mean and all other distribution statistics are determined by a single free parameter. The reason for examining one-parameter families is that once the observed mean is fit by choosing an appropriate value of the free parameter, the remaining distribution statistics have been fixed; this places each model on a footing comparable to the one-parameter mean shift model. We adopt the convention of calling the selection parameter $a$ in all cases. The functional forms of the various models are illustrated graphically in Figure 1.

- **Simple Cutoff.** The simplest possible data selection model is simply to reject all runs with a value less than some fixed cutoff $a$; $s(x) = 0$ for $x < a$, 1 otherwise. This model bears some theoretical importance in that it is provably the model with the smallest filedrawer quotient for any given mean shift. (The proof is almost trivial: it rejects all those and only those data with the greatest negative contribution to the mean. Changing the rule in any way will therefore reduce the mean shift generated per discarded data point, and thus require that more data be discarded to achieve the same mean shift.) The fact that the cutoff model also produces the greatest departures from normal distribution statistics of any model examined is therefore strong support for the general "tradeoff" thesis of the preceding section.

- **Fractional Rejection.** It seems overly simplistic to expect that the simple cutoff model ever would appear in actual data. Only the most naïve of frauds could imagine that the total disappearance of runs below some set value could be invisible; even the most biased of researchers would need phenomenal powers of self-delusion to convince themselves that all runs, and only those runs, lying below a particular value were methodologically invalid. The fractional rejection model supposes that a constant fraction $a$ of all unsuccessful ($x < 0$) runs are discarded:

  $$s(x) = 1 - a$$

  for $x < 0$, 1 otherwise.

- **Short Left Tail.** For a further increase in psychological plausibility, we can suppose that an increasing rate of exclusion is employed as the value of the run goes more negative. For a simple representation of this possibility, a short left tail selected distribution compresses the variance for all negative values of $x$. In other words, for some compression factor $a < 1$,
Statistical Consequences of Data Selection

\[ p(x) = \frac{2}{1 + a} \times \begin{cases} f(x), & \text{for } x > 0; \\ f(x/a), & \text{for } x < 0. \end{cases} \]

where \( p(x) \) has been renormalized to a proper probability distribution.

The selection function for negative \( x \) is

\[ s(x) = \frac{f(x/a)}{f(x)} = \exp\left\{ \frac{x^2}{2} (1 - \left( \frac{1}{\tau^2} \right)) \right\}; \text{ as before } s(x) = 1 \text{ for } x > 0. \]

**Increasing Bias.** Rather than presuming that only negative runs will be rejected, we may suppose that an experimenter preference for positive results may manifest itself as a graduated probability of rejection. For this model we presume that runs that exceed the positive one-tailed \( p = 0.05 \) significance criterion \( x = 1.645 \) are always retained, while runs with \( x < -1.645 \) are rejected with probability \( a \). In the region \(-1.645 < x < 1.645\), the rejection probability drops linearly from \( a \) to 0 as \( x \) increases.

Thus

\[ s(x) = \begin{cases} 1 - a, & \text{for } x < -1.645; \\ 1 - a/2 + ax/3.29, & \text{for } -1.645 < x < 1.645; \\ 1, & \text{for } x > 1.645. \end{cases} \]

This enumeration is not intended to imply that unconscious bias would work by any such explicit or exact means as the example selection functions. Indeed, it seems virtually certain that data selection due to unconscious bias would operate in a messy and inexact fashion driven by a host of psychological considerations. The intent of employing these several families of models, to the contrary, is to provide functional forms that are relatively amenable to calculations, for models that capture reasonable qualitative properties of the unconscious bias process. Taking the last as an example, it seems reasonable to suppose that a biased experimenter would be most eager to retain significant runs, and most willing to discard significantly negative runs, with an intermediate level of preference for intermediate cases. While it would be ridiculous to propose that such an experimenter would unconsciously be calculating the three-part \( s(x) \) function given above, this \( s(x) \) gives a computable quantification of the specified qualitative features; we therefore reasonably may argue that any psychological process fitting the given qualitative description should produce output statistics similar to those of our exemplar, chosen for its computational convenience.

Figure 1 displays examples of all of these functions. The uppermost plot shows the standard normal distribution in the interval \([-3, 3]\). The next plot is a mean-shifted normal, corresponding to the mean shift model which is our representation for a genuine anomalous effect. The mean of the distribution has been set at 0.3, a size appropriate to the scale of effect in some of the more successful PEAR datasets. The third plot shows the cutoff distribution with the same mean. (All of the selected distributions in Figure 1 have been normalized to have the same area as the shifted normal with which they are compared.) The next three plots show
the fractional rejection model, the short tail model, and the increasing bias model, all for the same mean of 0.3. Finally, the last plot shows all four selection models superimposed as dotted lines on the shifted normal with the same mean.

Comparison Methods and Statistical Power

Most selection processes produce non-normal selected distributions, and even the normality-preserving selection process produces a selected distribu-
Statistical Consequences of Data Selection

Table 1: Statistical Power Comparison — Required N

<table>
<thead>
<tr>
<th>Selection model</th>
<th>$\chi^2$ Goodness-of-fit</th>
<th>$\sigma$</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple cutoff</td>
<td>192</td>
<td>30</td>
<td>44</td>
<td>1.47 x $10^5$</td>
</tr>
<tr>
<td>Fractional rejection</td>
<td>1920</td>
<td>638</td>
<td>203</td>
<td>460</td>
</tr>
<tr>
<td>Short tail</td>
<td>1040</td>
<td>43</td>
<td>127</td>
<td>7551</td>
</tr>
<tr>
<td>Increasing bias</td>
<td>2.28 x $10^5$</td>
<td>638</td>
<td>3510</td>
<td>5870</td>
</tr>
</tbody>
</table>

Table 1 shows the statistical power of tests on the higher moments, for the same effect size 0.3 shown in Figure 1, by giving the number of data points required for an $\alpha = 0.05$, $\beta = 0.50$ test of each of the above selection models against the corresponding mean shift model. That is, if the mean shift model for $\mu = 0.3$ is taken as the null hypothesis, and the data actually are being generated by one of the selection models with a chosen to replicate the same value of $\mu$, Table 1 reports the N at which the expected value of the given test statistic is at the $p = 0.05$ significance level ($\alpha = 0.05$).* This corresponds to the point where, if the null hypothesis is false, the test is equally likely to produce results which are significant or nonsignificant under the aforementioned criterion, leading to $\beta = 0.5$ (probability of erroneously accepting a false null hypothesis). It should be noted that $\beta = 0.5$ is an unsatisfactorily high probability of Type II error; it is used here not prescriptively but for convenience of calculation, since Table 1 is intended solely to demonstrate the relative sensitivity of different tests for detecting the various selection models.

From Table 1 we see several features important to the identification of the most sensitive test:

* When the test statistic is a moment parameter, its expected value is found by applying Equations 1–5, given in the next section. The expected value of the $\chi^2$ is found by direct computation of the difference between each of the selected distributions and a normal distribution with the same mean and unit variance.
At least one of the three central-moment tests always outperforms the
distribution-based $\chi^2$ by a large margin, as would be expected from the
analysis in the Appendix.

In three of the four cases, the most sensitive moment test is on the standard
deviation $\sigma$; in the remaining case it is on skewness. The kurtosis test is never
the most sensitive.

The most sensitive moment test, for each selection model, requires appreciably
less data than the second-best test; the difference is sometimes considerable.

A feature which does not appear in Table 1 is the fact that the identity of the
most sensitive moment parameter depends on the scale of the effect. For example,
in Table 1 the cutoff model can most readily be detected by its change in the
standard deviation; however, for a mean shift of 0.03, an order of magnitude
smaller than that used in Table 1 (and more typical of general PEAR databases),
the most sensitive test for the cutoff model becomes the skewness, $\gamma_3$. In light of
this variability, the best procedure for testing whether data are consistent with
a particular selection model would seem to be to compute the statistical power
of each test for the given effect size and use the most sensitive.

**Selection Model Moment Parameters**

Both the statistical power calculations discussed in the previous sections and
actual comparisons with empirical data require that we calculate the higher
moments of a selected distribution from its mean shift. All of the selection rules
allow the mean and higher moments to be calculated from a given $a$; although
the mean shift in terms of $a$ is seldom given as an invertible function, procedures
such as a numerical binary search readily can be used to find the $a$ that
corresponds to a desired mean shift, and the higher moments then can be
calculated directly. Figure 2 illustrates the results of such a process for the cutoff
distribution, with the cutoff parameter, standard deviation, skewness, and
kurtosis presented as functions of the mean shift.

The functional forms of the mean and higher moments of each selection
model, as functions of the free parameter $a$, can be calculated by straightforward
integrations. For example, for a given cutoff parameter $a$ the cutoff distribution
has the following distribution moments and filedrawer parameters:

- Mean: $m = f(a)/(1 - F(a)) = f(a)/F(-a)$
- Variance: $\sigma^2 = 1 + ma - m^2$
- Skewness: $\gamma_3 = \frac{2m^3 - 3am^2 + (a^2 - 1)m}{\sigma^3}$
- Kurtosis: $\gamma_4 = \frac{3 + m(a^3 + 3a) - m^2(4a^2 + 2) + 6m^3a - 3m^4}{\sigma^4}$

(1)

- Filedrawer: $S = 1 - F(a)$; $Q = \frac{F(a)}{1 - F(a)}$. 
It can be seen that, even for this relatively simple function, the functional form of the higher central moments becomes somewhat involved. For the remaining distributions it will be somewhat more straightforward to report their moments $\langle x^2 \rangle, \langle x^3 \rangle, \langle x^4 \rangle$ rather than the statistical parameters, which are determined by the central moments. That is, for any distribution whatsoever, the variance, skewness, and kurtosis are given by:

$$
\sigma^2 = \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2
$$

$$
\gamma_3 = \frac{\langle (x - \langle x \rangle)^3 \rangle}{\sigma^3} = \frac{1}{\sigma^3} \left[ \langle x^3 \rangle - 3\langle x \rangle \langle x^2 \rangle + 2\langle x \rangle^3 \right]
$$

$$
\gamma_4 = \frac{\langle (x - \langle x \rangle)^4 \rangle}{\sigma^4} = \frac{1}{\sigma^4} \left[ \langle x^4 \rangle - 4\langle x \rangle \langle x^3 \rangle + 6\langle x \rangle^2 \langle x^2 \rangle - 3\langle x \rangle^4 \right].
$$

The standard normal distribution has $\langle x \rangle = 0$, $\langle x^2 \rangle = 1$, $\langle x^3 \rangle = 0$, and $\langle x^4 \rangle = 3$, hence its expected values of mean, standard deviation, skewness, and kurtosis are 0, 1, 0, and 3, respectively.

With Equations 2 in place, we may describe the parameters of the other distributions somewhat more concisely. For the fractional rejection distribution, described by the rejection rate $a$, the first four moments and filedrawer quotient are given by:
For the short-left-tail distribution, described in terms of the contraction factor $a$, the corresponding values are

\[
\langle x \rangle = \frac{af(0)}{1 - a/2}, \quad \langle x^2 \rangle = 1 \tag{3}
\]

\[
\langle x^3 \rangle = \frac{2af(0)}{1 - a/2}, \quad \langle x^4 \rangle = 3 \tag{4}
\]

\[
Q = \frac{a}{2 - a}.
\]

For the short-left-tail distribution, described in terms of the contraction factor $a$, the corresponding values are

\[
\langle x \rangle = 2f(0)(1 - a), \quad \langle x^2 \rangle = 1 - a + a^2 \tag{1}
\]

\[
\langle x^3 \rangle = 4f(0)(1 - a + a^2 - a^3), \quad \langle x^4 \rangle = 3(1 - a + a^2 - a^3 + a^4) \tag{2}
\]

\[
Q = \frac{1 - a}{1 + a}.
\]

And finally, for the increasing bias distribution, if for conciseness we define $b = 1.645$ for the transition points of $s(x)$, the moments and filedrawer quotient are

\[
\langle x \rangle = \frac{a}{\sqrt{2}} \left( 2F(b) - 1 \right), \quad \langle x^2 \rangle = 1 \tag{5}
\]

\[
\langle x^3 \rangle = \frac{a}{2 - a} \left( 6F(b) - 2af(b) - 3 \right), \quad \langle x^4 \rangle = 3
\]

\[
Q = \frac{a}{2 - a}.
\]

**Relation to General Results**

Figure 3 applies Equations 1–5 to illustrate the ease of detection, and the filedrawer parameter required, for the four models as the mean shift $\mu$ is increased from 0 to 0.5. The "distortion index" plotted in the top figure is the rate of growth with $N$ (the number of observed data points) of a $Z$-score describing the departure of the most sensitive parameter from its theoretical value. (When $N$ is large enough for the usual normal approximations to the variation in $\gamma_3$, and $\gamma_4$ to be useful, then the distortion index, multiplied by \sqrt{N}, gives the expected value of $|Z|$ for the most sensitive statistic.) The bottom graph simply plots the filedrawer quotient for that model at that mean shift.

It is conspicuous from the figure that the models, despite having grossly different distributions and being best detected by different distribution parameters, obey a generalized form of the results rigorously derived for normality-preserving selection functions. As $\mu$ increases, both the distortion and the filedrawer of any particular model increase monotonically. At every $\mu$, the order of models ranked by increasing distortion is exactly the reverse of their order ranked by increasing filedrawer; changing to a model with less statistical distortion always increases the filedrawer quotient.
**REG Data**

The PEAR REG experiment involved the collection of data from a device with no known non-anomalous channels for the operator's influence. While full
details of the experimental protocols and controls are available elsewhere (Jahn et al., 1987; Dunne & Jahn, 1995), a brief summary may be in order.

- **Redundant Recording.** The raw data were printed on a continuous paper tape, and concurrently entered into a computer file. Summary data were also recorded by the operator in a logbook. In any case where a discrepancy appeared among the three records, the paper tape record was given precedence.

- **Advance Designation of Intention.** The operator was required to declare an intention before data were generated. Data collection could not be initiated until the intention was entered and logged in the experimental computer.

- **Continuity of Record.** The paper tape record was required to be continuous, without gaps or breaks, as a safeguard against precisely the sort of selection discussed in this analysis. As an aside, it should be noted that this requirement of physical integrity of the paper tape and the primacy of the tape record over the other redundant records in case of disagreement also provided strong safeguards against the alteration of extant data or the introduction of spurious data. Such interventions would be considerably more difficult than the already challenging task of making data disappear from the records.

Data were collected in runs of 50, 100, or 1000 trials, where one trial is the sum of successes in 200 binary $p = 0.5$ events. If spurious data selection were attempted, the finest possible scale of intervention would have been the run level; picking and choosing what data to retain at the level of individual trials would require a massive invasion of all the data recording systems, both hardcopy and electronic, and involve far more labor than fabrication of the entire database from whole cloth.

The three different run lengths mentioned above must be treated separately in order to discriminate properly between selection and mean shift models. The reason is that a mean shift model, in the absence of additional qualifying hypotheses, predicts an effect that is constant at the trial level, and therefore predicts that the average Z-scores of runs depend on how many trials comprise them. Therefore, a mixture of Z-scores for runs of different lengths would be an intrinsically heterogeneous database under one of the hypotheses being compared, rendering all statistical comparisons suspect.

Table 2 displays the statistics for the three run lengths present in the primary REG database. The two active intentions have been combined, with the sign of deviations in the low intention reversed to produce a uniform measure of deviation in the direction of intention. The statistical uncertainty ($\sigma$) for each parameter also is given, along with the Z-score for its deviation from the expected value. Note that all of the Z-scores for higher moments are nonsignificant, indicating that the data are at least consistent with a mean shift model.

In accordance with the discussions in "Comparison Methods and Statistical Power", we may proceed now to calculate the statistical power of testing the various moments of the selection models on each of these databases. Table 1 presented a minimum N for having a probability $\beta = 0.5$ of failing to distinguish
Statistical Consequences of Data Selection

TABLE 2
REG Run Distributions

<table>
<thead>
<tr>
<th>Run length</th>
<th>N_{\text{runs}}</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>12,849</td>
<td>0.0277 ± 0.0088</td>
<td>1.0077 ± 0.0062</td>
<td>0.0085 ± 0.0216</td>
<td>2.9553 ± 0.0432</td>
</tr>
<tr>
<td></td>
<td>Z = 3.1369</td>
<td>Z = 1.2316</td>
<td>Z = 0.3939</td>
<td>Z = -1.0344</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>3340</td>
<td>0.0402 ± 0.0173</td>
<td>0.9932 ± 0.0122</td>
<td>-0.0371 ± 0.0424</td>
<td>2.9514 ± 0.0848</td>
</tr>
<tr>
<td></td>
<td>Z = 2.3252</td>
<td>Z = -0.5593</td>
<td>Z = -0.8763</td>
<td>Z = -0.5732</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>700</td>
<td>0.0485 ± 0.0378</td>
<td>0.9881 ± 0.0267</td>
<td>-0.0052 ± 0.0926</td>
<td>3.2692 ± 0.1852</td>
</tr>
<tr>
<td></td>
<td>Z = 1.2828</td>
<td>Z = -0.4449</td>
<td>Z = -0.0561</td>
<td>Z = 1.4536</td>
<td></td>
</tr>
</tbody>
</table>

The statistical power for rejecting the mean shift hypothesis is assumed. Unfortunately, it also is clear that, for these effect sizes and database sizes, the statistical power is too low to distinguish some of the models from the mean shift model. In particular, the lowest β value that appears for the increasing bias model is 0.923; that is, even with the most sensitive test on the best database for the purpose, there is a 92.3% likelihood that a database actually produced by the increasing bias process would fail to produce a test statistic significantly different from the expected value for a mean shift model.

Table 4 lists the predictions for the various models on these datasets. In each case the single free parameter of the model is used to fit the observed mean; the standard deviation, skewness, and kurtosis then follow from the functional form of the model. In addition to the model predictions for these parameters, Table 4 gives the Z-scores for the empirical value of the given parameter (as reported in the undisturbed mean shift hypothesis). Since the amount of formal data in hand is fixed and cannot be modified, Table 3 instead presents β for each parameter test, on each database, where a standard α = 0.05 criterion for rejecting the mean shift hypothesis is assumed.

The listing of β in Table 3 values allows easy identification of the most powerful test for detecting the presence of each selection model in a given dataset: simply choose the parameter with the smallest β value for that model in that dataset. Unfortunately it also is clear that, for these effect sizes and database sizes, the statistical power is too low to distinguish some of the models from the mean shift model. In particular, the lowest β value that appears for the increasing bias model is 0.923; that is, even with the most sensitive test on the best database for the purpose, there is a 92.3% likelihood that a database actually produced by the increasing bias process would fail to produce a test statistic significantly different from the expected value for a mean shift model.

Table 4 lists the predictions for the various models on these datasets. In each case the single free parameter of the model is used to fit the observed mean; the standard deviation, skewness, and kurtosis then follow from the functional form of the model. In addition to the model predictions for these parameters, Table 4 gives the Z-scores for the empirical value of the given parameter (as reported in
TABLE 4
Model Predictions

<table>
<thead>
<tr>
<th>Model</th>
<th>σ</th>
<th>Z(σ)</th>
<th>Skewness (γ3)</th>
<th>Kurtosis (γ4)</th>
<th>Z(γ4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50-trial runs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean shift</td>
<td>1.000</td>
<td>1.2316</td>
<td>0.0000</td>
<td>0.3939</td>
<td>3.0000</td>
</tr>
<tr>
<td>Cutoff</td>
<td>0.9670</td>
<td>6.5141</td>
<td>0.1393</td>
<td>−6.0508</td>
<td>2.7973</td>
</tr>
<tr>
<td>Fractional rejection</td>
<td>0.9996</td>
<td>1.2930</td>
<td>−0.0277</td>
<td>1.6740</td>
<td>3.0031</td>
</tr>
<tr>
<td>Short tail</td>
<td>0.9827</td>
<td>4.0005</td>
<td>0.0282</td>
<td>−0.9091</td>
<td>3.0006</td>
</tr>
<tr>
<td>Increasing bias</td>
<td>0.9996</td>
<td>1.2930</td>
<td>−0.0104</td>
<td>0.8752</td>
<td>3.0012</td>
</tr>
<tr>
<td>100-trial runs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean shift</td>
<td>1.000</td>
<td>−0.5593</td>
<td>0.0000</td>
<td>−0.8763</td>
<td>3.0000</td>
</tr>
<tr>
<td>Cutoff</td>
<td>0.9549</td>
<td>3.1240</td>
<td>0.1794</td>
<td>−5.1088</td>
<td>2.7743</td>
</tr>
<tr>
<td>Fractional rejection</td>
<td>0.9992</td>
<td>−0.4931</td>
<td>−0.0402</td>
<td>0.0722</td>
<td>3.0065</td>
</tr>
<tr>
<td>Short tail</td>
<td>0.9749</td>
<td>1.4893</td>
<td>0.0413</td>
<td>−1.8497</td>
<td>3.0012</td>
</tr>
<tr>
<td>Increasing bias</td>
<td>0.9992</td>
<td>−0.4931</td>
<td>−0.0151</td>
<td>−0.5206</td>
<td>3.0024</td>
</tr>
<tr>
<td>1000-trial runs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean shift</td>
<td>1.000</td>
<td>−0.4449</td>
<td>0.0000</td>
<td>−0.0561</td>
<td>3.0000</td>
</tr>
<tr>
<td>Cutoff</td>
<td>0.9474</td>
<td>1.5217</td>
<td>0.2029</td>
<td>−2.2479</td>
<td>2.7655</td>
</tr>
<tr>
<td>Fractional rejection</td>
<td>0.9988</td>
<td>−0.4009</td>
<td>−0.0484</td>
<td>0.4669</td>
<td>3.0094</td>
</tr>
<tr>
<td>Short tail</td>
<td>0.9698</td>
<td>0.6838</td>
<td>0.0500</td>
<td>−0.5959</td>
<td>3.0018</td>
</tr>
<tr>
<td>Increasing bias</td>
<td>0.9988</td>
<td>−0.4009</td>
<td>−0.0181</td>
<td>0.1395</td>
<td>3.0035</td>
</tr>
</tbody>
</table>

Table 1) relative to the model prediction. It is divided into three sections, for the three different datasets.

It is clear that the cutoff model is rejected for all three datasets. The short tail model is also strongly rejected for the largest dataset, that of 50-trial runs. Both the fractional rejection model and the increasing bias model are consistent with the statistics of the actual data, just as the mean shift model is. This ambiguous result is only to be expected, given the β values listed in Table 3, where the most sensitive tests have β = 0.751 for the fractional rejection model and β = 0.923 for the increasing bias model. We simply do not have enough data to distinguish these models reliably from the mean shift model on the strength of any statistical parameter of the distribution. Therefore, the resolution of the possibility of data selection must turn, not on the statistical parameters of the formal data distributions, but on the filedrawer quotients describing the amount of absent data.

Missing Data: Void Runs

Aside from the published data reported in Table 2, some data have of course been discarded. The formal protocol, although it has changed over time, always has mandated the invalidity of data collected under certain protocol-violating conditions. To the greatest extent possible, these void criteria have been designed with the intent of eliminating the human decision factor, and therefore the possibility of biased preferences.
The standard protocol requires that there be some record of the existence of every occasion on which formal data were generated, or even when an attempt was made to generate formal data. Violation of this protocol condition would require a deliberate attempt to deceive; the consequences of such efforts will be discussed further below, but for the moment we are concerned only with the possible impact of bias on the decision to reject data.

In the majority of cases, the void run generated data which were recorded, and statistical summaries were computed. Generally these are cases where some protocol violation mandated that the data be considered invalid; e.g., during a period when the formal protocol required a minimum of 5 runs per session, some sessions of fewer than 5 runs were generated due to operator misunderstandings of the protocol. These perforce were declared void and excluded from the formal database. During the same period, operators were permitted to complete a series over the course of multiple laboratory visits (since a series might include up to 300 runs, requiring over 5 hours of the operator's time). Some operators never returned to complete a series, and in these cases also the experimenters were obliged to mark the data as void.

In some cases data were declared void due to an equipment malfunction of such nature as not to preclude data generation or recording. For example, there were several occasions on which runs were generated with internal (inter-trial) standard deviations of 14–18 rather than the theoretical $\sqrt{50} = 7.071$; this grossly aberrant output was taken as sufficient demonstration that the noise source had suffered a breakdown and was no longer emitting properly conditioned random values. (Indeed, in these cases physical intervention was required to restore proper operation of the device.) In other cases, individual runs were declared void due to protocol violations such as the unexpected and disruptive arrival of visitors.

In some cases, no data values were recorded for the void runs. Much of the time this was due to equipment failures that made it impossible to record data, as for example during a period when the automated data collection was handled not by a local computer but by a remote connection to a departmental server, which was prone to unpredictable downtime and sometimes failed during an experiment. Sometimes, however, these episodes were due to operator errors in the conduct of the experiment, or were caused by a problem such as a disruptive visit actually prevented the recording of data rather than merely interrupting the operator.

The formalism discussed above addresses the statistical features left behind in a population of observed and recorded data, as a consequence of the construction of that population, by discarding and concealing a selected component of the total source distribution. It is clearly fatuous to apply this technique to the population of void runs with recorded values: the impact that their removal has had on the data can be calculated directly, simply by restoring them to the experimental population. On the other hand, since at least some of the void data with known values are products of a random source known to be malfunctioning
at the time, including them as part of the data under analysis violates one of the assumptions of the formalism, namely that the output of the experimental apparatus follows a standard normal distribution.

The best resolution of the situation with the two classes of void runs would seem to be as follows. The distribution of those voids with values can be computed; it can be compared, both with the null hypothesis of zero effect and with the observed effect size in the formal data, for any evidence of bias in its removal from the database, and it can be recomputed with the formal data to establish its impact, if any, on the scale and significance of the anomalous effect. The voids without values, in contrast, comprise a population of missing data which properly should be compared with the filedrawer quotients predicted for the various selection models. These predictions should, however, be based on the mean shift and population size of the formal data alone, not the formal data recomputed with voids of known value, since these latter are not in all cases drawn from the same distribution.

Table 5 summarizes the void data present in each of the three databases. It gives the numbers of each type of void run; for the voids with values, it additionally gives the number of voids with results in and contrary to the direction of intention and the Z-score of this count imbalance. (These totals do not add to the total count of voids with values in the 50-trial runs, because six of these runs had means of exactly 100.00 which is neither in nor contrary to the direction of intention.) Also, the mean and standard deviation of the population of voids with values, the Z-score of this population against the null hypothesis, and a two-population T-score for the difference between the voids and the formal data are given. Finally, the overall Z-score for the anomalous mean shift is recomputed with the void data added to the formal data.

The uniformly negative means of the void populations suggest that, despite all efforts, some degree of bias was present in the rejection of these data. None of the three void populations differs significantly from a null hypothesis, however.

| TWO-POPULATION T-SCORE FOR THE DIFFERENCE BETWEEN THE Voids AND THE FORMAL DATA | TABLE 5 | Void Runs, by Database |
|---|---|---|---|---|---|---|
| 50-trial runs | 100-trial runs | 1000-trial runs |
| N without values | 184 | 1 | 3 |
| N with values | 590 | 80 | 24 |
| ... Matching intention | 277 | 44 | 9 |
| ... Against intention | 307 | 36 | 15 |
| Z of count imbalance | -1.2414 | 0.0844 | -1.2747 |
| Mean Z of voids | -0.0422 | -0.0080 | -0.2759 |
| σ | 1.0504 | 1.0604 | 0.7319 |
| Z vs. null | -1.0247 | -0.0715 | -1.3483 |
| T vs. formal | -1.5824 | -1.0021 | -2.1020 |
| Z, formal only | 3.1369 | 2.3252 | 1.2828 |
| Z, formal + void | 2.8526 | 2.2869 | 1.0158 |
The composite $Z$ for a difference from the null, across all three sets, is $-1.2197$, entirely consistent with chance variation.

The population counts of void runs in and contrary to the direction of intention provide a secondary check of the existence of simple forms of bias. It may be noted that in the 100-trial run length the number of void runs in the direction of intention actually exceeds the number of void runs contrary to intention. Of the total population of void runs across all three categories, there are 358 contrary to intention, 330 in the direction of intention, and 6 null, producing a net $Z = -1.0675$ against a hypothesis that a void run is equally likely to be in, or contrary to, the direction of intention.

The void 1000-trial runs do differ significantly ($p = 0.036$, two-tailed) from the formal 1000-trial runs, and the meta-analytic combination of these three $T$-scores produces a marginally significant composite $Z = -1.9867$ ($p = 0.047$, two-tailed). While this significant difference could be taken as evidence for real bias in the void selection process, it also should be noted that the combination of no significant deviation from the null with a significant deviation from the formal data is consistent with the existence of a genuine effect in the formal data, provided that the circumstances which led to the rejection of a run as void also were such as to impede any anomalous effect. Since the most important criteria for voiding involved equipment breakdowns of various sorts, and unavoidable external interruption or distraction of the operator, this last would seem a reasonable expectation.

The recalculated 2-scores that include the voids along with the formal data are, for obvious reasons, slightly decreased. As a result, the composite $Z$-score representing the overall evidence for an anomaly, which is 3.8087 for the formal data, is reduced to 3.4439 when the voids are included, a reduction of approximately 9.6%. We may conclude from this that there is marginal evidence for a bias in the selection of void runs with values, but that it does not substantially impact the evidence for the existence of an effect. While it might produce a small distortion in the observed effect size, this distortion is smaller than the statistical uncertainty in that observation.

Having resolved the interpretation of those void runs which have recorded values, the next stage is to consider the voids with no values by applying the selection formalism. As discussed above, these voids without values are the appropriate population of missing data to be compared with the filedrawer quotients computed for each model. Table 6 compares the actual filedrawer quotients $Q$, as computed from the formal data populations in Table 2 and the void populations in Table 5, with the theoretical values required by the fractional rejection, short tail, and increasing bias models. (The cutoff model has been dropped from Table 6 since its statistical predictions have already been shown to be completely incompatible with all datasets.)

The 2-scores presented in Table 6 are computed from the mean and standard deviation of the binomial distribution for the theoretical rejection rate mandated by the model. That is, the predicted rejection rate for a given model, in
TABLE 6
Filedrawer Populations

<table>
<thead>
<tr>
<th>Item</th>
<th>50-trial runs</th>
<th>100-trial runs</th>
<th>1000-trial runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of formal runs</td>
<td>12,849</td>
<td>3340</td>
<td>700</td>
</tr>
<tr>
<td>Void runs without values</td>
<td>184</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Real filedrawer Q</td>
<td>0.0143</td>
<td>0.0003</td>
<td>0.0043</td>
</tr>
<tr>
<td>Fractional rejection model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predicted Q</td>
<td>0.0347</td>
<td>0.0504</td>
<td>0.0608</td>
</tr>
<tr>
<td>Predicted void population</td>
<td>446</td>
<td>168</td>
<td></td>
</tr>
<tr>
<td>Z, real vs. prediction</td>
<td>12.3182</td>
<td>12.8824</td>
<td>6.0869</td>
</tr>
<tr>
<td>Short tail model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predicted Q</td>
<td>0.0176</td>
<td>0.0259</td>
<td>0.0313</td>
</tr>
<tr>
<td>Predicted void population</td>
<td>226</td>
<td>87</td>
<td>22</td>
</tr>
<tr>
<td>Z, real vs. prediction</td>
<td>2.7740</td>
<td>9.2188</td>
<td>4.0422</td>
</tr>
<tr>
<td>Increasing bias model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predicted Q</td>
<td>0.0506</td>
<td>0.0735</td>
<td>0.0886</td>
</tr>
<tr>
<td>Predicted void population</td>
<td>650</td>
<td>245</td>
<td>62</td>
</tr>
<tr>
<td>Z, real vs. prediction</td>
<td>18.1485</td>
<td>15.5863</td>
<td>7.4770</td>
</tr>
</tbody>
</table>

conjunction with the known population of retained data, gives us both an expected population of voids, and a standard deviation for that expected population. The observed population of void runs can then be compared with that theoretical mean and standard deviation to obtain a Z-score, which is the source of the Z values listed in Table 6. The sign of the difference was ignored in computing these Z-scores, so if p-values are calculated for them, the two-tailed form must be used.

It is evident that all of the models predict rejection rates grossly in excess of the actual number of void runs without recorded values. In many cases the mismatch is so extreme that p-values cannot be calculated readily by conventional techniques. The least significant result is the value of 2.774 for the short tail model on the 50-trial runs, with a two-tailed p-value of 0.006. This particular model, however, can already be rejected on the basis of its distribution statistics as discussed above. No model examined here can plausibly accommodate both the observed distribution statistics and the known rate of run rejection.

Additional General Arguments: Models with More Parameters

Except for the cutoff model, which demonstrably has the smallest possible filedrawer for a given mean shift, the selection models examined above are not obviously optimal or extremal in their properties. While they show the same relationship between distortion and filedrawer expected from the analysis of normality-preserving selection, it seems worthwhile to sample the space of possible s(x) in somewhat more detail to test the generality of this property.

An immediate generalization can be obtained by noting that the cutoff model and the fractional rejection model are two extreme members of a single, two-parameter family of selection processes. This family of "fractional cutoff"
models rejects a fraction \( r \) of all data falling below a minimal cutoff boundary \( b \).

The simple cutoff model is then the fractional cutoff model with \( b = a \) and \( r = 1 \);
the fractional rejection model is the fractional cutoff model with \( b = 0 \) and \( r = a \).

A choice of a specific mean shift does not, of course, identify a specific member
of this two-parameter family, but rather determines a one-parameter subset of it.

For more general investigations, one may begin to approximate the freedom
of the full, arbitrary \( s(x) \) by adopting a sectional selection model. For some \( n \),
choose \( n - 1 \) boundaries \( b_1, \ldots, b_{n-1} \) in the real line, along with the notional
"boundaries" \( b_0 = -\infty \) and \( b_n = +\infty \). For convenience, and without appreciable
loss of generality (since \( n \) is not fixed), these boundaries can be placed at
uniform quantiles of the inverse normal distribution, so that \( \int_{b_{k-1}}^{b_k} f(x) dx = 1/n \)
for each \( k \in \{1, \ldots, n\} \). The selection function \( s(x) \) is then taken as the piecewise
continuous function defined by \( s(x) = s_k, \quad b_{k-1} < x < b_k, \quad \text{for each} \quad k \in \{1, \ldots, n\} \).

It is obvious that for this \( n \)-section selection function \( S = (1/n) \sum_{k=1}^{n} s_k \).

The higher moments are given by \( \langle x^m \rangle = (1/S) \sum_{k=1}^{n} s_k I(b_{k-1}, b_k, m) \), where the integral
function \( I \) is defined as \( I(a, b, m) = \int_{a}^{b} x^m f(x) dx \); note that the \( I \) terms depend
only on the set of \( b_s \) and are the same for any selection function using the same
set of sections.

Figure 4 shows numerous selection models plotted against their distortion
index \( D \) and filedrawer quotient \( Q \). All models in this figure are constructed
to have \( \mu = 0.0277 \), identical to the 50-trial runs subset. The shaded area in
the lower left corner identifies the limits of \( Q \) and \( D \) imposed by a \( Z < 2 \) criterion
of consistency with the observed characteristics of the 50-trial runs; any model
outside that area will be rejected at \( p < 0.05 \) (two-tailed) or better for its
distribution shape, its filedrawer prediction, or both.

The four filled markers show the four one-parameter models, as labeled on the
figure. The smooth solid line shows the behavior of the normality-preserving
selection model, where the distortion index is purely driven by the change in
variance. The dotted line shows the evolution of the two-parameter fractional
cutoff model as it is smoothly changed from the cutoff model to the fractional
rejection model while maintaining constant mean shift. It is noteworthy that this
curve passes very close to the point characterizing the short left tail model,
despite their very different functional forms. It is also intriguing that as the
cutoff boundary \( b \) migrates toward zero, the index of distortion passes through
a minimum and then begins to increase again, while the filedrawer quotient
increases monotonically throughout.

The remaining elements of this figure are based on sectional selection models
with 100 sections. The cross icon is the sectional model that most nearly matches
the optimal (maximal \( S \)) normality-preserving selection; as one might expect it
lies on the curve for such models, at its minimum in \( Q \) (which is its maximum in
\( S \) given \( Q = (1 - S)/S \)). The open circle is the 100-section sectional model
defined by \( s_k = \alpha_0 + \alpha_1 k \), where the linear parameters \( \alpha_0 \) and \( \alpha_1 \) are determined by
the joint constraints \( s_{100} = 1 \) and \( \mu = 0.0277 \).

The dots are the result of an iterative optimization procedure applied to a 100-
section model. At several starting points, including the smoothly increasing model, the normal model, and three models close to the different members of the fractional cutoff family, a gradient-following optimization algorithm was used to try to reduce $D$, reduce $Q$, or both. For each starting point one algorithm attempted to reduce $D$ without regard to effects on $Q$, another minimized $Q$ without regard to $D$, and others attempted to minimize linear combinations of $Q$ and $D$ with varying weights. These optimization algorithms halt when their attempt to follow a downward gradient produces an increase rather than a decrease in the target parameter, indicating that a local minimum lies within the algorithm's finest resolution for adjustments to the individual $s_k$.

It is notable that the scatter of dots at the upper left represent one model—the one minimizing $D$ without regard to $Q$—from each of the five starting points. Similarly, several of the larger dots appearing along the "fractional cutoff family" dotted curve result from efforts starting at different points to minimize $Q$ without regard to $D$. One, obscured by the large block showing the simple
cutoff model, essentially rediscovered that model to within the resolution of the 100-section parameterization. Other dots represent other local minima for the different linear combinations $\alpha D + \beta Q$ being minimized.

The endpoints of these optimization attempts, along with the parametric curve of the fractional cutoff model family, strongly indicate the existence of some limiting curve below which the distortion index and the filedrawer parameter cannot simultaneously be reduced. Moreover, these facts suggest that the fractional cutoff family curve is either on or very near that limit at least up to its inflection point, and that the limiting curve can be expected to have some smooth continuation into the region populated by the D-minimizing endpoints at the upper left. Finally, it is clear that the four one-parameter models examined in detail are all quite near the joint lower limit of distortion and filedrawer, and span much of the possible range of distortion values.

Given the failure of any selection model to occupy the shaded region statistically consistent with the observed REG data, even if the selection model is allowed to optimize 99 free parameters in an effort to achieve such consistency, it seems reasonable to conclude that the REG observations are inconsistent with any form of selection hypothesis whatever. A graph similar to Figure 4 could be drawn for the 100-trial runs; it would show even more dramatic inconsistencies due to the larger effect size and smaller filedrawer. The same cannot be said for the 1000-trial runs due to the poor statistical resolution resulting from their relatively small population. However, even if this subset is regarded as suspect due to the possibility of selection, the 50-trial and 100-trial runs between them produce an aggregate composite $Z = 3.9044$ (using the standard per-trial weighting), a result actually slightly stronger than that of the REG database as a whole.

**Conclusions from Model Comparisons**

On the basis of the statistical parameters of the data distribution, we are able to reject the data-selection models producing the strongest distortion of the source statistics as an explanation for the anomalous mean shift in the REG data (Table 5). When we examine models with lower distortion indices, we find that the distribution statistics become indistinguishable from a mean shift hypothesis, but the required rejection rate for these models is so large as to be completely incompatible with experimental records (Table 6).

It has been proven above that for selection models preserving normality of the output statistics, there is a tradeoff between the variance (the only free parameter once the mean is fixed) and the filedrawer; the more closely the selection process tries to preserve the original variance while imposing a nonzero mean shift, the more data it must discard. Figures 3 and 4 illustrate that the tradeoff between increasing filedrawer vs. increasing distortion appears to be a general feature of selection models, even when the number of adjustable model parameters is increased and when optimizing searches are made to try to
improve their performance in these features. If this generalization, supported by all currently available evidence, is valid, then the fact that every selection model considered is either strongly rejected by its distribution statistics, strongly rejected by its predicted filedrawer population, or both, means that no form of selection model can account for the data.

As noted above, the experimental protocol involved logging and recording all rejected or missing data along with the reasons for their rejection. Any experimenter being misled by bias into making an invalid decision to discard a dataset thus would leave a record of this act, even if the data values themselves went unrecorded. The only possibility for selection rates large enough to induce the effects therefore requires deliberate deception on the part of the experimenters, rather than simple bias. Experimenter fraud of this sort is frequently invoked as a last-resort accusation for explaining away anomalous results. A drawback of this "explanation" is that it is innately unfalsifiable: once it has been decided that a given experimenter is fraudulent, there is no reason to believe anything that experimenter says, nor any argument that the experimenter can make to refute the accusation. Perhaps more to the point, experimenters who set out to conduct a fraudulent experiment have far less labor-intensive ways to do so than carefully hiding a selected subset of the experimental data after they were generated.

A Subset of Special Interest

Aside from the general question of data selection in the REG experiment, there is a specific subset where the issue is of extra interest. As has been noted in the past (Dunne & Jahn, 1995; Dobyns & Nelson, 1998), the operator assigned to ID code 010 produced impressively large effect sizes in the early period of the experiment. This early period is distinguished from this operator's later data, not only by a temporal hiatus of over a year in which no data were generated, but also by a change in device (a portion of the hiatus, for this operator and all others, was caused by the delay of qualifying and calibrating the replacement REG machine), and by a change in protocol (it was decided during this period that all secondary parameters, such as volitional vs. instructed assignment of trial intentions, must be held constant throughout a series, rather than being variable on a session-by-session basis as had been the case previously). The early data for Operator 010 have an effect size more than an order of magnitude larger than any other database; they are statistically distinguishable not only from the vast bulk of other operator performances, but from the contemporaneous early data of other operators, and from the later performance of the same operator as well. The reasons for the larger effect are not clearly understood.

The distinctive character of this early 010 dataset, and its lack of explanation, mandate that all reasonable hypotheses for its outcome must be carefully scrutinized. Therefore, it seems appropriate to apply the formalism developed herein to the possibility that this extraordinary database was produced by deliberate deception on the part of the operator.
The formal data in this set comprise 503 runs, with a mean Z-score of 0.2556 (composite Z = 5.732). All runs are in the 50-trial length. There are 38 void runs with values in this database; these voids have a mean value of -0.2037 (composite Z = -1.256, nonsignificant). Combining the voids with known values with the formal data would lower the effect size (mean Z-score) to 0.2233, with an associated composite Z-score of 5.194. As in the general database examination, the voids with known values have an apparent negative bias that is nevertheless well within the range of plausible chance variation; the change between effect size with and without the void data is within the statistical uncertainty of the measured effect size in the formal data, and the Z-score of the recombined data remains highly significant. We may conclude that, as in the general analysis, the voids with values do not appreciably impact the experimental conclusions, and they need not be considered further.

The database also contains 10 voids with no recorded values. Unfortunately, since we are here considering the possibility of deliberate concealment of data in addition to experimental bias, this does not give us a value for the actual filedrawer. Indeed, the actual filedrawer in a case of deliberate data selection is both unknown and unknowable. Our interpretation of theoretical filedrawer predictions of model fits must, rather, be based on the credibility of the operator's having managed to discard the requisite amount of data without detection by the experimenters — assuming a selection process that is consistent with the statistical parameters of the data.

Table 7 presents the standard deviation, skewness, and kurtosis of the actual database, as contrasted with the selection models developed previously. The filedrawer quotient Q required for the selection model to fit the observed mean shift is also reported.

An analysis of statistical power indicates that the most sensitive test parameter is the skewness, in the case of the fractional rejection model, and the standard deviation for all other models. In contrast to the examination of the general

<table>
<thead>
<tr>
<th>Model Comparisons for Special Subset</th>
<th>( \sigma )</th>
<th>( \gamma_3 )</th>
<th>( \gamma_4 )</th>
<th>( Q )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual data</td>
<td>1.0402</td>
<td>0.1037</td>
<td>0.1538</td>
<td>NA</td>
</tr>
<tr>
<td>Cutoff</td>
<td>0.8100</td>
<td>0.5547</td>
<td>-0.0482</td>
<td>0.160</td>
</tr>
<tr>
<td>Z-score</td>
<td>-7.300</td>
<td>4.129</td>
<td>0.925</td>
<td></td>
</tr>
<tr>
<td>Fractional rejection Z-score</td>
<td>0.9668</td>
<td>-0.2459</td>
<td>0.2698</td>
<td>0.320</td>
</tr>
<tr>
<td>Increasing bias Z-score</td>
<td>-2.329</td>
<td>-3.201</td>
<td>0.531</td>
<td></td>
</tr>
<tr>
<td>Short tail Z-score</td>
<td>0.8467</td>
<td>0.2980</td>
<td>0.0639</td>
<td>0.191</td>
</tr>
<tr>
<td>Increasing bias</td>
<td>-6.136</td>
<td>1.779</td>
<td>-0.412</td>
<td></td>
</tr>
<tr>
<td>Increasing bias Z-score</td>
<td>0.9668</td>
<td>-0.0697</td>
<td>0.0834</td>
<td>0.467</td>
</tr>
<tr>
<td>Increasing bias Z-score</td>
<td>-2.329</td>
<td>-1.587</td>
<td>-0.322</td>
<td></td>
</tr>
</tbody>
</table>
database, we generate a $p < 0.05$ rejection of the selection model in every case. Moreover, those models for which the rejection is weakest ($p = 0.0013$ for fractional rejection, $p = 0.020$ for increasing bias, both two-tailed) predict large filedrawer quotients; generating these data by the increasing bias selection model would require discarding very nearly one run for every two which were recorded. While experimenter vigilance might be less than perfect, the redundant measures deployed to prevent operators from concealing the fact that they have generated data (the continuous hardcopy of the data is perhaps the most relevant to the current instance) make it difficult to credit that an operator could succeed in concealing one experimental run out of every three.

**Final Summary**

The space of possible selection processes is, essentially, the space of all functions of $x$ bounded by $[0, 1]$. A set of four relatively simple one-parameter families of selection processes nevertheless allows some conclusions to be drawn. The cutoff model provably demonstrates the existence of a minimum level of discarded data for any target level of mean shift. The performance of numerous sectional selection models allowing multi-parameter optimization indicates a lower limit to the filedrawer for any given level of statistical distortion, and indicates further that the minimal filedrawer increases as the distortion decreases, and that the four one-parameter models are close to if not actually occupying this minimal limit.

For the effects in the database as a whole, the effect size is small enough that some selection models can produce statistics indistinguishable from those observed in the data. The possibility that the apparent effect was constructed by biased selection can be refuted, however, by comparing the actual population of discarded runs with the population required by those selection models that produce adequate fits to the data statistics. Surveying the space of optimized multi-parameter selection models confirms that the conclusions drawn from the single-parameter models can be generalized to all selection models with high confidence. Since failure to record the existence of a discarded run would require the deliberate circumvention of protocol rather than mere biases of judgment, the thesis that the anomalous effect as a whole could be due to unconscious selection of favorable data can be rejected.

For a database that has been regarded as suspect due to its origin in the exceptional performance of a single operator, all models examined predict statistics significantly different from those actually present. Even the least ill-fitting model requires a "filedrawer" of discarded data so large that its successful concealment from the experimenters by a malevolent operator becomes incredible. Since improving the statistical fit would require an even larger filedrawer quotient, the hypothesis that Operator 010 produced a spurious effect by concealing negative runs also can be regarded as refuted.
Appendix: Relative Statistical Power of Tests

It is common practice to compare distributions by using distribution tests such as $\chi^2$ goodness-of-fit tests or Kolmogorov-Smirnov tests, to the extent that the use of a moment-based test may strike some readers as archaic. However, moment tests offer better statistical power than such distribution tests when a specific hypothesis regarding a moment value is available.

Since in the worst case of a selection effect we may be confronted with a normal distribution that merely has a smaller standard deviation than expected, the detection of reduced variance is used as a test case. Table A.1 presents, for a range of $N$, the probability of Type II error for an optimally sensitive $\chi^2$ test, given a change in $\sigma$ such that a simple variance test has $\alpha = \beta = 0.05$. In other words, each line of the table was computed by calculating, for that $N$, the value of $\sigma$ that would produce a 5\% chance of failing to be rejected by a $p < 0.05$ criterion on a variance test. The $\chi^2$ tests were based on uniform-population binning; the number of bins was chosen by finding the bin number which minimized the Type II error probability $\beta$. The $\beta$ values given in the table assume $\alpha = 0.05$, that is, that the $\chi^2$ test will reject the null hypothesis on a $p < 0.05$ criterion.

Recalling that $\alpha = \beta = 0.05$ for the direct moment test on the variance, it is obvious that the $\chi^2$ test has a much higher chance to overlook the same effect on the same data. It is notable that as $N$ increases, the loss of performance in the $\chi^2$ test becomes worse.

Similar considerations apply to the Kolmogorov-Smirnov test for distribution differences. For a simple demonstration, a Monte Carlo test was run by generating 10,000 sample distributions, each comprising 100 normal deviates with $\sigma = 0.767$. As noted above a simple variance test will reject such samples at the $p < 0.05$ level 95\% of the time. The K-S test, in contrast, produced $p < 0.05$ rejection on only 2164 of the samples, indicating a Type II error probability of approximately 78\%. It was considered redundant to extend this investigation to larger sample sizes.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\sigma$</th>
<th>Bins</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.767</td>
<td>5</td>
<td>0.363</td>
</tr>
<tr>
<td>200</td>
<td>0.836</td>
<td>6</td>
<td>0.449</td>
</tr>
<tr>
<td>500</td>
<td>0.896</td>
<td>7</td>
<td>0.511</td>
</tr>
<tr>
<td>1000</td>
<td>0.926</td>
<td>7</td>
<td>0.535</td>
</tr>
<tr>
<td>5000</td>
<td>0.967</td>
<td>8</td>
<td>0.565</td>
</tr>
<tr>
<td>10,000</td>
<td>0.977</td>
<td>8</td>
<td>0.569</td>
</tr>
</tbody>
</table>

Acknowledgments

This work was supported by donations from the Fetzer Institute, the Institut für Grenzgebiete der Psychologie und Psychohygiene, and numerous private donors including Laurance Rockefeller, George Ohrstrom, and Donald Webster.


**Comments from Mikel Aickin**

The Dobyns article gives the impression that all reasonable subconscious ("unconscious" in his terms) distortions of the PEAR REG data through data selection have been ruled out. I believe this implication is untrue. I programmed the simulation of a very simple automated strategy, which is oriented toward moving the mean of the data, while arranging things so that a statistical test of Normality would be passed. I used a better Normality test than Dobyns did, so that my simulation provides stronger evidence than his. In my simulations, I counted a success when I could produce the results cited by Dobyns for the data that he retained, without non-Normality being detected. I then recorded the percent of data that had to be deleted, among the successful cases. Here are the results for the data presented by Dobyns.

<table>
<thead>
<tr>
<th>Type of Experiment</th>
<th>Reported % Data Deleted</th>
<th>Simulated chance of successfully distorting the data</th>
<th>Simulated % of data deleted among successful distortions</th>
</tr>
</thead>
<tbody>
<tr>
<td>50-run</td>
<td>5.7%</td>
<td>1.00</td>
<td>1.0%</td>
</tr>
<tr>
<td>100-run</td>
<td>2.3%</td>
<td>0.09</td>
<td>0.5%</td>
</tr>
<tr>
<td>1000-run</td>
<td>3.7%</td>
<td>0.33</td>
<td>1.0%</td>
</tr>
</tbody>
</table>

It seems clear from this that it is possible to produce the PEAR REG results through data selection. This says almost nothing, of course, about whether any such selective distortions occurred. I would rely on the professional reputation and integrity of the PEAR investigators, which I regard as beyond question, for the validity of the data. Further, since Dobyns reports that, whether one includes or excludes the "void" data, the overall conclusions about the experiments are the same; it is not clear to me why any of this is of any importance.
LETTERS TO THE EDITOR

The Wave Function Really Is a Wave

I would like to point out what I believe are certain misunderstandings and inaccuracies in the recent article by M.G. Hocking, "Linking String and Membrane Theory to Quantum Mechanics and Special Relativity Equations, Avoiding Any Special Relativity Assumptions", Journal of Scientific Exploration Vol 21, Number 1, 13-26 (Spring 2007).

There are a number of relatively inconsequential errors:

In the first sentence, of the 10 (11 in more recent theories, as noted) dimensions in string/M theory, six (not seven as stated) are believed tightly coiled up, leaving four (not the stated three) dimensions, including time, that we perceive as uncoiled. In the subsequent paragraph, approximately 96%, not 90% as stated, of the matter in the universe is "missing" (i.e., invisible.)

The term "normalising" on page 16 is not simply "squaring" (sic) of the wave function, but comprises setting the probability for finding a particular extant particle somewhere in space to be 1 (quite a reasonable proposition.) Further, this "squaring" does not necessarily create a "matter wave" (sic), as stated, since it can result in a non-moving probability density distribution (and waves, by definition, move).

On page 17, the quote attributed to Huxley "Nature is not only stranger than we have thought, It is stranger than we can think!" may possibly have been said by him, but it seems what was meant was the oft referenced remark by Sir Arthur Eddington, "Not only is the universe stranger than we imagine, it is stranger than we can imagine."

The term "quark string theory" is used often, but string theory is not limited to solely to quarks, and is never referred to in this way. Besides quarks, it also includes leptons (such as electrons and neutrinos), bosons (such as photons, gravitons, Higgs particles), and more.

On page 24, force is said to be "an energy field or gradient". It is not an energy field, but it is the gradient of the potential, quite a different thing. An energy field has units such as ergs/cm^3. Force has units such as dynes.

More serious are the following:

The hypothesis that the square of the absolute value of the wave function is the particle probability density is denigrated without mentioning the many experiments that appear to have clearly demonstrated it not only is so, but that this density has been shown to be typically continuous, and not point like. Indeed, scattering experiments, such as those done daily in particle...
accelerators around the world, show this to be the case. How about some references here?

On page 18, the article states "On the basis of the Big Bang theory with its residual microwave radiation . . . there is an absolute reference point of origin . . . in space. This negates the 1st Principle of Special Relativity, which denies an absolute reference point in space." This is a common misunderstanding, and is wrong for several reasons.

For one, special relativity is an idealization that deals with space devoid of matter and unaffected by gravity, and this obviously does not include the Big Bang universe. For another, within general relativity (the part of the theory including matter sources of gravity), one often prefers certain reference frames because they are more convenient for calculation (as is the case for the Big Bang universe), but they are definitely not preferred in the sense that Nature is making them the sole fundamental reference frame (as in Newtonian physics.) For yet another, the first principle of special relativity does not address the issue of reference points, but of reference frames. Indeed, Newtonian physics and relativity hold the exact same position with respect to reference points, i.e., there is no absolute such point. Finally, there is no "center" to our universe. Fundamental symmetries of its expansion dictate that every point in the universe is like every other point, in the sense that from any such point, all of the rest of
the universe appears to move away in precisely the same manner. That is, every point in creation looks like its own center of the universe.'

But the most fundamental issue I have is with the interpretation of the wave function as "imaginary" and thus not real, at all but discrete points in our spatially three dimensional world. The truth is that the wave function is complex, not simply imaginary, at virtually all of those other points not shown in Hocking's Fig. 1. Complex numbers have both real and imaginary components.

So as time evolves, the wave function $\psi$ has both an imaginary part that varies sinusoidally, and a real part that does so, as well. Implying it only has a real part at particular 3D points is not correct. And thus it does not follow that the particle somehow "jumps" from each of these points to the next.

Fig. 1 illustrates the differences between real waves (such as pressure waves with only a real number magnitude such a pressure) and complex waves (such as the wave function with real plus imaginary components of its magnitude.)

What I refer to in the figure as "Pressure vs x" and "Wave Function vs x" spaces are what Hocking refers to as "configuration space". Note that the corkscrew function in the lower right hand side is the wave function. Since this intersects the $\text{Re } \psi$ vs x plane at specific points (of equal values separated by one wavelength),
one can see how some might be led to believe that $\psi$ only has real values at those points. This is not true, however, as can be seen with the aid of Fig. 2 herein.

In Fig. 2, only a short section of the "corkscrew" curve of Fig. 1 is displayed, and that particular section does not intersect the Re $\psi$ vs x plane. However, as shown, it has both real and imaginary components for every point along that section. That is, it has a real value even though the curve of $\psi$ does not intersect the plane formed by the Real $\psi$ axis and the real world spatial direction x.

Thus, the plot of the real component of $\psi$ vs x is a sinusoidal curve, which has continuous, not discrete point, values. That is, the real values do not appear only at spatially separated points, as claimed.

The article (page 15) states that $\psi$ "is not a wave." But it is. It is simply a complex wave, rather than a real one. Not only theory, but an enormous number of experiments, demonstrate the wave nature of $\psi$.

As an aside, I have long wished that authors of books on quantum mechanics would include figures such as those herein when introducing the concept of the wave function. Perhaps then, we would have less confusion about what is really going on with the solution to the Schroedinger equation as it evolves through time.

And thus, though there may be some merit in the subsequent arguments made by the author, those arguments may need to be reconsidered in light of the aforemade remarks.

Robert D. Klauber
Fairfield, Iowa
rklauber@iowatelecom.net
rklauber@netscape.net

Note

1 Consider the commonly employed analogy for our expanding universe of a balloon being blown up. Imagine that the balloon expands from being very small, almost a point originally. If you live on the surface of the balloon, where would you determine the "original center" of the balloon to be? It is actually no one point on the surface. Rather, every point would seem to be the original center to someone located at that point.
The Wave Function $\Psi$ has an extensive imaginary component:
so it is just our imagination that thinks it is really a wave?

Dr. R. B. Klauber's Letter to the Editor is replied to, below. Some of the
comments are agreed with and the letter does not invalidate the paper [1]
referred to.

$\Psi$ is not a wave in 3-D space, as clearly shown by Dr. Klauber's Fig. 1 & 2.
My thesis is that $\Psi$ can be interpreted as particle excursions into a higher
dimension, instead of the equally odd idea that matter is wavelike (which is the
probable cause of the difficulty of reconciling quantum theory and relativity).

In numbers of dimensions, I was counting spatial dimensions, excluding time.
Dr. Klauber mentions "four (not the stated three) dimensions, including time"
are perceived as uncoiled, but my "stated three" were the 3 spatial dimensions
excluding time (see first line of Abstract).

Dr. Klauber comments 96% of the matter in the universe is "missing" or
invisible. I had stated "About 90%". I am happy to adopt with his figure.

I agree "normalising" is not simply "squaring" and I did write "in effect
to square" (page 16) for this reason, doing so to keep the paper concise.
Dr. Klauber also says it is a "reasonable proposition" to set the probability of
finding a particle somewhere in space to be 1, but this is just what the many
authors which I cited did before the notion of multidimensional space appeared.
My thesis is that this very assumption "forces" on Nature what our presumption
is of what Nature should be. If the probability is not unity for a series of short
times along a trajectory, this could mean the particle has moved to another
dimension for those short times, rather than become a "wave".

Dr. Klauber comments that a non-moving probability density distribution can
occur but waves move. But such a distribution can be equated to a standing
wave.

In saying on page 24 [1] that a force is defined in basic physics as an energy field
or gradient, dE/dx, this was intended as a general physics statement, not meaning
an electric potential field which Dr. Klauber assumes. Electric potential is not
involved here. Another simple non-electrical example is that an atom diffuses in
the direction in which its Gibbs Free Energy is reduced and the diffusing force on it
is $d(\Delta G/dx)$ where $d(\Delta G/dx)$ is a field (an energy field, specifically a Gibbs Free
Energy field) with units of J/m or erg/cm. The units of an energy field (a force) are
not erg/cm³. Force can also be expressed in dynes and a dyne is 1 erg/cm.

On the points described as "more serious":

(1) Taking $\Psi^2$ to be the particle probability density and that it is shown to be
continuous by scattering experiments is a matter of interpretation of the
meaning of the experimental results: as mentioned above, I took a simple model of one particle moving through free space. Taking $\Psi$ to be "point-like" is only my basic model for a particle moving in free space and the results will obviously be modified in other situations. Once interactions occur with other objects, as in scattering experiments, then disruption of the simple trajectory will occur but scattering phenomena do not invalidate the existence of particles on the proposed model.

(2) Comment made: "Special Relativity is an idealization ... that obviously does not apply to the Big Bang universe". This is true, but even so, Special Relativity is widely used in practice in physics! I do not disagree of course that Relativity does not allow a "sole fundamental reference frame".

Comment made: "Newtonian Physics and Relativity hold the exact same position with respect to reference points, i.e. there is no absolute such point."

I would disagree that the Newtonian position excludes using reference points.

Comment made: "There is no centre to our universe". I would disagree that there is enough evidence to suppose that every point in the universe is like every other point and each would look like its own centre of the universe. Space may have existed before the Big Bang, which would void the argument that there is no centre of the universe, which would then obviously be at the point of origin of the Big Bang.

I do not accept the balloon surface inflation model of space in Dr. Klauber's footnote.

These points in this heading (2) do not, in any case, invalidate the derivations of the Special Relativity equations given assuming absolute motion.

(3) Interpretation of wave function as imaginary:

Again, I would say that the model given in my paper [1] is for a particle moving in a free-space trajectory. The ordinary graph of $\Psi$ given in [1] is certainly a series of points separated by gaps, for the equation $\psi = \exp[-2i\pi((xm/2h) - tE/h)]$ which is a well-known solution of Schroedinger's Equation, $d\psi/dt = (\hbar i/4\pi m) [d^2\psi/dx^2]$.

All I am saying here is that the gaps between the points are where the value of $\psi$ is imaginary in the sense that can mean that the particle can exist in a higher dimension in between the real points. A distinction between "imaginary" and "complex" would not invalidate my thesis. I thus do not argue against the Figure 1 given by Dr. Klauber, which does not negate the interpretation that I have made. I am just trying to interpret the meaning of the imaginary/complex $\psi$ as being a periodic excursion into a higher dimension. That there is an imaginary component cannot be
discounted and I also do not argue against Fig. 2 of Dr. Klauber. The plot of $\psi$ requires an imaginary axis at right angles to any of the 3 directions in 3-D, which is suggestive of a 4-D involvement. The real component shown in Dr. Klauber's Fig. 2 corresponds to a fractional probability which is also suggestive of a transit into 4-D for the curve segment in Fig. 2.

There could be a semantic problem here, as Dr. Klauber says, "the plot of the real component of $\psi$, vs $x$ is a sinusoidal curve, which has continuous, not discrete point, values", whereas I am saying a plot of the total value (not the real component) of $\psi$ (my Fig. 1) is a series of discrete points.

(4) Comment on the point saying that $\psi$ "is not a wave". Again, I am referring to the total value of $\psi$, not what Dr. Klauber refers to as the real component of $\psi$. I do not deny that $\psi$ appears to be a wave in many experiments but it is possible to give an alternative explanation of this apparent wave nature in terms of particle excursions into a higher dimension.

M.G. Hocking

Erratum: In reference [1], on page 16 line 7, "where and when $x$ and $t$ are both integers" should read, of course, "where and when $x/\lambda$ and $tv$ are both integers".

Reference

OBITUARY

IN MEMORIAM GEORGE SASSOON 1936–2006

Beyond compare is the intellectual stimulation and companionship I have enjoyed through the Society for Scientific Exploration. Editing the Journal of Scientific Exploration brought me into contact with even more fascinating characters than I had been able to encounter at the Society's meetings. Among those extraordinary individuals was George Sassoon, to whom the much-misused term "unique" happens to be literally appropriate.

Obituaries in the Daily Telegraph, The Independent, and elsewhere give a sense of the man's diverse talents. He "attained distinction as a scientist, electronic engineer, linguist, translator of scientific papers, player of the piano accordion and investigator into extra-terrestrial phenomena. His book The Manna-Machine (1978), and its companion volume The Kabbalah Decoded of the same year, investigated the origins of the manna that sustained the Israelites in the desert; Sassoon went back to Jewish texts, particularly the Zohar, a collection of 13th-century writings which he translated from Aramaic". Sassoon commanded a knowledge not only of Aramaic (and of course English, French, and German) but also of Serbo-Croat, Hebrew, and Klingon. He was a radio ham as well as a professional engineer; his book, The Radio Hacker's Codebook (1980), was not welcomed by the authorities because of its insights into encryption.

Various of the obituaries mention—or more than mention—the idiosyncratic and difficult family circumstances in which George Sassoon grew up. He was the only child of the poet Siegfried Sassoon, enduringly famous for his depiction of the personally experienced horrors of World War I; a well-received biography of Siegfried was published quite recently. (It was rather a surprise to find a "Siegfried" to be the scion of Sephardic Jews.)

George Sassoon and I had corresponded by e-mail about the pendulum studies of Allais, about cold fusion and sonoluminescence, about Tunguska, and about much else. At times he would come up with the sort of puzzlers that are featured in The Scientist Speculates, I. J. Good's collection of "partly baked ideas"; for instance:

"Space is curved", said Einstein. If this is so, the value of pi will be different in practical measurements to that obtained by calculation, which assumes a flat universe. Has anyone considered doing a practical measurement to the required degree of precision? If so, how?

George once told me about a German book about possible dinosaur survivals and sent me a copy of his translation of it. He had been unsure about the German titles "Freiherr" and "Rittmeister", and I was able to check his guesses—which
were correct—in the old Muret-Sanders dictionary\(^3\) that I had inherited from my father. In response, George mentioned that he had once taught himself to read the old German handwriting script in order to translate a letter for his uncle.

Sassoon was not just a passive observer, he drew on his knowledge of engineering to make measurements of a variety of phenomena. Concerning David Deming’s review of the "Hum", Sassoon wrote that his wife experiences it, but only at the shore; and he prepared to carry out measurements to check the possibility that very-low-frequency electromagnetic waves might be the source. As to "cold fusion"—which has come to be more commonly described as "low energy nuclear reactions" (LENR) or "condensed matter nuclear science"—Sassoon carried out experiments himself to check the claims that welding could give rise to nuclear transformations, as revealed by resulting radioactivity. Also concerning cold fusion, a neighbor of Sassoon’s at his residence in southern England was the electrochemist Martin Fleischmann, discoverer of cold fusion; they would occasionally get together at a local pub, and via George I was able to recall with Martin the year I had spent in his Department at Southampton University.

George Sassoon had a residence on the Isle of Mull in Scotland as well. His neighbors there included Lionel Leslie, who had explored for lake monsters in Irish loughs, and Christopher James, son of David James who had led important expeditions to Loch Ness in the 1960s and 1970s. Again via George, Christopher and I exchanged interesting information about Loch Ness matters.

I had hoped to meet George Sassoon in person during one of my trips to Scotland, but it was not to be, to my great regret. Yet I learned much through our correspondence and his exceptional range and depth of interests.

Ron Bracewell knew George Sassoon personally, and offers the following further recollections.

**Henry Bauer**

**Sources**

2. Hartwig Hausdorf, Die Rückkehr der Drachen (The Return of the Dragons), Herbig (Germany), 2003

**Episodes from the Life of George Sassoon**

My old friend George was well known in many contexts, first of all as the son of Siegfried Sassoon, the World War I poet. George, however, had a technical bent, and became mathematically proficient at Cambridge, where he attended
Publication, for him, did not exert the driving force that animates the inhabitants of refereed journals. This is not to say that he did not write. Indeed, to illustrate one of the outstanding features of his intellect, an interest in languages, he published a translation of the Zohar, a component of the Kabbalah so abstruse that his was the first English translation. Why had this centuries-old Jewish text been so neglected? It appeared to consist in part of a record of the dimensions of God, the distance from His knee to His elbow, the number of hairs in His beard, and other data deemed to be without wide interest. Thinking about this, George recalled that in the Sioux language the parts of an automobile are named after parts of the human body or of some animal or vegetable. A few parallels with this can be found in English; the word "nut" (that which screws onto a bolt) is a mechanical word borrowed from botany. But in Sioux the headlights are eyes, the wheels are legs, the doors are wings, and even the exhaust pipe has an anatomical name. Possibly then, the Zohar could be understood as a specification of a machine. To explore this hypothesis George learnt mediaeval Aramaic, the language of the Zohar. That is not something that you or I would undertake lightly, nor would we have known about Sioux automobile mechanics. The companion of this published translation was his book *The Manna Machine* in which he hypothesises that the text specifies the construction of the Ark of the Covenant. In support of this interpretation he noted that Uzzah was smitten dead when he inadvertently put out his hand to steady the Ark, that the Philistines, who captured it from the Israelites were stricken with "emerods", that the unfortunates who carted it back to the owners were smitten, and that when Israelites of the Exodus pitched camp, access to the tent protecting the Ark was restricted to Moses and Aaron and that the tent had no roof—the air above it glowed red, indicating dangerous radioactivity. Possibly then, the Ark was of extraterrestrial origin and dangerously radioactive. Later, in the Temple of Solomon, Aaron wore a breastplate famous for its twelve glowing jewels. Was this the computer keyboard with which Aaron communicated with the supernatural entity? The present whereabouts of the Ark are uncertain. Many copies exist because its precise dimensions are on record. Possibly it is the one guarded today in a church in Axum, Ethiopia; George reasoned that it may be on the bed of the Tiber. As a learned student of the Torah and Talmud, George could discuss exotica such as whether rhinoceros meat was kosher and whether the ban on killing insects on the Sabbath applied to head lice.

To my knowledge George spoke Serbo-Croatian fluently, was fond of his ability to persuade his computer to print Armenian script, lundly provided me with a disc containing a grammar and dictionary of Maltese (which I was happy to pass on to a Maltese-American friend for the edification of his son). He gave me a Latin-Sorbian dictionary in case I ever encountered Sorbs—who, as every Ukrainian knows, survive only in Germany; and he provided me with a grammar and dictionary of Klingon, a language hardly heard of in Central Europe.
Nobody knows much about the origin of Basque but George was of the opinion that "bai eta ez" (yes and no) sounded a bit like Georgian.

On the technical side George consulted for the oil-well drillers in the North Sea. He gave much attention to encryption and, when the United States government restricted the publication of codes that the CIA could not decrypt, became unpopular by publishing "The Radio-Hackers Codebook". His program, written in BASIC, for generating prime numbers, is a gem. It goes far beyond the nominal 12-digit limit of your everyday computer. At home in Mull he studied the radioactivity of pitchblende and looked for correlation with terrestrial and solar weather, magnetograms, and seismograms. A web-camera on a hilltop outside his house provided internet users a view of the current weather on Mull, should they need to know. He lived not far from a tower built by the Knights Templar in the days of the Crusades. When the Knights were banished from France and Spain they travelled to Ireland and then on to Mull in the 14th century where, George concluded, they seeded the practice of freemasonry in Britain.

In the thirties, when the radio transmitter in Luxembourg was the most powerful in Europe, listeners reported strange echoes. They would hear what the announcer said, and then about eight seconds later would hear it faintly again. Since it takes only one-seventh of a second for radio waves to travel all round the Earth, some strange extraterrestrial phenomenon seemed to be at play. The measured echo delay was consistent with an echo from something nearly as far away as the Moon. One candidate for such reflections would be an accumulation of interplanetary flotsam at the Lagrangian point L1, where the gravitational fields of the Sun, Earth, and Moon cancel to zero. An object at such a point is in unstable equilibrium; if it moves away a little, it will keep going. But particles moving under the net gravitational field of three attractors will slow down as they pass the equilibrium point; therefore an accumulation of matter might be expected in that vicinity from time to time. Efforts to observe long-delay echoes by direct experiment were conducted at the Cavendish Laboratory in the late fifties but with no success. However, reports of long-delay echoes had come from French naval ships in French Indochina. Having in mind the earlier thinking, George calculated the elevation angle above the horizon at the times and dates of the detection of echoes. This was a most ingenious idea; he found that on all occasions the point L1 was above the local horizon. This investigation was well worthy of publication in an international scientific journal, an exercise that was not on George's list of priorities.

He was a jovial, good-natured, and inspiring friend and is very much missed.

RONALD N. BRACEWELL
Termin Professor of Electrical Engineering
Stanford University
REVIEW ESSAY

Stagnant Science: Why Are There No AIDS Vaccines?

HENRY H. BAUER

Professor Emeritus of Chemistry & Science Studies
Dean Emeritus of Arts & Sciences
Virginia Polytechnic Institute & State University
e-mail: hhbauer@vt.edu


AIDS Vaccine Research by Flossie Wong-Staal and Robert C. Gallo (eds.). Marcel Dekker, 2002. 342 pp, $165.00 (hardcover).

In 1984, the Secretary for Health and Human Services (Margaret Heckler) announced that Robert Gallo had discovered the virus—that later designated HIV, the human immunodeficiency virus—that causes AIDS. On Gallo’s advice (Cohen, p. 8), she forecast that a vaccine would likely be available in a couple of years.

More than two decades later, there is no vaccine. Nor has there been credible progress toward a vaccine. Scores of attempts using a variety of approaches have all failed to show promise—or, rather, successive claims greeted initially as promising have all failed to bear fruit. Two books published in 2001 set out to describe for a general audience the search for a vaccine; an edited volume published in 2002 addresses a specialist audience.

The salient question is, why has so much effort failed so resoundingly? Cohen sees the answer in institutional and organizational terms: there has been no coordinated effort drawing on every idea and experience. By contrast, Thomas finds the answer in contingent personal experiences and fluctuating commercial demands that happened to sabotage various efforts at critical times; her book ends with readers left hanging as to the impending results of a large trial, but it is now known that the tested vaccine showed no sign of efficacy whatsoever. The technical specialists collected by Wong-Staal and Gallo cannot agree on what approach might work, but they do agree on the central difficulty: HIV mutates so
prodigiously that even "Within a single HIV-1 infected human host, HIV-1 population represents a complex mixture, or swarm, of mutant virus variants, in which all viruses are genetically related yet virtually every virus is unique".

Cohen and Thomas describe the mutability in less drastic terms, mentioning only that designers of potential vaccines need to make an initial decision as to which of the clades of HIV—A through E—a vaccine is intended to counter, or which mixture of 2 clades; B is the most common in the USA, E in South-East Asia; though Thomas (p. 161) mentions a "Thai E" with "A-type innards . . . encased in an E-type envelope". The fact remains that no vaccine has shown efficacy against even a single clade. ("Clades" are much the same thing as "species" or "sub-species". For a fascinating discussion, see David Hull's magisterial, sadly neglected account of the origins of this concept2.)

In principle, the possible types of vaccine—used against polio, for example—include:

- "Whole killed virus": HIV inactivated in some way.
- "Attenuated" virus: HIV reproduced through stages that progressively weaken it.

Most researchers regard these as too dangerous: not all the virus might be inactivated or attenuated, or it might re-activate in the body. So most efforts have been directed toward stimulating production of antibodies that might neutralize the virus, or stimulating production of immune-system killer cells that could recognize and attack cells infected with the virus, or finding ways to safeguard cells against entry of virus.

Mainstream discussions make it appear that much is known about the composition and structure of the virus, and Cohen and Thomas echo that uncritically: "AIDS researchers have turned the virus inside-out and carefully detailed how it destroys the immune system"3; "molecularly cloned . . . HIV—in other words, pure virus" (Cohen, p. 125); "epitopes of the viral envelope are too variable. Furthermore, the functionally important epitopes of the gp120 protein are masked by glycosylation, trimerisation and receptor-induced conformational changes making it difficult to block with neutralising antibodies"4. Yet the plain fact of the matter is that all these specifics are mere inferences based on indirect experiments on mixtures of substances under complex protocols: pure HIV has never been isolated. All published electron micrographs of "isolations" of HIV reveal a motley array of different-sized particles, only some of which have the shape and size of a retrovirus5. "Cloned" virus is synthetic RNA assumed, inferred, to be a "clone" of actual viral RNA—begging the question, of which strain of which clade?

In addition to the direct evidence of electron micrographs, clues have long abounded, that "isolates" are actually mixtures; for example, that "the viral surface" supposedly contains not only viral protein but also proteins from the cells from which the virus was thought to have budded (Cohen, pp. 131–2). But
these books leave no room to doubt orthodox HIV/AIDS theory: "August 14, 1984 ... scientists had conclusively proven three months earlier [that HIV] was the cause of AIDS" (Cohen, p. 200). Not only is Cohen unreliable in this manner as to scientific substance, he is also Panglossian — to put it most mildly — on the ethics of research on humans. Referring to clinical trials that have been universally condemned as unethical, Cohen writes: "Zagury's Zairian trials, although no one would dare say it, had a positive impact on the field ... : without, apparently, hurting anyone, they offered a lesson to an inexperienced world about how to identify and prevent unethical AIDS vaccine trials" (Cohen, p. 339). The world did not need Josef Mengele to discover that experimenting on live human beings is abhorrent.

Contradictions of logic or fact in HIV/AIDS research go unremarked in these books. "HIV tests" have been (until recently, exclusively) tests for antibodies — the presence of antibodies was taken as sign of active infection, not immunity; yet the initial hoped-for sign that a vaccine might be effective in producing immunity is the production of antibodies. Furthermore, Robert Gallo sneered at Peter Duesberg's statement that antibodies in healthy people are generally taken to show that an infection has been encountered and defeated. The mainstream's left hand does not know what the right hand is doing, apparently, as to antibodies and vaccines.

Lacking scientific insight, these books nevertheless have points of historical interest; verging at times on prurient interest, as when remarking that the former Secretary for Health and Human Services was not aware that gay sex may involve anal intercourse (Cohen, pp. 6–7). It may be useful to have Robert Gallo on record with the absurd claim that "Every single retrovirus in every single species, from chicken to man, causes disease ... 'Every. Single. One.'" (Cohen, p. 346). One learns how Donald Francis became one of Gallo's arch enemies (p. 61); and about Daniel Zagury's irresponsible vaccine initiatives in Africa, carried on in collaboration with Gallo (p. 66 ff.); about Fauci's bureaucrat-typical behavior (Cohen, pp. 133 ff., 189; Thomas, pp. 300, 314); and much about the statistically incompetent claims of Army researcher Robert Redfield, MD (Cohen, p. 158 ff.; Thomas, p. 168 ff.). There are copious illustrations of the intemperance of all-purpose gadfly John P. Moore (Cohen, pp. 170, 258–9, 274–5, 286; Thomas, pp. 285–6, 298, 301, 365, 378). Bureaucratic infighting within and between the HIV researchers of the Army and of the National Institutes of Health, and conflicts with commercial interests, are described at some length by both Cohen and Thomas.

A few points of substance are worth noting. Chapter 10 in Cohen's book describes the many people known to have been in frequent sexual contact with HIV-positive people without becoming positive themselves, including several percent of a group of prostitutes in Nairobi (Kenya). Experts estimate a rate of new infections of 2% in the high-risk group of urban gay men (Cohen, p. 250), which should surely raise eyebrows in wonder as to how so low a rate could have brought on a suddenly exploding epidemic. In several places, there are
illustrations of the salient difference between scientists and MDs (Cohen, pp. 100, 344), the latter being more ready to try, on hunches, things that might work, whereas the former seek to understand what is going on. Clinical trials are carried on in Africa and Thailand that would not be approved in the United States (Thomas, p. 370). The National Institutes of Health "has a formula for constituting committees so that women, minorities, a spectrum of academic institutions, and different parts of the country are represented" (Thomas, p. 246), bringing to mind the late Senator Hruska’s assertion: "Most of the American people are mediocre. And they have a right to be represented on the Supreme court 7".

It is dismaying to find that Cohen, who has covered HIV/AIDS for years for Science, is so brainwashed by orthodox views that by hindsight he dismisses as "fanciful" and "far-flung" perfectly sensible ideas as to the possible causes of AIDS that were bruited when the illness was first noted (Cohen, p. 9). He also misleads (p. 10) by saying that Gallo had isolated the virus from 48 patients without mentioning that those 48 were only one third of the sample; Gallo had failed to "isolate" virus from two thirds of AIDS patients. Cohen states sheer speculation as fact: "HIV's gp160 uses a secret handshake ... opening the cellular doors" (p. 47)—but to this day, the scientific literature does not describe a proven mechanism by which HIV attacks and kills cells. Cohen's answer to the lack of progress is naïve: "inability of the primate research world, the NIH, and industry to work together" (p. 81). If anyone anywhere were to demonstrate the merest credible inkling of substantive progress, cooperation would immediately follow.

Thomas, too, is substantively misleading on important points. She reports (p. 187) that AIDS cases doubled in a single year (1993) in the USA. Instead of pointing out that this was a statistical artefact reflecting a changed definition of AIDS, she describes it as follows: the "CDC was finally counting women more accurately" (p. 188). Counts of CD4 cells are said to correlate with clinical condition (pp. 223-4) even though previously published reports contradict that (and confirmation of those earlier reports accumulates: the now-standard "cocktail" HAART therapy does increase CD4 counts and decrease viral load, but the treated patients fare worse, moreover, there is no correlation between CD4 counts and viral load). Thomas (p. 395) also adopts the egregiously inappropriate epithet of "holocaust deniers" for those, like the British Sunday Times, who dare point out that the facts do not support the notion that HIV causes AIDS.

The technical experts assembled in AIDS Vaccine Research can be just as misleading as the journalists. "It seems paradoxical that a human pathogen that medical science knows more about than virtually any other causative agent remains without a solution", according to the volume's editors (p. iii). Paradoxical only if one is not aware that all this claimed knowledge does not include how HIV is supposed to kill cells, nor what a genuine particle of native virus looks like. The chapter titles in this book illustrate that their content is hunches, wishful thinking, and padding, not useful knowledge. Chapter 1,
"AIDS vaccines: challenges and opportunities", is clearly enough about things waiting to be tried. The title of Chapter 2, "Immunopathogenesis of HIV infection", is misleading, however, because the text describes no proven mechanism: the mass of detail is all speculation, not knowledge. Chapter 3, "The genetic diversity of HIV-1 and its implications for vaccine development", already cited above—"virtually every virus is unique"—carries the clear implication that no vaccine may be feasible. Chapter 4, "The role of cytotoxic T lymphocytes in protection against HIV infection and AIDS", argues that killer cells rather than antibodies might be a better goal for vaccinology, but again this is speculation and not demonstration. Chapter 5, "Immune reconstitution in HIV infection", describes the phenomenon that patients do less well under therapy even though their blood tests are "better". Chapter 6 speculates about the "Design of engineered vaccines for HIV", though once again none of the attempts along that line have been successful, nor have those with the "DNA vaccines for immune deficiency viruses" discussed in Chapter 7. Other unproven approaches are the "Replication-deficient, pseudotyped HIV-1 vectors" of Chapter 8 and the "Mucosal DNA vaccines" of Chapter 9. Chapter 10, "Innate immunity in HIV infection", acknowledges that large numbers of people exposed to HIV never become infected, and that large numbers of HIV-positive people never progress to AIDS; yet many years of study have not revealed what properties of their immune systems have apparently rendered them impervious. "The role for non-human primate models in the development and testing of AIDS vaccines", Chapter 11, would better be, "The lack of any role . . .", because no monkey or ape model for studying HIV infection exists despite numerous attempts to find one. The concluding Chapter 12, "International perspectives on HIV vaccine development", is a shopping list by the director of the International AIDS Vaccine Initiative, specifying how cheap as well as effective a vaccine needs to be.

Like the journalists, the experts can be glib where details seem called for, say, when reporting results on "newly infected individuals" (p. 96). How could a useful sample of such people ever have been assembled? It would require continual testing of many people, waiting for them to become infected, as well as coping with such complications as that seroconversion—becoming HIV-positive—is only supposed to occur at some ill-defined time following infection: weeks or months, up to a year. Again, directly contradicted by the abundant examples of spontaneous reversion to HIV-negative status, including among the majority of children born HIV-positive and among recovering drug addicts, "No HIV-infected individual has yet been shown to eradicate the virus" (p. 97).

The lack of progress toward a vaccine has been accompanied incongruously by a refrain of optimistic utterances. In 1989, Fauci praised "a giant step forward for AIDS vaccine research" (Cohen, pp. 92–3). In 1990, a leading researcher opined that "we've cracked open the door on the optimism of a vaccine" (Cohen, p. 100). In 1997, "it is no longer a question of whether we can develop an AIDS vaccine, it is simply a question of when" (Thomas, p. 317). In Wong-Staal and Gallo one encounters "HIV vaccine development . . . poised to make
significant advances" (p. 6), "growing optimism . . . that an AIDS vaccine may be possible" (p. 121), "great promise in the field of HIV vaccinology" (p. 311). But privately held opinions have been less sanguine: it had taken nearly 2 years to hire a leader for the vaccine project at the National Institutes of Health (Cohen, p. 332). There have also been serious discussions over highly hypothetical matters: for example, if a reasonably successful vaccine eventuates, would that impair the quest for genuine cures for AIDS (Cohen, p. 242)?

Like other aspects of HIV/AIDS activity, lack of substantive achievement is coupled with economic benefit to a host of people and organizations. The International AIDS Vaccine Initiative provides many employment opportunities13, as of October 2006, there were at least two dozen further openings for policy assistants, grants analysts, communications personnel, writers, research associates, and more!4.

These books tell much about the contemporary state not only of HIV/AIDS but also of book publishing. Cohen and Thomas are not writing popularized science: readers learn nothing useful about scientific aspects of HIV/AIDS through this discussion of a quest that has yielded no successes. Nor do these writers take the obvious opportunity to ask whether there may be a fundamental flaw to a quest that gets nowhere. Cohen’s book is replete with words that used to be eschewed by well-brought-up people, words in no way essential to his story. Thomas’s book could serve as a text to illustrate what is meant by over-writing: inept similes and metaphors abound, all the main actors are referred to by their first names throughout, and the focus is almost exclusively on "human interest"—one reads much about researchers' children, illnesses, career problems. Evidently this is what publishers believe can sell books. Fortunately, they are sometimes wrong; amazon.com reveals that the Thomas book was soon remaindered. The Staal-Gallo volume, of course, represented guaranteed sales to libraries simply because of the subject and its celebrity editors.

Was there ever a golden age where publishers were concerned primarily whether a book was good, believing that if it was, it would earn its keep?

As to HIV/AIDS, once it can no longer be denied that Duesberg and other "dissidents" were right all along, Leo Szilard will (once more) prove prophetic: "They'll never forgive you for being right" (Cohen, p. 220).

References


REVIEW ESSAY


DAVID DEMING
College of Arts and Sciences
University of Oklahoma
e-mail: ddeming@ou.edu

Introduction
The Man Who Predicts Earthquakes purports to be the story of California geologist Jim Berkland. Berkland first attracted notoriety when he successfully predicted a magnitude 6.9 earthquake that struck the San Francisco Bay area on October 17, 1989. The prediction was published in a local newspaper, the Gilroy Dispatch, on October 13. Berkland had specifically forecast that an earthquake with a magnitude between 3.5 and 6.0 would occur between October 14 and the 21st.

Berkland was rewarded for his successful prediction by being summarily suspended from his job as Santa Clara County (California) geologist. After two-and-a-half months, he was allowed to return to his duties, but with the admonishment that he cease predicting earthquakes on government time. On January 18, 1990, The San Jose Mercury News ironically noted that "Berkland's return to work coincided with [another] successful quake prediction" (Calvert, 1990).

Jim Berkland is a real geologist. According to the biographical sketch given on his website (www.syzygyjob.com), he holds an Masters degree in geology from San Jose State University. His employment history includes six years with the U.S. Geological Survey (USGS) and twenty years as a geologist with Santa Clara County in California. A search I conducted using the GEOREF database brought up 25 publications in the scholarly geological literature authored by James O. Berkland.

Earthquakes
The cause of earthquakes is one of the great geological questions that has intrigued people for thousands of years. In Meteorologica, Aristotle (384-322 BC) attributed the cause of earthquakes to the movement of wind within the Earth: "not water nor earth is the cause of earthquakes but wind" (Aristotle,
1923). The Roman Senator, Lucius Annaeus Seneca (3 BC–65 AD), also supposed that earthquakes were caused by the sudden and violent movement of air within the Earth: "No one, I suppose, can doubt that there is nothing so restless, so capricious, so fond of disturbance as air" (Clarke, 1910, p. 245).

Since the advent and acceptance of plate tectonic theory c. 1963–1970, it is now accepted that the ultimate cause of most earthquakes is the dissipation of stresses generated by the interaction of tectonic plates. The standard explanation for the immediate occurrence of earthquakes is known as the elastic rebound theory. The theory was originated by Harry Fielding Reid (1859–1944), a professor of geology at Johns Hopkins University who was called upon to investigate the great 1906 earthquake in San Francisco. Reid believed that an earthquake resulted from the gradual buildup of stress within the Earth and its sudden release in an elastic rebound.

It is not going too far to say that whenever ruptures occur, they result from elastic strain, and the sudden movements produced are merely elastic rebounds; and moreover, except in the case of earthquakes connected directly with volcanic action, the strains have not been set up suddenly, but are gradually developed by the slow displacements of adjacent areas. (Reid, 1910, p. 31)

USGS geophysicist Ross Stein has constructed a marvelous mechanical analog that illustrates the buildup and release of stress that occurs in earthquakes. His simple machine consists of a wooden board covered with sandpaper, about a meter in length, with a winch at one end. A length of elastic cord, such as surgical tubing, is attached to the winch. The other end of the elastic cord is affixed to a brick resting on the sandpaper-covered surface of the board. As the winch is wound, the cord tightens, increasing the stress on the brick. Eventually, the tension in the cord exceeds the frictional resistance between the brick and sandpaper, and the brick suddenly moves, releasing the stress in the cord through an elastic rebound. If the winch continues to be operated the cycle repeats itself, but neither the timing or length of brick movements is predictable.

**Earthquake Prediction and Precursors**

Although the buildup of stress that eventually produces a rupture and movement may occur at a uniform rate, earthquakes are as unpredictable as the movements of the brick in Ross Stein's earthquake machine. To be sure, we must differentiate between long-term and short-term prediction. We can state with certainty that the continual increase of stress in the elastic cord will result in the eventual movement of the brick. But it is impossible to predict precisely when that movement will occur. It is a virtual certainty that major earthquakes will continue to occur in the vicinity of the San Andreas Fault Zone in California. But predicting the date of specific events is highly problematical.

The USGS, an organization noted for its credibility and dedication, has had its own embarrassing debacle with earthquake prediction. In 1984, the USGS
predicted that a major earthquake would occur on the San Andreas Fault in central California near the town of Parkfield by 1993. The prediction was based on a simple extrapolation of historical experience. Parkfield had experienced six major earthquakes between 1857 and 1966, with an average recurrence interval of 22 years (Scholz, 1997). In anticipation of collecting data, the USGS set up a rich array of experimental apparatus near Parkfield. But the anticipated temblor did not arrive until September 29, 2004, eleven years after the predicted deadline (Perlman, 2004).

Short-term earthquake prediction is a black art, based on the interpretation of indicators known as precursors. The simplest example of an earthquake precursor is a foreshock, a small temblor that precedes a larger event. Changes in stress and strain that precede an earthquake can also produce changes in hydrological parameters such as water levels in wells, stream and spring flow rates, and the periodicity of geyser eruptions (Roeloffs, 1988; Deming, 2002).

More contentious is the claim that earthquakes can be preceded by electric currents in the Earth, or changes in the terrestrial magnetic field. Physicist Panayiotis Varotsos and his colleagues in Greece have claimed to be able to make successful earthquake predictions by monitoring electrical currents in the crust (Kerr, 1999). The idea is based on the phenomenon of piezoelectricity, discovered by Pierre and Jacques Curie in 1880. When mineral crystals such as quartz are subjected to differential or changing stresses, mechanical energy can be converted into electricity.

Perhaps the most controversial earthquake precursor is changes in animal behavior. Although the concept smacks of pure superstition, there are physiological reasons that make such precognitions theoretically possible. The sensory capabilities of animals may exceed human beings in several areas, including the perception of odors, low-frequency sounds, and sensitivity to electromagnetic waves (Kerr, 1980).

Western scientists have traditionally categorized short-term earthquake prediction as impossible, and its practitioners as fakers. The attitude is epitomized by a well-known quote from the Dean of American Seismologists, Charles F. Richter (1900–1985).

I am often asked about prediction. Since my first attachment to seismology, I have had a horror of predictions and of predictors. Journalists and the general public rush to any suggestion of earthquake prediction like hogs toward a full trough. It is a parallel to the obsession with a cure for cancer, or with the question of life on other worlds. There is nothing wrong with aiming toward prediction, if that is done with common sense, proper use of correct information, and an understanding of the inherent difficulties. Otherwise, the subject provides a happy hunting ground for amateurs, cranks, and outright publicity-seeking fakers. The vaporings of such people are from time to time seized upon by the news media, who then encroach on the time of men who are occupied in serious research. (Richter, 1977, p. 1246)

Richter’s characterization of earthquake prediction seemed to strike home in 1990 when iconoclast Iben Browning (1918–1991) predicted a major earthquake
would occur on the New Madrid Fault Zone on December 3, 1990. Although the midwestern U.S. is tectonically quiescent, there is a section of the Mississippi Valley that has the potential for devastating earthquakes. During the winter of 1811–1812, the largely unpopulated New Madrid area in the southeast corner of the State of Missouri was struck by three enormously powerful earthquakes. The classic description of the destructive power of these temblors was made by Myron L. Fuller in USGS Bulletin No. 494.

The ground rose and fell as earth waves, like the long, low swell of the sea, passed across its surface, tilting the trees until their branches interlocked and opening the soil in deep cracks as the surface was bent. Landslides swept down the steeper bluffs and hillsides; considerable areas were uplifted, and still larger areas sunk and became covered with water emerging from below through fissures or little "craterlets" or accumulating from the obstruction of the surface drainage. On the Mississippi great waves were created, which overwhelmed many boats and washed others high upon the shore, the return current breaking off thousands of trees and carrying them out into the river. High banks caved, and were precipitated into the river, sand bars and points of islands gave way, and whole islands disappeared. (Fuller, 1912, p. 10)

Browning, the publisher of a business newsletter who held a Ph.D. in zoology, based his prediction on the idea that strong tidal forces could trigger an earthquake (Booth, 1990). But there was no earthquake, and Browning was subsequently criticized for unnecessarily alarming people.

Earthquake prediction enjoyed a short-lived renaissance in the West during the latter half of the 1970s. In February of 1975, Chinese scientists made a successful short-term prediction of a major earthquake in Liaoning Province of the People's Republic of China. The warning was issued a mere 13 hours before a magnitude 7.3 earthquake destroyed the city of Haicheng. Western seismologists were impressed. The journal Science characterized the prediction as "the first known instance of a major quake successfully predicted and disaster prevented or mitigated on such a scale" (Hammond, 1976, p. 538). One of the striking aspects of the Haicheng earthquake was the report of strange animal behaviors, apparently validating the idea that animal behavior could predict imminent seismic disturbances.

Some instances noted at this time [before the earthquake] were of snakes being found frozen on the road ... geese flying, chickens refusing to enter their coop, pigs rooting at their fence, cows breaking their halters and escaping, and goats as well as cows being unusually restless. Rats appeared to behave as though drunk. Three well-trained police dogs howled, refused to obey commands, and kept their noses to the ground as though sniffing. (Molnar et al., 1977, p. 254)

But the search for reliable earthquake precursors has been elusive. Post mortem examination of the 1975 Laicheng prediction suggested that the "prediction" had been made after the fact, and that an intense series of foreshocks prior to the main earthquake had saved lives by inducing spontaneous evacuations.
(Geller et al., 1997). In recent years, seismologists have become very pessimistic regarding the possibility of making reliable short-term earthquake predictions. In a 1997 Science article frankly titled "Earthquakes Cannot Be Predicted", the authors stated that reliable precursors probably did not exist and that it was difficult to formulate and test earthquake-prediction hypotheses.

Empirical earthquake prediction would require the existence of observable and identifiable precursors that would allow alarms to be issued with high reliability and accuracy. There are strong reasons to doubt that such precursors exist. Thousands of observations of allegedly anomalous phenomena... have been claimed as earthquake precursors, but in general, the phenomena were claimed as precursors only after the earthquakes occurred... each new claim brings a new set of proposed conditions, so that hypothesis testing, which is what separates speculation from science, is impossible. (Geller et al., 1997, p. 1616–1617)

Jim Berkland bases his earthquake predictions on three primary methods: tidal forces, hydrologic disturbances, and notices of missing dogs and cats published in local newspapers.3 There is nothing unreasonable about using any of these indicators as earthquake precursors. The upper crust of the Earth is a porous medium, and changes in stress are likely to be reflected in hydrologic parameters. The existence of such precursors is extensively documented in the scientific literature (Roeloffs, 1988). And counting missing dogs and cats is a straightforward and repeatable way of quantifying "animal behavior," something that is otherwise elusive and subjective.

Other than the newsletter he publishes, Berkland has never documented his methodology in publication.4 But the primary precursor he uses in his predictions appears to be tidal forces. Tidal forces on Earth are a maximum at syzygy, the time when the Sun, Moon, and Earth are in a straight line and tidal forces on the Earth are a maximum. Syzygy is synonymous with the full and new Moon phases.

There is nothing new in the idea that earthquakes can be triggered at syzygy by tidal forces. In 1912, USGS geologist Myron L. Fuller noted that the strongest shocks of the 1811–1812 series of New Madrid earthquakes occurred during new or full Moon phases.

The attraction exerted by the sun and moon on the earth's surface at times of new and full moon is decidedly greater than during the intermediate periods and has been thought to be an appreciable factor in determining the times of earthquakes... With one marked exception the groups [of major earthquake shocks] occur[ed] approximately either at times of new or full moon. (Fuller, 1912, p. 37–38)

There is no consensus in the scientific literature as to whether or not tidal forces can act as earthquake triggers. Some studies have had positive results, others negative. In 1975, geophysicist Thomas Heaton concluded "there is good
reason to believe that larger ... earthquakes are tidally triggered" (Heaton, 1982, p. 2181). But a follow-up study published in 1982 failed to confirm the initial work, "analysis ... fails to confirm an earlier hypothesis that the origin times of ... earthquakes correlate with solid-earth tidal shear stress" (Heaton, 1982, p. 2181).

In 1981, geologists at the USGS published an apparent refutation of Berkland's claim that earthquakes could be predicted by using tidal forces. It has been suggested by geologist James Berkland ... that the eight-day period commencing with each syzygy (new or full moon), constitutes an "earthquake-prone period" for the San Jose area. These eight-day periods, called "seismic windows", form the basis for this system of earthquake prediction. Berkland requested that the U.S. Geological Survey perform a computer evaluation of his prediction method. The results from this study conclusively indicate that the seismic window theory fails as a reliable method of earthquake prediction. (McNutt and Heaton, 1981, p. 12, 16)

Scientists at the USGS have been Berkland's most outspoken critics. In 1990, a USGS seismologist told the Wall Street Journal "we've known this guy throughout his professional life, and he's distinguished himself by being a clown the entire time" (Dolan, 1990, p. B1).

But there is evidence that tidal forces can trigger earthquakes. In 1990, the journal Geology published a paper with the self-explanatory title "Strong Correlation of Major Earthquakes with Solid-Earth Tides in Part of the Eastern United States" (Weems and Perry, 1990). More recently, researchers in Japan found that tidal forces could trigger earthquakes, but "only when the tidal stress ... acts in the same direction as the regional tectonic stress" (Stein, 2004, p. 1248).

The question has not yet been settled, nor are scientific controversies ever settled conclusively.

Berkland's methods are scientific, in the sense that there is some empirical evidence for the precursors on which he bases his predictions. If American seismologists can swoon over the insightful use of animal behavior as an earthquake precursor by the Chinese, I don't know why it's unreasonable for Jim Berkland to count newspaper notices of missing dogs and cats.

This is not to say that Berkland is capable of accurately predicting earthquakes. The claim that he successfully predicted the Loma Prieta Earthquake in 1989 can be attacked on two fronts. First, Berkland predicted the maximum magnitude of the temblor would be 6.0, but it was 6.5. A magnitude 6.5 earthquake is much larger than a magnitude 6.0. Earthquake magnitude scales are logarithmic, and a magnitude 6.5 earthquake has approximately five times the energy of a magnitude 6.0. Second, Berkland has been predicting earthquakes since 1974. He was bound to get one or more predictions right, just by chance. This is known as the Texas Sharpshooter's Fallacy. The Texas Sharpshooter randomly shoots holes in the side of a barn. He then walks up to the barn and paints a bullseye around one hole and proclaims himself a sharpshooter. To
evaluate the accuracy of Berkland's methodology his predictions must be analyzed in a sound statistical sense.

Berkland himself can be frustrating and disappointing. In a telephone interview (January 2, 2007), he cited anecdotal evidence to me obsessively. He could not produce a document as simple as a listing of the scientific literature on earthquakes and tidal forces. He claimed that attempts to publish his ideas in the scientific literature resulted in rejection. This is not difficult to believe. Anyone who works in the area of anomalies is well aware that manuscripts that fail to conform to arbitrary consensus positions are apt to be rejected, no matter how high the scientific quality. But Berkland has not done so much as self-publish a pamphlet. He claimed that documentation existed in the form of his newsletters, but I found this answer less than satisfying. Berkland was quick to cite his persecution by the scientific establishment, but that does not validate his ideas.

Kooks and Kookery

Whatever the case may be with Jim Berkland, *The Man Who Predicts Earthquakes* is an execrable book. Setting aside the poor, disorganized prose, the primary problem with *The Man Who Predicts Earthquakes* is the conflation of science with kookery.

The second edition of the *Oxford English Dictionary* (OED) defines "kook" as "a cranky, crazy, or eccentric person." According to the OED, the first documented use of the word in the English language occurred in the *Daily Mail* newspaper in 1960, "A kook, Daddy-0, is a screwball who is 'gone' farther than most."

I am not sure that the dictionary definition of kook is sufficient, but I am aware that kooks and kookery are a significant problem for those of us attempting to use genuine scientific methods to investigate anomalous phenomena. In 2004, I published a manuscript in the *Journal of Scientific Exploration* on an anomalous sound known as the "Hum" (Deming, 2004). I also moderate an internet discussion forum dedicated to the scientific discussion of this phenomenon (http://tech.groups.yahoo.com/group/humforum/). Kooks are a persistent problem on the forum. I can't precisely define what constitutes a kook or kookery, so in practice I apply the standard originated by U.S. Supreme Court Justice Potter Stewart (1915–1985). In the case *Jacobellis v. Ohio* (1964), Justice Potter explained that although he could not precisely define "hard-core pornography," he knew it when he saw it.

Under the First and Fourteenth Amendments criminal laws in this area are constitutionally limited to hard-core pornography. I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description; and perhaps I could never succeed in intelligibly doing so. But I know it when I see it, and the motion picture involved in this case is not that. (*Jacobellis v. Ohio*, 1964)

Although there is always some ambiguity involved, I can usually recognize a kook when I see one. When they pop up on the Hum forum, I immediately ban
them. Many of these people write as if they have psychological or emotional problems. They are inexorably attracted to anomalies of all types, but their behavior is fundamentally irrational. A kook has little to no understanding of how to think scientifically or perhaps even logically. They lack the mental discipline that is necessary to do science, and are unwilling to accept or evaluate facts that may contradict preconceived ideas. Scientific thinking is a discipline that the average person is not prepared to accept.

Science is not a little repellent to the average man ... anything sensational, abnormal, shocking, impossible, fascinates him. Irrationality must necessarily appeal to beings who are normally irrational. Science is a joykiller. (Sarton, 1924, p. 81)

There is nothing more harmful to the legitimate scientific investigation of anomalies than for the field to become littered and infused with kooks. On internet discussion forums, these people relentlessly drive out good posters, and ruin everything they come into contact with. They need to be condemned swiftly and mercilessly.

*The Man Who Predicts Earthquakes* is rife with kookery. On page 46, the reader is advised: "Next time your pooch howls at night, give Rover a doggie treat to reinforce psychic behavior." On page 74, we are introduced to the "popcorn" precursor: "Got a headache or craving for popcorn? It may be that you like popcorn, or it could be a warning that an earthquake is coming." On page 86, the author describes a man who predicts earthquakes by listening to static on his television and radio. This nonsense is seamlessly fused and presented in the same discussion with legitimate precursors such as changes in groundwater levels or radon emissions.

Author Cal Orey describes herself as a "longtime fan and friend of Jim Berkland" (p. 47). With friends like this, who needs enemies? I can't imagine a more effective hatchet job on Jim Berkland than *The Man Who Predicts Earthquakes*. Berkland is discussed in the same vein as self-described psychics who hear voices. Pages 77 and 78 are devoted to discussion and promotion of a woman named Carol Shumaker who is otherwise known as "Calico the Psychic." One afternoon, Calico heard a voice say "China. 7.2." Two days later, there was a magnitude 7.2 earthquake in China. The reader is referred to Calico's website (www.calicothepsychic.com), where anyone can "talk live" with Calico for only $3.79 a minute. Calico also takes Mastercard, Visa, Discover, and American Express.

Who is Cal Orey, and how could she write such a terrible book? Prior to venturing into the world of seismology, her literary efforts were limited to such titles as *The Healing Power of Vinegar*, 202 Pets' Peeves: Cats and Dogs Speak Out on Pesky Human Behavior, and Lose 10 Lbs. (& 10 years off your age) In Just 2 Weeks! It is apparent from reading *The Man Who Predicts Earthquakes*, that the writer has no understanding of the limitations of anecdotal data, the nature of proof, or how science operates. She is completely lost. Potential
Earthquake Prediction, Kooks, and Syzygy

authors take note: writing articles on pets for the magazine *Woman’s Day* is probably inadequate preparation for comprehending the issues contemplated in Popper’s *Logic of Scientific Discovery* (1959).

In a revealing statement on page 155 of *The Man Who Predicts Earthquakes*, Orey confesses her inability to understand science. Discussing the disastrous tsunami that occurred in the Indian Ocean on December 26, 2004, Orey states that she was unable to distinguish whether the tsunami had been caused by an earthquake or "aliens from another planet."

I received dozens of emails that contained a variety of opinions about why this disaster happened. Some folks said it was the beginning of the end of the world; some said that deadly earthquakes are linked to human’s disobedience of God; others insisted that the quake-tsunami was linked to aliens from another planet or military explosions. I felt dazed and confused. I didn’t know whom to believe.

Under the same circumstances, I do not think most readers of this review would have had much trouble knowing "whom to believe."

In summary, *The Man Who Predicts Earthquakes* is a terrible book. It systematically misinforms the reader by conflating legitimate efforts to predict earthquakes with kookery. Writings like this discredit the entire field of anomaly investigation. This book in particular will be devastating for Jim Berkland's reputation. More importantly, the work discredits the entire field of earthquake prediction, an area where scientific progress has the potential to save human life.

Notes

1 I should disclose that I met Jim Berkland about ten years ago when I invited him to speak at the University of Oklahoma. Berkland gave an informative presentation on earthquake precursors.
3 I confirmed this in a telephone interview with Berkland on January 2, 2007.
4 Confirmed in a January 2, 2007, telephone interview.

References


REVIEW ESSAY

The Man Who Predicts Earthquakes—Jim Berkland, Maverick Geologist: How His Quake Warnings Can Save Lives by Cal Orey

PATRICK MCCLELLAN
557 Posada Way
Fremont, CA 94536
e-mail: p.mcclellan@comcast.net

Cal Orey has written numerous books and magazine articles over the past two decades on topics ranging from pets and cooking to natural health, including several short articles about Jim Berkland and his earthquake predictions. This is her first book on the "maverick geologist" or on the subject of earthquakes. Her choice of Berkland as a biographical subject is a good one, as he has remained a colorful, charismatic, and controversial character at the boundaries of earthquake science for over 30 years. Her timing is impeccable. With publication in early 2006, just months before California's centennial celebration of the Great San Francisco Earthquake of April 18, 1906, the book's wide exposure was assured.

Berkland established his credibility in the earth sciences during the 1960s and 1970s, as a geologist in the U.S. Department of Interior (the Geological Survey and the Bureau of Reclamation), then briefly as an Assistant Professor of Geology at Appalachian State University, and finally as the first official County Geologist for Santa Clara County (aka, "Silicon Valley"), where he served from 1973 until his retirement in 1994. He began predicting earthquakes in 1974. Since then, the notoriety of his predictions gained him a reputation among West Coast news media and non-scientists as an earthquake expert. Among "mainstream" seismologists his predictions are regarded as "simply amusing—unless you happen to be the U.S. Geological Survey (USGS) spokesperson repeatedly called upon to debunk them" (Scholz, 1997), and the delusions of a misguided dilettante, at best, or a "crackpot or clown" (Berkland, 2006) or "fraud" (quoting one who posted on Berkland's web forum), at worst.

Except for an American Geophysical Union meeting abstract a quarter-century ago, Berkland has chosen to self-publish his predictions and his basis for them through his subscription newsletters and on his website (www.syzygyjob.com), issue them through undocumented speaking engagements, and promote
them through the often-fallible filters of local news coverage, television documentaries, and talkshow broadcasts. He eschews publication in peer-reviewed journals, asserting that "high science" unfairly blocks his message through an ostensible conspiracy of prejudicial reviewers with unrealistic research standards, as the book suggests.

Therefore, this book is likely as close as Berkland (now in his late 70s) will come to a formal publication of his theory of earthquake prediction. Insofar as he provided Orey with the factual material for her research and informed her perspective, reviewed her manuscript, and now, along with the author, tours bookstores to support the book's sale, we can safely assume that this treatise closely represents his scientific viewpoint. For that reason, I take this opportunity both to review Orey's book and to briefly assess Berkland's theory, to the extent possible.

I approached the book as a sympathetic observer. Like Berkland, I have an interest in pre-earthquake anomalies, including some that have gotten short shrift by the mainstream over the past three decades (e.g., McClellan, 1980, 1984, 1986). I am also sensitive to the politics and fashions that have influenced U.S. earthquake prediction research. Pre-earthquake anomalies are notoriously inconsistent. Research dollars are limited. Grant review committees giveth and taketh away, not always with cold objectivity. During the first few years of the USGS prediction research program (in the late 1970s and early 1980s) some "premonitory" phenomena, such as unusual animal behavior, received scant funding and then were summarily discredited. In contrast, other phenomena, such as seismicity, fault creep, crustal strain, ground tilt, ground-water fluctuations, soil radon emissions, and magnetic field fluctuations, continued to receive millions of dollars in funding, year after year, into the 1990s and beyond—in spite of proving equally useless for prediction. The institutional bias seemed to define research "success potential" as the pursuit of purely geophysical precursors and I found it galling, both as a scientist and a U.S. taxpayer. So, I share Berkland's frustration—to a degree.

The Book

The volume's contents are organized into six parts of several chapters each: (1) "The Predictor", about Berkland, the development of his "seismic window" theory, and the premonitory muses he consults (e.g., the unusual behavior of animals, geysers, rainfall, etc.); (2) "The New Age Movement", covering human "earthquake sensitives" and the physiological symptoms they relate to seismic activity; (3) "Shakers and Seismic Windows", applying Berkland's theory to regional earthquake activity in the U.S. and around the world; (4) "The Big Wave", reviewing anomalous phenomena reportedly observed before the Indian Ocean tsunami of 2004; (5) "Politics and Quakes", addressing the clash of mainstream science and government, on one hand, with unofficial and unpopular views of earthquakes past and future, on the other hand; and (6) "Bracing for the
Review of *The Man Who Predicts Earthquakes*

Big One", a primer on preparedness and applying Berkland's lessons to save lives in the earthquakes yet to come.

A seven-page Foreword written by psychobiologist and author, David Jay Brown, sets the tone. Brown, who has himself explored pre-earthquake animal behavior (see his website at www.animalsandeathquakes.com) and done background research for books by biologist and philosopher, Rupert Sheldrake, adds his own interesting views about Orey's subject.

In general, the book is fairly well organized, with a limited glossary, brief bibliography, and serviceable index. I found the section and chapter titles to be annoyingly cute and trendy, but perhaps understandably so, considering the author's background in "pop" magazine literature and her non-technical target audience.

**Berkland's Earthquake Prediction Theory**

The cornerstone of Berkland's proposition is his "seismic window theory", introduced by subtitle on page 24 but never concisely or completely explained. Inferring from the clues scattered across the chapters, his thesis is this: Earthquakes worldwide are more likely to happen near times of the maximum fortnightly tide at new and full moon (i.e., "syzygy") than at other times, especially near the time of a lunar or solar eclipse. And at those times, quakes are most likely when the moment of syzygy closely coincides with the moment of the moon's closest approach in its elliptical orbit around the earth (i.e., "perigee"), which slightly strengthens the maximum fortnightly tide. Connecting syzygy to earthquakes, Berkland contends, is the periodic tidal force on faults which are already stressed to near the point of slipping. The extra applied stress around syzygy ultimately triggers these near-critical faults to slip and causes earthquakes. (It is an old notion, which he freely acknowledges.)

To bracket his intervals of elevated earthquake likelihood, Berkland recognizes a "seismic window" eight days in length, which begins from one to three days before syzygy. (The lunar cycles of syzygy and perigee, while fortnightly, differ somewhat in period, hence, the time between these two events differs with each fortnight—and the start of Berkland's eight-day window does likewise, by up to a day or two.) Of the two syzygies per month, the one nearest perigee raises the greater tides and defines Berkland's "primary" seismic window. The other window is "secondary". As a proxy for orbital physics, his reference for the timing and efficacy of the seismic windows is the tide table for San Francisco's Golden Gate. The greatest twice-daily tides there are about 9.0 feet in tidal range and occur, he states, when syzygy and perigee are less than 25 hours apart (called "perigean spring tides"), which happen just two to five times each year.

On that foundation, he formally predicts for each primary window: (1) one or more earthquakes of specified magnitude (M) with an epicenter located in each of three U.S. West Coast regions, and (2) one or more earthquakes globally "of 7.0-plus... most likely around the Pacific Ring of Fire where 80 percent of the
world's strongest quakes hit" (p. 253–254). The West Coast regions are: (a) within a radius of 140 miles (approximately two degrees) from Los Angeles City Hall (34N, 118W), (b) within 140 miles of Mt. Diablo in the San Francisco Bay area (37.9N, 121.9W), and (c) anywhere in Washington or Oregon. In those U.S. regions, he predicts earthquakes of 3.5M to 5.5M (or up to 6.5M depending on region and perigean spring tide condition).

So Far, So Clear

But then, as other indicators suit him, Berkland complicates his theory with additional inference from (1) unusual animal behavior (i.e., counting lost and found animals in local newspaper ads), (2) misbehaving geysers and hot springs, (3) extreme seasonal rainfall, (4) reports of symptoms from human "seismic sensitivies", (5) magnetic fluctuations, (6) seismic quiet periods, (7) personal intuition (his so-called "MOSS predictions", or "monthly outright seismic speculations"), etc. He grants special exemptions for quakes that violate his syzygy theory in certain situations (near California's Salton Sea, for example), and modifies his "standard" syzygy predictions ad hoc (e.g., increasing the magnitude, or shifting the time) as he deems the present circumstances warrant.

As candidate pre-earthquake anomalies, these additional and disparate sources of inference have scientific merit in their own right and each deserves a proper hearing. However, Berkland accounts for none of them scientifically, offering speculations and anecdotes instead.

He frustrates an assessment of his theory by obscuring what it actually predicts. He asserts that his "actual predictions involve the primary windows" only (p. 185) — however, should an earthquake occur in his secondary window, "it was predicted by my seismic window theory, but not directly by me" (p. 38) and he likewise refers to those as "hits" (e.g., p. 110). This ambiguity confuses readers and even Orey herself, who occasionally promotes earthquakes in his secondary windows to "predicted" status.

Through the kaleidoscope of seismic windows, local exceptions, multifarious indicators, and alternative explanations, he sees a triggering condition here or premonitory sign there that seemingly enables his theory to account for almost any significant earthquake. If his primary-window prediction fails, the secondary window provides a back-up, and if no "hits" there, then the lost pet ads may show an anomaly, and if not that, then he can point to equinoctial tides, or a "geyser gap", or record rainfall in the preceding months, or a solar flare, an "ear tone", or a phone call from a seismic sensitive with a migraine . . . thus, beguiling the unsuspecting reader about his theory's broad predictive power.

Throwing Darts

Orey's thesis, and the book's title, pivot on Berkland's track record of earthquake predictions: an astounding "75 percent accuracy rate" which he has claimed since the mid-1970s. How he computed this statistic before 1999 is left
Review of *The Man Who Predicts Earthquakes*

entirely to the reader’s imagination. Since then he has measured his success with a simple "dartboard" metaphor of his own invention—a convenience which, he explains, "permits evaluation beyond hit or miss and allows meaningful comparisons with predictions for various years and various predictors" (p. 187). Instead of a binary success or failure, his dartboard allows partial credit for near-misses. The predicted time, magnitude, and location define the "bulls-eye" (a "100%" hit). Outside of that are three "rings" that represent a miss by successive increments of one day, 0.1 magnitude unit, or a 10% increase in the radius of the circular prediction area. For rings 1, 2, and 3, he states, the probability of a hit "is somewhat greater, so the score is lowered to 90, 80 and 70 percent" (ibid.), respectively. The scoring method is simple. It's straightforward. And it's unfair, by arbitrarily and generously rewarding the predictor for random acts of seismicity.

For example, Berkland rewards himself with 70% partial credit for a miss by three days on either side of his eight-day window. However, crediting a miss of one, two, or three days in this manner actually adds two, four, and six days to the length of the eight-day window, and this raises the chance odds of a hit by 218 (25%), 418 (50%), and 618 (75%), respectively, above the odds of a bulls-eye. Therefore, a more objective discount for a miss by three days should be 100% minus 75%, or just 25% partial credit.

Similarly, missing the magnitude by 0.3 units earns him 70% partial credit. As with the eight-day window, allowing a miss by 0.1, 0.2, or 0.3 units on either side of his predicted magnitude range extends the range by double that margin—or 0.2, 0.4, and 0.6 units. Now, because small earthquakes are much more frequent than big ones (logarithmically, about 10 times more frequent for each whole unit decrease in magnitude, according to the "Gutenberg-Richter Law"), enlarging the predicted 3.5-to-5.5 magnitude range by those amounts increases the chance odds of a hit by roughly 26%, 59%, and 100%, respectively. Therefore, the partial credit for a missed magnitude prediction should be roughly 74%, 41%, and 0% for the three rings (a far cry from his 90%, 80%, and 70%!).

And it's the same for location. An earthquake falling outside of the circular predicted location, by 10%, 20%, or 30% of his standard 140-mile radius, earns Berkland a 90%, 80%, and 70% partial credit, whereas in fact, those increments of increased radius expand the circular land area by 21%, 44%, and 69%, thus, creating more spatial opportunity to capture hits by chance alone. So, the three-ring partial credit for missed location should be more like 79%, 56%, and 31%.

Orey illustrates how he scores the overall prediction (p. 187): "if Berkland predicts a 3.5-to-5.5 quake for a 140-mile radius and an eight-day window, he might get a hit of 3.4, 14 miles beyond the given circumference, and one day late. His total score would be 72.9 percent". In other words, her example is a "ring 1" hit for each parameter (magnitude, location, and time), and the summary score is the product of the separate parameter scores, or 0.9 × 0.9 × 0.9 = 0.729, or 72.9% for the overall prediction. If the summary score is less than 70%, "it's off the board", states Berkland (ibid.), and his prediction a "miss".
Clearly, Berkland’s dartboard rings increase the random probability of a hit more than "somewhat". His arbitrary partial scores inflate his prediction record substantially. Using the objective partial credits instead, the summary score for Orey’s "ring 1" example would be $0.74 \times 0.79 \times 0.75 = 0.44$, or 44%—a "miss", accepting Berkland's 70% pass-fail threshold. In fact, if more than one parameter misses the bulls-eye—or if just one misses beyond ring 1—then the prediction fails the 70% test.

(Any "objective" scores above are obviously over-simplified. Earthquakes are not random in space and time, but concentrate around fault zones and sometimes occur in temporal clusters of foreshocks, aftershocks, and swarms, for example. Moreover, reliable historical catalogs are short, typically just decades long, while large earthquakes in a given area are infrequent, often centuries apart—making estimates of their low rates and likelihoods uncertain. Nevertheless, the objective scores illustrate that Berkland’s "partial" credit is, in fact, partial to him. His dartboard method may be convenient, but that does not make it correct.)

"75 percent accuracy"?

Berkland computes his long-term success as the average of the summary scores for all of his predictions. Unfortunately, the book includes scores for only those predictions which he and Orey selected to showcase. Reasoning backward, however, it is apparent from his "90-80-70" percent dartboard system that a 75% long-term accuracy rate means his predictions, on average, missed the bulls-eye (a) on more than one parameter (magnitude, location, or time), or (b) by more than one ring. Depending upon which parameters he happened to miss, his "75 percent" average performance would fall to between 59% and 0% using the objective partial scores above.

Even ignoring objective scores, if his accuracy claim is to be believed, the reader must assume there has been no "file drawer effect" (i.e., that no "misses" are missing), but there is reason to wonder. Berkland has published his official predictions in no particular archive, so, a complete tally of his successes and failures, and the data supporting them, are known only to him. Orey’s book is not intended to be that archive; however, she does present a few short lists of his predictions and results, out of an obligation to establish the claim on the front cover. Where she includes them, her tabulations of data are often incomplete and error-prone. The result appears to be the selective reporting of his successes.

For example, to document Berkland’s accuracy in the U.S. Midwest, under the heading "4.0-Plus Quakes Since 1900" the author lists 22 earthquakes in the New Madrid seismic zone, each labeled "hit" or "miss" (p. 119). This, implies Orey, establishes his success rate for this region: "13/22 (59.0 percent) were in a predictable primary window (chance allows 26 percent)" (p. 120). No account is given of the composition, source, or integrity of the tabulated data beyond her exclamation, "take a look at these numbers".

On close inspection, errors and omissions abound. Deducing from the Midwest earthquakes that are listed, the list's coverage area is bounded by latitudes 34.0 to 36.5 degrees North and longitudes 89.78 to 94.2 degrees West. Searching the authoritative seismological databases for this region (CERI, 2006; NCEER, 2006), it is clear that Orey or Berkland, whether inadvertently or otherwise, omitted 11 earthquakes of magnitude 4.0 or larger that occurred within the same area—including the largest earthquake there since 1900 (a magnitude 4.9 event on March 25, 1976). Moreover, six events in the author's list (including one mistyped as “1076/02/15”) do not appear to be in the official catalogs. Of the 11 quakes she omits, all are misses (and only two are in a secondary seismic window). Of the six uncatalogued quakes she includes, all are labeled as primary-window hits. The revised record, based on the catalog search, shows that 7/27 (25.9%) of the Midwest earthquakes occurred in a predictable primary window, which is what chance predicts.

Confounding efforts to check his claims, Berkland calls for a magnitude that is generic, then he shops around for the premium brand. Measured seismographically, an earthquake's size is expressed in terms of several types of magnitude scales, all quite different: local ("Richter") magnitude (abbreviated ML), surface-wave magnitude (Ms), body-wave magnitude (Mb), moment magnitude (Mw), duration magnitude (Md), etc. Somewhat like apples and oranges, each scale represents a different physical point of view of the earthquake (e.g., ground motion amplitude of short-period pressure waves vs. long-period vibrations only vs. the total seismic energy released by the fault). For the same earthquake, the magnitude values assigned on these distinct scales may differ from each other by 0.5 units or more. Rather than calibrate his predictions with reference to a specific magnitude scale or to one authoritative earthquake catalog, Berkland canvasses the world's major seismographic centers after his predictions expire to find the maximum magnitude reported for each candidate hit and he accepts the highest number as the quake's official magnitude—regardless of scale (ML, mb, Mw, or whatever). (His egalitarian regard for magnitude scales is not explained in Orey's book, but was gleaned from his website.) Days and weeks later, as the preliminary magnitudes are revised upward or downward with more and better seismographic data, he accepts revisions that satisfy his predictions (p. 22), but there is no hint whether his "hits" become "misses" when revisions go the other way.

Of course, Berkland's scale-irrelevant magnitude search risks pushing the occasional large earthquake beyond the predicted magnitude range and off of his dartboard. However, it also admits a disproportionately vast number of "hits" onto the dartboard at the lower end of the predicted range because the smaller earthquakes are "orders of magnitude" more frequent, according to the Gutenberg-Richter distribution. (This likewise increases the number of misses, which err on the predicted time or location or both, but Berkland spends no effort searching them out for acknowledgment.) Hence, his magnitude validation approach also biases the results in his favor.
Ambiguity in the book's terminology further distracts the reader from Berkland's real track record. The terms "hit" and "miss" are nowhere strictly defined. An earthquake is variously called a "hit" for the time parameter when it occurs (a) within his eight-day primary window (p. 39 ff.), (b) within his eight-day primary window plus the six-day partial score period (p. 39), and (c) within his eight-day secondary window (pages 110-111). The lenient labeling thus allows "hits" on 22 days per month—assuring an apparent 70+ percent success rate even using real-world probabilities.

"Do-It-Yourself" Earthquake Prediction

With this book, Orey aims to acquaint her lay audience with Berkland—the man, his methods, and his mark on the world—in a popular narrative style. If that were her only aim, then she has succeeded. It is an easy read, and likely would be enjoyable and engaging to non-technical readers. (Amazon.com ranks the prose as easier to read than about 90% of science books, and only about 10% as complex.) She captures Berkland's affable and affecting personality, his breadth of curiosity, his sincere concern for public safety, and his courage and obvious enthusiasm for predicting earthquakes (qualities I found when I met him in person in the 1980s).

However, her unabashedly passionate treatment of him—"this dedicated geologist" (p. 18), "a progressive thinker" (p. 20), "this dedicated avant-garde scientist" (p. 27), "this brilliant man" (p. 32)—is poorly balanced against her view of the scientific establishment. Some of Berkland's critics get their say (e.g., Chapter 15), but she tends to cast them as a pack of axe-grinding, narrow-minded, and almost villainous prigs whom Berkland relentlessly brands "high science" (at least a dozen times throughout the book), rather than regard them merely as disciplined researchers seeking clarity through precise communication and acknowledged rules of logic. The heavy-handed character development, with its conspicuous icons of good and evil, are better suited for the fiction shelf than science writing.

Unfortunately, biography is not her only aim. The book is decidedly dogmatic in its claims. Berkland asserts speculations as if they are vetted scientific facts (e.g., "The magnetic field gets distorted locally because of upcoming earthquakes . . . dogs and other animals seem to pick up these rather abrupt changes in the magnetic field . . . for small quakes they get warnings of seven to ten days, but up to three weeks prior to large quakes", p. 33). And Orey manufactures "facts" of her own (e.g., "Cat's have an amazing ability to sense earthquakes", p. 36; "animals (both domesticated and wild) do have a sixth sense", p. 159). The mix of fact, fiction, error and hyperbole cloaked under the veil of "theory" leads the unwary reader to accept Berkland's claims as gospel.

Her title, her title, is "The man who predicts earthquakes". The book begins and ends on that premise. And in between—at the back of every chapter, in fact—is a sidebar with bulleted "Earthquake Facts" purporting to teach the reader
Berkland's "user-friendly do-it-yourself" quake prediction steps (e.g., "Each month note on your calendar the dates of new and full moons and eight-day seismic windows"), "Tally up the number of lost pets in the classified ads", "Next time your pooch howls at night, give Rover a doggie treat to reinforce psychic behavior", "A headache or craving for popcorn ... could be a warning that an earthquake is coming", "Eartones [may be] potential precursors for an oncoming quake", "Pay attention to visionary Clarisa Bernhardt's earthquake predictions").

This tutorial format reveals Orey's other aims. Through her long association with Berkland she has become a convert, a true believer, and recently a self-proclaimed "earthquake sensitive" (posted on www.calorey.com, accessed 19 September 2006). Shortly after publishing her book she co-founded her own earthquake prediction website (www.earthquakeepi-center.com) where she posts her own predictions and invites others to post theirs—and liberally promotes her new book and her newer, fee-based "do-it-yourself earthquake prediction" seminars conducted with Berkland himself. She is a Berkland disciple, a defender, a fanatic. I am left with the impression that this book is merely one of several bullets in a covert business plan, a slanted sales pitch on multiple levels, and should be read in that light. Unfortunately, her intended audience likely will not see it that way.

An Opportunity Lost

Berkland's unconventional approach to prediction is as eclectic as it is controversial. Orey attempts to cover the whole spectrum, including the data-laden debates. But, as "the girl who flunked math" (p. 180) she manages to do so only superficially and occasionally admits defeat (e.g., p. 167). As a formal document of Berkland's theory and methods of earthquake prediction, the book is frustratingly fragmentary. What is clear is that his "theory" is not theoretical, in the sense of being a unified set of unambiguous testable hypotheses. It is, rather, a tangle of vague speculations each of which, while plausible and perhaps appealing to the uncritical observer, resists diagnosis and definition.

For students of anomalous phenomena willing to extend Berkland the benefit of doubt in the face of his critics, and seeking once and for all to pin down his prediction hypotheses long enough to examine and perhaps test them, Orey's book is an opportunity lost. It is a compendium of unfounded claims and unchallenged data, wrapped in a dubious diary of untraceable anecdotes. Its many inconsistencies, generalizations, allusions, omissions, and errors, along with her wide-eyed embrace of personal testimonials and media stories, undermine both its credibility as a serious treatment of Berkland's method, and his credibility as a disciplined researcher.

For a scholarly treatment of unconventional earthquake prediction science, interested readers are referred instead to the likes of Rikitake (1976) and Ikeya (2004). The latter focuses on unusual animal behavior and was reviewed recently in the Journal of Scientific Exploration (JSE; Huang, 2005).
Changing Tide

Over the past three decades, a billion tax dollars (more or less) have been spent on USGS-sponsored earthquake research, much of it building and maintaining dedicated instrument networks along the nation's most active faults to monitor geophysical processes in which earthquake precursors were expected. These systems have captured many earthquakes including magnitude 6-plus events, some of them tightly within the world's most dense instrument arrays. Nevertheless, "We have never found an (earthquake) precursor", says Lucy Jones, USGS Scientist-in-Charge of seismology in southern California and Coordinator of USGS-funded earthquake research in that region (quoted in Kaufmann, 2006).

The initial confidence of the 1970s and 1980s led Western science to support a few official predictions (e.g., Bakun and Lindh, 1985; CISN, 2004), but all proved to be failures, and some very publicly so (Lubick, 2004a, 2004b).

Today, the official conclusion, on the USGS website, is: "Neither the USGS nor Caltech nor any other scientists have ever predicted a major earthquake. They do not know how, and they do not expect to know how any time in the foreseeable future" (USGS, 2006).

Outside of the U.S., scientists have not been quite so pessimistic. Numerous examples of pre-earthquake anomalies have been documented in retrospective case studies (e.g., Rikitake, 1976). But pessimists point out that for each case there are hundreds of similar earthquakes portended by no apparent anomalies.

The most widely acclaimed prediction remains that of the magnitude 7.3 earthquake in Haicheng, China, in 1975 (Molnar, et al., 1977). Reportedly based on foreshocks, anomalous fluctuations in the magnetic field and well-water levels, and unusual animal behavior, this prediction evolved over months and days and saved thousands of lives when vulnerable buildings were evacuated hours before the quake struck, destroying most of the city. It was minor consolation for the Chinese people, however, as the following year a similar earthquake in Tang Shan, near Beijing, was not predicted, and killed at least 250,000 people (as many as 750,000 according to some reports).

At that time, the sociopolitical environment in China limited Western access to the evidence and the discipline behind the Haicheng prediction. That vacuum, plus numerous "false alarms" and the terrible surprise at Tang Shan, raised doubts in the West about the Haicheng prediction claim. So, pessimists eventually chalked it up to coincidence—a well-timed earthquake swarm that coincided with "a sort of widespread public hysteria associated with a Cultural Revolution declaration that earthquake prediction could be accomplished through the unfailing efforts of the 'broad masses of the people'" (e.g., Scholz, 1997). Thirty years later, however, an analysis of declassified Chinese documents confirms the legitimacy of the Haicheng prediction as the only documented case of a successful precursor-based earthquake prediction (Wang, et al., 2006; Yauck, 2006).
The pessimists win. Berkland's claims face an impassable tide without all oars in the water. His unwillingness or inability to document and explain his method and reveal his record for peer review leave him stranded in the backwater of this science. Orey's book does nothing to advance his position.

Just now, ironically, seismologists are finding that tides really do trigger earthquakes (Cochran et al., 2004; Tanaka, et al., 2005)—but only the strongest ocean tides, and only on shallow faults of a certain type near coasts, and only when the tidal force acts in the same direction as the regional tectonic stress on those faults. As has been observed in other areas of anomalies research, small effect size thwarted this discovery for over a century. The concrete evidence remained elusive until large samples (tens of thousands of earthquakes) were carefully sorted and sifted by fault type, geometry, and location, and the tidal phase vector at the time of each earthquake was resolved on the theoretical fault plane. The tide effect is small under optimal conditions, and smaller still, if detectable at all, in inland areas (where most earthquakes in California, Oregon, and Washington happen).

Conclusion

Evaluating a scientific hypothesis should not be a scavenger hunt. It is in Berkland's case. On his website, curious and skeptical posters have tried to clarify the essential components of his theory, often just to understand them. Like a game of "20 Questions", they tease from his posted replies details of his methods that remain unclear, undefined, or undeclared. They consolidate the clues into coherent statements of hypothesis amenable to independent testing. And then they beg him for revisions. Hunter (2006), for example, based on as much as Berkland would post about his syzygy theory, tested seismic window predictions against random expectation. Chance did about as well as Berkland, according to Hunter.

How Berkland formulates predictions from lost pet ads in local newspapers has remained a mystery since he began this practice in 1979. To clarify that part of his theory, an anonymous poster (cited in Brown's Foreword, p. xiv) formalized the method through another online chat session. Before Berkland lost interest in the exchange, he had helped the veiled investigator revise the evolving "Missing Pet Ads Window Hypothesis" and even yielded his lost cat ad data compiled from a San Francisco area newspaper for the year 1989—the year punctuated by the disastrous Loma Prieta earthquake. The poster proceeded to test the hypothesis as far as it could be ascertained ("v2,1") against the lost cat data, with equivocal but tantalizing results. His/her online paper (Ped, 2005), in addition to providing a useful context for Orey's book, stands on its own as a worthy (and unique) contribution to the literature of earthquake prediction, and will likely be of interest to many JSE readers!

In the end, if Jim Berkland's theory remains misunderstood and unappreciated in the scientific world, he has only himself to blame. In that world, the path to
knowledge and understanding is by way of peer-reviewed publication, not through online interrogation, inaccessible newsletters, untraceable anecdotes, and adoring biographers. In the interest of science, he might take that path by publishing (on his website at least) a comprehensive record of his predictions dating back to 1974—both successes and failures—and packaging up his database of missing pet ads for public download, so that his past claims can be openly cross-examined. Going forward, for the record, he might consider issuing his online predictions by a method that is more tamperproof than a simple website, such as via a listserver or RSS (Real Simple Syndication), to minimize a reason for skepticism by his critics.

As JSE readers well know, anomalous phenomena are too often harshly judged by association with suspicious data, unsupported claims, and unreliable witnesses. The hypothesis of premonitory animal behavior has yet to be fairly judged, and "the Berkland effect" may be one reason why.

References


McClellan, P. (1986). Springtime seismicity in Central California before the 1906 San Francisco Earthquake—and its recent return. Paper presented at the Fifth Annual Meeting of the Society for Scientific Exploration of Anomalous Phenomena, San Francisco. (Note: Three years after this meeting the Bay Area was struck by the largest earthquake since 1906, the Loma Prieta earthquake disaster.)


BOOK REVIEWS


This two-volume set is the book version of two issues of the Journal of Consciousness Studies devoted to introspection; the title is a game of words, where "subject" refers both to the subjects used in research and to the subject of introspection. A detailed review of 24 different chapters and two introductions by the editors would be too long and tedious for most readers, so I will focus instead on some common themes across various papers. One strength of this anthology is its multidisciplinary approach, with works by eminent psychologists, philosophers, and neuroscientists. Besides its focus on introspection, Trusting the Subject can also serve as a good introduction to current research on consciousness, although predictably the quality and depth of the contributions vary.

The editors introduce the volume with a good overview of many of the debates on introspection, calling for multiple methods to supplement but not replace it (i.e., triangulation; Jack & Roepstorff, 2003). An important issue mentioned by almost every author is that introspection is a very complex concept, subsuming different processes and practices. The typical account that introspectionism was tried early in the history of psychology and failed as a method disregards a far more complicated situation in which different types of introspection were used, and it is by no means clear that introspection was an invalid or unreliable method, despite what psychology students are typically told (Goldman, 2004; Hurlburt & Heavey, 2004). Prinz's contribution (2004) elucidates how the term encompasses different types of awareness (e.g., non-verbal or verbal; primary or reflexive), and it might refer to many distinct mental processes, such as reports about perceptual experiences, subjective states, various forms of memory, and so on. Prinz cautions against assuming that these different types necessarily share "a common essence between them" (p. 55).

Another major topic is whether introspective reports should be obtained from "naïve" participants or from those who have had extensive training in introspecting and reporting on their mental occurrences. On the one hand, we educate people in all fields to provide "expert" observations instead of "untrained" ones; on the other, asking people to observe their mental events might affect them, a psychological analogue to Heisenberg's uncertainty principle, if you will. Schwitzgebel's paper (2004) shows how the attempts by Titchener to provide exhaustive introspective training in the early 20th century did not provide unambiguous results, although part of the blame could be placed on the structuralists' notion that conscious experience could be decomposed the same way as a chemical substance. In contrast, the ambitious and philosophically
sophisticated proposal of *neurophenomenology* (Lutz & Thompson, 2003) requires training in a phenomenological attitude towards attention and communication of mental occurrences, but framed within a model of research that includes behavioral and neural triangulation of data. Lutz and Thompson provide an exemplary study showing that trusting their participants' reports helped make sense of EEG data by differentiating their mental strategies. Gallagher (2003) makes the similar point that introspective reports can be used to design experiments. Other responses to such problems in introspection as reactivity, memory failures, and so on include the very important work of Ericsson (2003; Ericsson & Simon, 1980) on thinking out loud as people work on problems and Hurlburt and Heavy's (2004) use of experience sampling, where people are "beeped" at random times and asked to report the content of their consciousness just before the probe. This method has an ecological validity that is rarely seen in psychological research and has been used by Hurlburt to map the experience of "normal" and "schizophrenic" people (for its use on research with Vipassana meditators engaged in development of their metacognition, see Easterlin & Cardena, 1998–99). Schooler and Schreiber (2004) discuss how the validity and reliability of introspective reports depend on the specific phenomenon studied and can be established by looking at how they covary with behavioral and physiological indices. They also provide fascinating data on how sometimes introspecting can be harmful to memory consolidation (e.g., when asking a subject to describe complex objects such as faces) and how participants' minds wander in the midst of various tasks. Rather than just providing arguments, they show in what circumstances introspection is and is not helpful in the description of mental occurrences, a proposal that is similar to that of Goldman (2004), who argues for "wary acceptance" of introspective reports, unless there are good reasons (empirical or theoretical) to question them. Piccinini (2003) builds a case for testing the validity of introspective reports, in a similar way to the common sense evaluation of reports but formalized for the purpose of science (and what is the scientific method in general but the formalization and systematization of common sense evaluation of biases and competing explanations?).

An important limitation of introspective reports is that individuals do not necessarily have access to the cognitive processes underlying their beliefs, decisions, and so on. A very influential paper by Nisbett and Wilson (1977) purported that "The accuracy of subjective reports is so poor . . . (and) not sufficient to produce generally correct or reliable reports" (p. 233). This damning conclusion was questioned early on by Smith and Miller (1978), who indicated that the research question was not whether people have accurate access to their mental process, but under what conditions they have such access (see also Overgaard & Sorensen, 2004). Wilson (2003) now considerably softens his original conclusion and also discusses how many uncontroversial areas in psychology, such as the study of perception or attitude, are founded on introspective reports. He also makes a strong case for a dual system of human cognition involving both conscious and non-conscious processes.

Baars (2003), a long-time champion of consciousness studies, provides
fascinating examples of how conscious reports and cognitive and neurophysiological measures can provide clear answers to many issues, such as the nature of imagery, whereas Marcel (2003) elegantly discusses conscious and non-conscious processes, integrating both experimental and clinical phenomena, such as Anton's syndrome, in which blind individuals confabulate that they can still see. Cytowic (2003), an expert in synaesthesia, emphasizes his area and other clinical phenomena and concludes that sometimes a clinician must trust a patient's report, but not literally. Besides the works mentioned, a number of chapters focus conceptually and empirically on different explanations for autism (Hill, Sally, Frith, 2004; Robbins, 2004; Zahavi & Parnas, 2003); the experience of control and agency (Haggard & Johnson, 2003; Hohwy & Frith, 2004; Nahmias et al., 2004); and visual (Leopold, Maier, & Logothetis, 2003) or taste (Snyder, Fast, & Bartoshuk, 2004) perception.

Despite the many fascinating contributions in these two volumes, there are also some problems. Some contributors seem to share the same historic amnesia as the field of consciousness in general and assume that the "science of consciousness" is a very recent endeavor, thus disregarding decades of important theoretical and empirical work on, among other topics, introspection (e.g., Natsoulas, 1981), states of consciousness (e.g., Tart, 1975), and the reliability of trained introspective reports (e.g., Siegel & West, 1975). There is also almost no mention of introspective research on states of consciousness, despite considerable work in this area and the fact that without introspective reports the theme would be incoherent (Pekala & Cardeña, 2000). I also missed the perspective of someone like Allan Wallace (e.g., Wallace & Shapiro, 2006), an expert meditator who can comfortably present this perspective from both an experiential and an academic perspective.

In my opinion, too many pages were dedicated to a refutation of Dennett's heterophenomenology, his "phenomenology of another not oneself" (Dennett, 2003: 19), to which the dictum that "what is new is not good, and what is good is not new" may apply. What may be considered new is Dennett's notion that the contents of consciousness should be considered as just beliefs, which clashes against the actual data that our conscious experiences are not just beliefs (the "indisputable fact," according to Schooler and Schreiber, 2004). And what is good of Dennett's stance is that any phenomenological report could be misleading and thus may require additional forms of corroborative evidence. However, besides the fact that all forms of evidence, introspective or not, may be misleading and are subject to corroboration, this insight about introspective reports is not at all new and is "nothing other than good scientific practice," as Gallagher points out (2003: 90). Goldman (2004: 11) also shows how Dennett misrepresents actual research (in this case the work on mental rotation), and Schooler and Schreiber (2004) even quote Dennett warning others not to use specific quotes of his own work!

On the editorial side, these volumes show their provenance as journal issues. There is no index, the papers are not integrated by obvious topics (e.g., work on autism), and it is bizarre to have a summary of the contributions for the first volume at the beginning of the second one (Roepstorff & Jack, 2004). Also, some
contributions are regular research reports rather than comprehensive surveys of an area, and there are more typos (including authors' names in the table of contents!) than should be acceptable in a work of this caliber. I hope that in future reprints the editors will really make this more of a book rather than two related journal issues. Despite all of that, the volumes are chock-full of interesting and important discussions and data on current work on conscious and unconscious processes, and they affirm that one can trust introspection as much (or as little) as other sources of scientific knowledge, behaviorists and Dennett notwithstanding . . .

_Etzel Cardena_

_Thorsen Professor_  
_Department of Psychology_  
_University of Lund_  
_P.O. Box 213 SE-221 00_  
_Lund, Sweden_  
_Etzel.Cardena@psychology.lu.se_  
_http://www.psychology.lu.se/Personal/e_cardena/

**Notes**

I thank Devin Terhune, M.A., for his editorial assistance.

**References**


D. Scott Rogo was one of the few researchers who brought scientific parapsychology to the general public during the 1970’s and 1980’s. Although he was not employed by an academic or research institution, he made a career out of parapsychology, both as a researcher and a journalist. In his short life (1950–1990), he wrote over 26 books on parapsychology and related subjects, as well as publishing full papers in refereed parapsychology journals. The rapid publication of his popular books supported him financially, but writing at a pace of 20 pages a day sometimes led to errors, resulting in opposition, and sometimes outright hostilities, from his colleagues (Hansen, 1991). Although his popular books frequently contained far more scholarly references than comparable books of his time, these efforts were not always appreciated by other researchers.

Before embarking on this series of reviews of the D. Scott Rogo collection (Ventola, 2007), I searched the Journal of Parapsychology and the American Journal for Psychical Research in an effort to gauge how Rogo’s work was originally received by scholars in the United States. I found that aside from some scathing reviews (Akers, 1974; Pratt, 1975; Price, 1974; Stokes, 1977) and critical reviews (Alvarado, 1988; Anderson, 1987; Stanford, 1979), the majority of his books were unfortunately ignored. However, some scholars valued Rogo’s encyclopedic knowledge of the history of psychical research and his talent for communicating the complexities of parapsychological research to the general public. In a review of Minds and Motion: The Riddle of Psychokinesis, Gertrude Schmeidler (1979) emphasized the strengths of Rogo’s writing, stating that his work “forces on us a broad perspective that may make us shake ourselves out of our research grooves . . . and for the general reader, its presentation balanced with its critical approach, is likely to be not only stimulating but enthralling” (p. 156).

Being an independent researcher, Rogo’s position in the field was betwixt and between the public and academic spheres and this is reflected in his writing. This position allowed him the freedom to discuss topics that might be considered too fringe even for parapsychologists. Miracles, originally published in 1982, provides a parapsychological interpretation of wondrous phenomena recorded in religious traditions. Well-researched and thoughtfully written, it remains one of the few introductory books available on this topic. Rogo’s 1985 book, The Search for Yesterday, provides a critical examination of the evidence for reincarnation and freely discusses material outside of the current reincarnation research paradigm.
Our Psychic Potentials, written in 1984, looks at how the findings of experimental parapsychology can be used to develop one's own psi abilities. Embracing the roles of both experient and experimenter, Rogo communicates the results of psi research to the public in a meaningful way.

The purpose of Miracles is to provide an introduction to the study of religious miracles from a critical and parapsychological point of view. The book is divided into three broad sections, Miraculous Talents, Miraculous Events, and Miraculous Interventions. Within each of these sections, Rogo explores several representative genres, such as levitation, stigmata, divine images, bleeding statues, weeping Madonnas, manifestations of the Blessed Virgin Mary, and miraculous healings. Specific cases, such as the hailstones of Remiremont and the miracles at Garabandal, Spain, and Zeitoun, Egypt, are also discussed. Most of the material is taken from the records and testimonies used in the canonization of Catholic saints. However, throughout the book Rogo attempts to compare and contrast the miracles of Christianity to miracles recorded in other religious traditions and similar effects produced by individuals in the course of their secular lives.

Miraculous phenomena are untidy and elude simple classification. They are physically messy, often involving blood, sweat, and tears, unlike the more cerebral phenomena of parapsychology. They are conceptually messy as well, requiring one to tease apart that which appears supernatural from ostensibly paranormal phenomena. Rogo rather enthusiastically gets up to his elbows in it. In the end, he suggests that the power of belief might be strong enough to create an independent spiritual reality for the culture that supports it. This interpretation opens the door for a parapsychological inquiry into miracles without totally diminishing the purely religious significance of the events. Along the way, he states his case in language that is accessible to the general public, yet fully cited, using some unusual and unexpected sources.

The work of Ian Stevenson has become the mainstay of reincarnation research, but in The Search for Yesterday Rogo attempts to go beyond this research paradigm, introducing the public to a variety of approaches to the question of reincarnation, including hypnosis, regression, past-life therapy, and the past-life recall reported by researchers studying the properties of psychoactive drugs. The annotated bibliography also points the reader toward case reports originating in Western countries as well as the work of Dr. Stevenson, promoting a well-rounded examination of the evidence. The book has been more thoroughly reviewed elsewhere (Anderson, 1987), but it is worth noting that Rogo attempts to shake researchers out of their 'research grooves' by questioning scientists' very definition of reincarnation. Acknowledging the differences between Hindu, Buddhist, and Western conceptualizations of reincarnation, Rogo adopts a somewhat Buddhist theory about the nature of the self and suggests how the current evidence might be reconsidered.

There are a multitude of books and courses on the market purporting to enhance one's psychic abilities, but Our Psychic Potentials is not merely a 'how to' book on psychic development. It is equally devoted to the laboratory evidence for psi. The book is set up according to several overarching themes in parapsychological
research, looking at the roles of dreaming, mental imagery, relaxation, suggestion, and feedback. Each of these themes is discussed in a full chapter. The first half of each chapter summarizes the laboratory research, while the second half gives step-by-step strategies for how the reader could adapt this material into a program of self-training. In addition to detailing relaxation exercises and techniques for improving dream control and mental imagery, Rogo also instructs readers on how to simulate their own ganzfeld setting, or determine their statistical success over a series of Zener card trials.

Derived from laboratory procedures, Rogo's strategies could very well form the basis for a formal self-training course in ESP. However, Our Psychic Potentials is an ambitious work. It is not likely that casual readers would follow the entire program as laid out by Rogo, because he asks a lot of the solitary practitioner. However, I sometimes receive emails from visitors at the Public Parapsychology website seeking suggestions for how they can develop their own psychic abilities, and I have recommended this book to them. Unlike most other books on the subject, this book is devoid of the New Age platitudes typically found in such discussions and provides a very good introduction to the parapsychological literature, complete with references for further exploration.

Currently, the Anomalist Books' re-released Rogo series totals eight volumes. Having spent the better part of six months reading them, I find myself infected by Rogo's enthusiasm for the field of parapsychology and the promise of its findings. I admire how he communicated even the most arcane details of laboratory experiments in a cogent yet personal manner. In doing so, he took on the underestimated role of public scholar, a role that parapsychologists have largely neglected in the 21st century. Though one could easily become mired in the discussion of the types of inadequacies categorically exposed in some of the earlier reviews of his work, doing so might cause one to overlook Rogo's greatest contribution to the field: the wonder that he fostered in the American public.

Annalisa Ventola
www.publicparapsychology.org
Columbus, Ohio

References


In the past, analyses have made it seem unlikely that the Earth's rotation has affected global tectonics. Recent data suggest otherwise.

With this statement, albeit, only partly true, Robert Bostrom opens his chapter on Historical Perception, seeking the roots for thoughts about a connection between the Earth's rotation and global tectonics. Under the reign of plate tectonics, the potential link between the long-term inertial slowing of the Earth and its geological evolution has not been a matter of serious debate. Regardless of this ignorance, the strongly latitude-dependent lunar tidal friction is being assumed currently to be the dominant cause of both secular deceleration of the planet's spin rate and acceleration of the Moon's rate of recession. At present, the gravitational influence of the Sun-Moon system on the Earth, having its maximum effects at low latitudes, is relatively significant. However, there is hardly any geological evidence that deep ocean basins existed prior to the middle Mesozoic, some 150 million years ago, probably implying that oceanic tidal slowing of the Earth was not significant in earlier geological times.

The moment of inertia of a rotating body is a way of expressing the concentration of its mass about its centre of gravity. The greater the concentration of mass, the greater its moment of inertia, and the faster the body will spin. Thus, any net inward motion of mass increases the Earth's moment and its rate of rotation. Similarly, any net outward (equator-ward) mass transport decreases the moment of inertia as well as the planet's spin rate. An observation supporting the hypothesis of a relatively faster spin velocity in the geological past (compared to now) is that the present Earth has a certain excess flattening, making its equatorial bulge some 200 m larger than that expected. Munk and MacDonald (1960) suggested that this non-hydrostatic bulge is a relic of a faster rate of rotation in the past. Also, lengths of day estimates based on growth rings in fossil shells demonstrate the overall reduction in spin rate for the last 600 million years or so (as per data compiled by Creer [1975]). If the Earth has been undergoing long-term internal degassing, giving rise to continual upward transfer of mass since Precambrian times (see Storetvedt, 2003), probably even reducing the size of the core, it may have been in a state of decelerating rotation during most of its history. In his book, Robert Bostrom hardly touches upon these important aspects of Earth's rotation.

The gravitational tug from the Moon and Sun are predominantly recognized by the rise and fall of the sea, but the solid Earth too undergoes an up-and-down motion caused by external gravitational pull. The latter (bodily) tidal response,
producing much less surface distortion than the oceanic part, constitutes one-way
body waves progressing unimpeded around the Earth—in a direction opposite
planetary rotation. However, since the solid Earth is not perfectly elastic the bulge
does not relax immediately, causing a certain torsional stress on the outer brittle
layer. Hence, any long-term structural effect on the Earth's crust (tectonics) would
not seem an impossible proposition. So, at the outset, Robert Bostrom has a viable
project.

The tidal dissipation mechanisms have remained speculative to this day, not
least regarding those that possibly operate on the Earth's 'rocky' part. Despite the
fact that the bodily and oceanic dissipations seem impossible to separate, the
essence of Bostrom's book is to attempt to evaluate their possible tectonic
consequences. Regrettably, the author leans entirely on purely speculative plate
tectonic presumptions, including the highly problematic issue of solid-state mantle
convection. Thus, by building on models and concepts that have continuously
been strained by ad hoc repairs for more than 40 years, the author's endeavour has
minimal chances of success. In fact, the book's opening citation, "There is no end,
but addition" (T. S. Eliot), neatly summarizes the author's impossible task of
finding his way through the plate tectonics deadlock.

In his review of past attempts at linking Earth rotation and tectonics, Robert
Bostrom seems to be unfamiliar with important work by German and Austrian
geoscientists around the turn of the 19th Century. Thus, by integrating
paleoclimatology and geophysics, the old notion of True Polar Wander had
already been substantiated at that time. Without having actually read the book,
Bostrom, in passing, refers to the 2nd edition of A. Damian Kreichgauer's Die
Äquatorfrage in der Geologie (1926), cited by Reginald Daly. However, already in
the 1st edition of this book (Kreichgauer, 1902), a dynamic link between tectonics
and the Earth's rotation was outlined in surprising detail. Kreichgauer (an Austrian
geologist) also discovered the Pole-fleeing Force, i.e., the combined effect of the
dynamics of Earth rotation and the principle of isostasy (Archimedes' principle of
floatation). Kreichgauer found that the equator-ward force of crustal motion,
directed away from the poles, would have produced fold belts aligned along time-
equivalent equators, while a second set of tectonomagmatic belts would have
evolved in meridional settings. In other words, owing to the westward-directed
inertia forces caused by planetary rotation, the second set of tectonic zones was,
according to Kreichgauer, oriented at steep angles to the corresponding palaeoequator. By the way, speaking of Earth rotation and inertia-triggered
tectonic effects, how can Robert Bostrom write a whole book about the subject
without including the latitude-dependent Coriolis Effect? This is a serious
omission. At the end of his book he states: "Lacking the Moon, Earth would lack
also a mobile lithosphere". This is going a bit far, because the Coriolis Effect
would have provided the observed pattern of crustal mobility, albeit with possible
reduced tectonic vigour.

For Kreichgauer (and some of his contemporaries, including Alfred Wegener),
the geological evidence for ancient climates (from rocks and fossils) seemed to
favour the existence of True Polar Wander. For dynamical reasons, the rotational
axis has to be aligned along, or remain in the vicinity of, the maximum moment axis. Hence, the inevitable conclusion was therefore that internal axial shifts had intermittently taken place during the Earth's history. In other words, the apparent displacement of the poles over the surface was the result of the globe intermittently turning over relative to the ecliptic. In the Kreichgauer dynamic system, the required changes in the Earth's moment, causing polar wander, were brought about by the equator-ward and westward movements of the entire crust, without relative continental motion. The possibility that internal changes of mass could alter the maximum axis of moment, and thereby instigate True Polar Wander, was not a discussion topic at that time. Nevertheless, the associated changes in the relative position of the ancient palaeoequators, whatever their cause, would have given rise to tectonic belts in variable orientations across the globe. This was really a breakthrough in tectonics, but it was basically disregarded by Kreichgauer's contemporaries. However, Kreichgauer's understanding of the phenomenon of True Polar Wander was, a decade later, picked up by meteorologist Alfred Wegener, the father of continental drift—and the 'forefather' of plate tectonics.

Regarding possible driving mechanisms for tectonic processes, Wegener followed Kreichgauer and other German workers by adhering to the combined tidal torques from the Sun and Moon, the Pole-fleeing Force and the Coriolis Effect. Such forces are indeed directed westward and towards the equator, but the tectonic consequences of their latitude-dependent variation have not been given proper consideration in the past. Robert Bostrom, referring to the famous Stanley Keith Runcorn and his early work on continental palaeomagnetism, mentions the tendency of Southern Hemisphere continental blocks to have rotated counter-clockwise, as against corresponding clockwise rotations in the Northern Hemisphere. So what if the by now accepted continental mobility consists basically of relatively modest inertia-driven rotations in situ—not of relative lateral translations of crustal segments, as invoked by traditional continental drift and plate tectonics? In fact, it can be easily demonstrated that, for the larger land masses, only a few tens of degrees of relative 'in situ' rotations are required to explain the palaeomagnetically-based longitudinal discrepancies of the classical polar wander paths. Such a change in basic presumptions leads to a fundamentally different global tectonic system, in addition to avoiding the multitude of grave conceptual problems that flourished in the wake of plate tectonics. And equally important, mantle convection is no longer needed as a driving force in tectonics (Storetvedt, 1990, 1992, 1997, 2003, 2005).

By combining the changing palaeoclimate pattern with the observed global tectonic system, for different geological epochs Kreichgauer was able to draw polar wander curves that show remarkable similarities to modern polar paths based on palaeomagnetic data. In terms of geodynamics, Kreichgauer was indeed far ahead of his time. But his books were not translated into English, so the scientific message did not reach out beyond German-speaking Europe. The simple pattern of global polar wander for the past 400 million years or so, based on palaeoclimatic data, later described by Wegener (1929), is depicted in Figure 1a.
For comparison, the corresponding master curve for Global Polar Wander (Storetvedt, 1997), based on palaeomagnetic data, is shown in Figure 1b. For Robert Bostrom, and indeed for the majority of contemporary geophysicists, this simple picture of polar wandering seems to have gone unnoticed.

Considering the Northern Hemisphere palaeoclimate system, Wegener argued confidently that for the last few hundred million years the climate of the Arctic had gradually changed from tropical to polar, the most marked cooling having taken place only some 35 million years ago. A reverse trend, from polar to tropical, had taken place in southern Africa. Wegener ascribed this observational fact to the phenomenon of True Polar Wander—a spatial turning-over of the globe, in this case adding up to 70° of latitude. Indeed, if Wegener had avoided the distorted scientific vision that his preconceived idea of lateral continental drift undoubtedly inflicted upon him, he would have noticed that both the Arctic and Antarctic regions had experienced closely similar climatic trends over at least the last 400 million years. Unfortunately, Wegener ignored the growing evidence for ancient tropical to sub-tropical conditions in Antarctica that became available at the beginning of the 20th Century—which would have falsified his Pangaea concept, and indeed the very foundation of the later plate tectonics. Furthermore, with plate tectonics-constrained assumptions out of the way, many arguments in Bostrom’s book would automatically become irrelevant.

In his discussion of True Polar Wander and its bodily tidal actions, the author seems to regard mantle convection as a naked truth. However, despite the fact that both the nature and scale of the postulated cells of slow mantle flow have become increasingly obscure, the idea of thermally-driven 'solid-state' flows in the mantle has had many lives. With the twin-concepts of plate tectonics and mantle convection taken for granted, the challenge has been to understand, evaluate and harmonize the whole range of observations with the constraints necessitated by different types of evidence and the different proposed models. In fact, conflicting...
arguments have stimulated the formulation of a number of mutually incompatible 'systems' of thermal convection.

Compositional models of the mantle generally rest on assumptions of either one-layer (whole-mantle) or two-layer convection (upper and lower mantle separately), but ready solutions seem more remote than ever. In fact, if convection with plate tectonic time constants had been in operation, compositional heterogeneities could be expected to have been erased long ago. The pronounced inhomogeneity of the mantle, including the deep mantle roots of the continents, actually speaks against convection. It is indeed essential to ask whether thermal convection is at all feasible within a high-viscosity material like that comprising the mantle. What is more, the current choice of geotherm and other physical parameters for the deep Earth continues to be conditioned by the relevance of the convection hypothesis—for the sole reason that without it, plate tectonics would be without a driving mechanism. In other words, there is now every reason to look at the Earth with fresh eyes.

A range of studies suggest that the core is a good deal lighter than it traditionally has been considered to be and that it is not in equilibrium with the mantle. The presence of irregular core-mantle boundary topography (e.g., Morelli & Dziewonski, 1987) provides further evidence that this layer is a thermochemically active and heterogeneous Lone. What if the core is in a state of degassing (of lighter material) and the core-mantle boundary represents the fundamental trigger of energy, eventually leading to the observed range of geodynamic and surface geological phenomena?

Figure 2 shows that, when projected onto the Earth's surface, upstanding
topographic regions of the core-mantle boundary zone correspond to deep oceanic depressions. This may indicate that processes at the outer core and/or the core-mantle boundary layer release energy as well as buoyant masses that on the surface of the Earth lead to formation of oceanic depressions. This is additional information arguing against classical continental drift and seafloor spreading. The mobile crust, as demonstrated by palaeomagnetism and space geodesy, then reduces to a kind of 'cogwheel' arrangement—a system of continental blocks undergoing variable rotations 'in situ'.

The relationship between the 'morphology' of the core-mantle boundary layer and the gross surface topography indeed argues against classical continental drift (Wegener's mobile crust)—the theoretical forerunner of plate tectonics. Instead, the early Earth may have had a pan-global continental crust, having gradually—through vertical mass transfer from the deep, in association with mineralogical changes en route (see, for example, Storetvedt, 2003)—turned the cover layer into its present mosaic of continental and oceanic regions. The mass reorganization associated with such a degassing Earth inevitably would have altered the internal distribution of matter, causing changes in both spin velocity and intermittent events of True Polar Wander. These changes in planetary rotation are most likely the triggers of geodynamic phenomena—the motor behind the range of surface geological processes. And not least, the mass of plate tectonics-inflicted artificiality, in present-day interpretations, will disappear.

Robert Bostrom acknowledges the moral support of the legendary Keith Runcorn, who, "more than any of his contemporaries (...) perceived the central importance of establishing the existence or otherwise of convection in the Earth's mantle; he identified, furthermore, means of observing the resulting surface displacement: palomagnetism".

For decades Keith Runcorn was indeed very influential, and during the late 50s and 60s he turned his Newcastle Physics Department into a leading geosciences centre—the uncontested Mecca in palaeomagnetism and global geophysics at the time. Runcorn had an enquiring mind that never rested with 'finished' solutions. Thus, despite his crucial role in establishing widespread interest in the twin-concepts of continental drift and mantle convection, Keith became increasingly worried about the growing bewildered state in global geology and geophysics. Actually, at the end of his life he seems to have opted out of plate tectonics altogether. I would like to take this opportunity to give some of my personal recollections of Keith Runcorn's shift of attitude in global geophysics.

The close contact between the Physics Department in Newcastle and the Geophysical Institute in Bergen lasted for more than 30 years. But despite his many visits to Bergen, and our frequent encounters elsewhere, it took many years before I really had deeper scientific communication with Keith. He always wanted my opinion on matters that preoccupied his own mind, but when it came to my own scientific problems, and not least my doubts about the validity of some fundamental assumptions in global geophysics, he was for many years surprisingly silent. When I pressed him on these matters he mostly looked thoughtfully into the air, often smiling disarmingly and saying nothing, but
occasionally he would comment: “Yes, you may be right”! It was through a growing interest in history and philosophy of science on the part of both of us that I eventually developed a closer personal contact with Keith. This new situation came very suddenly, in early February 1980, following a colloquium on the evolution of the Central Atlantic that I gave in his department. At that time I had become very critical of many aspects of conventional continental drift and plate tectonics, so it was not surprising that my conclusions were at variance with the ruling dogma of the day. Something in my talk must have triggered Keith's curiosity, because from that time on I never had problems getting his attention. The following day he invited me to go with him to Durham, where he was going to give an evening lecture in the Department of Earth Sciences. In his car we discussed some of the grave problems facing plate tectonics, and I found him much more open about these matters than he had been before. Then he switched to the more philosophical aspects, asking me what kind of response I got from colleagues when I confronted them with my doubts about the established views. When I told him about all the non-scientific high-blood pressure/high-temperature reactions I had experienced by then, he turned to me smiling and said that that kind of behaviour was unfortunately very common but extremely unscientific. He then shared with me a number of amusing stories from his own repertoire, notably from the late 50s and early 60s, when he travelled widely in Europe and North America trying to convince the geosciences community about the reality of continental drift. Keith was extremely open-minded that evening, and we had a most enjoyable time together.

During the 1988–89 academic year I had a sabbatical leave at Keith's Newcastle Physics Department. During the daily coffee breaks the discussion topic was frequently the many intriguing aspects of global geophysics, not least the shortcomings of plate tectonics. One day, during a morning coffee in the Physics Penthouse, Keith said to members of the geophysics group present: "Well, we must take Karsten's criticisms of plate tectonics very seriously. Just remember all the observations we could never satisfactorily explain"! And he added, "I, for one, have often wondered, when I see the rotating globe just before the BBC News, why the Pacific Ocean is of roughly circular shape and covers about fifty per cent of the Earth's surface"! I was very impressed by such an honest and open-minded attitude of a man who for many years had been the *primus motor* in the process leading to the final acceptance of continental drift/plate tectonics, a theory which I now argued had to be denounced.

During the 1990 Copenhagen Assembly of the European Geophysical Society (EGS), Keith and I had longer discussions on the possible close link between Earth rotation and crustal deformation (tectonics). He was particularly interested in hearing my views about the link between the Coriolis Effect and the continental motion system—how relatively modest inertia-driven individual rotations of the larger continental blocks, around their 'centroids', could account for the dispersal of polar wander paths. The consequences of my new geophysical thinking would be devastating for plate tectonics, Keith admitted. During his subsequent visit to Bergen, in February 1992, he strongly encouraged me to continue my battle against
occasionally he would comment: "Yes, you may be right"! It was through a growing interest in history and philosophy of science on the part of both of us that I eventually developed a closer personal contact with Keith. This new situation came very suddenly, in early February 1980, following a colloquium on the evolution of the Central Atlantic that I gave in his department. At that time I had become very critical of many aspects of conventional continental drift and plate tectonics, so it was not surprising that my conclusions were at variance with the ruling dogma of the day. Something in my talk must have triggered Keith's curiosity, because from that time on I never had problems getting his attention. The following day he invited me to go with him to Durham, where he was going to give an evening lecture in the Department of Earth Sciences. In his car we discussed some of the grave problems facing plate tectonics, and I found him much more open about these matters than he had been before. Then he switched to the more philosophical aspects, asking me what kind of response I got from colleagues when I confronted them with my doubts about the established views. When I told him about all the non-scientific high-blood pressure-high-temperature reactions I had experienced by then, he turned to me smiling and said that that kind of behaviour was unfortunately very common but extremely unscientific. He then shared with me a number of amusing stories from his own repertoire, notably from the late 50s and early 60s, when he travelled widely in Europe and North America trying to convince the geosciences community about the reality of continental drift. Keith was extremely open-minded that evening, and we had a most enjoyable time together.

During the 1988–89 academic year I had a sabbatical leave at Keith's Newcastle Physics Department. During the daily coffee breaks the discussion topic was frequently the many intriguing aspects of global geophysics, not least the shortcomings of plate tectonics. One day, during a morning coffee in the Physics Penthouse, Keith said to members of the geophysics group present: "Well, we must take Karsten's criticisms of plate tectonics very seriously. Just remember all the observations we could never satisfactorily explain"! And he added, "I, for one, have often wondered, when I see the rotating globe just before the BBC News, why the Pacific Ocean is of roughly circular shape and covers about fifty per cent of the Earth's surface"! I was very impressed by such an honest and open-minded attitude of a man who for many years had been the primus motor in the process leading to the final acceptance of continental drift/plate tectonics, a theory which I now argued had to be denounced.

During the 1990 Copenhagen Assembly of the European Geophysical Society (EGS), Keith and I had longer discussions on the possible close link between Earth rotation and crustal deformation (tectonics). He was particularly interested in hearing my views about the link between the Coriolis Effect and the continental motion system—how relatively modest inertia-driven individual rotations of the larger continental blocks, around their 'centroids', could account for the dispersal of polar wander paths. The consequences of my new geophysical thinking would be devastating for plate tectonics, Keith admitted. During his subsequent visit to Bergen, in February 1992, he strongly encouraged me to continue my battle against
the dogmatic attitude of the geoscientific establishment. He maintained that the
gerologists had made a complete mess of global tectonics and that the Earth
sciences were in need of a brand new physical framework.

After his tragic death in December 1995, the Executive Director of EGS, Arne
Richter, called me to discuss the memorial program for Keith at the forthcoming
EGS Assembly at The Hague in April 1996. Arne Richter held that Runcom’s
growing disregard for plate tectonics was no longer a secret, not least in EGS
circles. Keith’s crucial role in breaking the ground for the modern drift-inspired
global tectonics and his later lukewarm enthusiasm for these seemingly popular
ideas were important facts for the history of science and therefore had to become
publicly known. Consequently, in a couple of talks at The Hague EGS Assembly I
unveiled some of Keith’s criticism of plate tectonics—an exposition that created
some stir in certain circles.

Bostrom’s book does not pretend to give final answers. The author touches on
many important and interesting aspects in global geology and geophysics, but with
plate tectonics as the rigid frame for the mass of attempted interpretations and
discussions, it is impossible to see a connecting phenomenological link in his
analysis. I think the author is fully aware of the non-conclusive nature of his book,
a situation for which the ingrained plate tectonics thinking is largely responsible.
Nevertheless, despite my overall critical remarks, the author makes a number of
interesting geophysical, historical and philosophical points. I therefore regard my
copy a useful reference text.

KARSTEN M. STORETVEDT
Emeritus Research Professor of Geophysics
Institute of Geophysics
University of Bergen
Bergen, Norway
karsten.storetvedt@&.uib.no

References

Creer, K. M. (1975). On a tentative correlation between changes in the geomagnetic polarity bias and
reversal frequency and the Earth’s rotation through Phanerozoic time. In The History of the
Earth’s Rotation (pp. 293–318). Wiley.


Munk W. H., & Macdonald J. F., The Rotation of the Earth: A Geophysical Discussion, New York:

Storetvedt, K. M. (1990). The Tethys Sea and the Alpine-Himalayan orogenic belt; Mega-elements in


177–187.


With his doctorate in Physics from Princeton University, Prof. Emeritus of Environmental Science at the University of Virginia S. Fred Singer has been the first Director of the U. S. National Weather Satellite Service and for five years, Vice Chairman of the U. S. National Advisory Committee on Oceans and Atmospheres, as well as author of over 400 technical papers, among many other activities during a very long career. Dennis T. Avery was a Senior Analyst at the U. S. Department of State, and has been a Senior Fellow of the Hudson Institute since 1989. He currently writes a weekly column on environmental issues, which has wide circulation, and he lives on a farm in Virginia. The inspiration for this book, Unstoppable Global Warming (USW), was derived from a 1998 article in Hudson’s American Outlook magazine entitled “Global Warming: Boon for Mankind?” (pp. 259–260). USW is partly built on Singer’s 1999 book Hot Talk Cold Science, which is still worth a read for its accuracy and many graphs.

Mainstream media have generally crumbled to the fearful predictions on global warming, which is said by some extremists to be on the verge of being out of control, as in the book The Revenge of Gaia, which I have reviewed in the JSE (Kauffman, in press). Prevalent current opinion is that human emissions of carbon dioxide (CO₂), produced other than by breathing, have increased the concentration of CO₂ in the Earth’s atmosphere by about 50% in the last 200 years, and there is no argument here. But it is also rampant dogma that this CO₂ absorbs infrared radiation emitted by the Earth and that this is the main cause of warming of the lower atmosphere and the Earth’s surface. To make the dogma scarier, there is a further claim that the years from about 1980 to the present have been the hottest in history and that more storms and more severe storms have resulted. Extremists have made it clear that humans must stop burning most fossil fuels in favor of wind and solar power. The true costs of doing so are never admitted—turning the standard of living and dying in the developed countries back some 150 years. And extremists turn the knife by asserting that most scientists agree with them—that there is a consensus on warming caused by CO₂.

International panic over these alarms resulted in the Kyoto Protocol for reduction of emissions of CO₂, implemented by a few countries in 2005. According to UGW, the only result will be payments of billions of dollars to the government of the Russian Republic, with no reductions of CO₂ emissions into the atmosphere (p. 226) by the countries making the payments. So UGW gives the evidence from the Russian Academy of Science in 2004 to reject Kyoto because (1) the world’s temperatures do not follow CO₂ levels; (2) the world was warmer in the Roman Empire period and during the Medieval Climate Optimum (1000–1250 AD) than it is now; (3) there is a much better correlation between warming and solar output than with CO₂ levels; (4) sea levels are not rising faster than 15 cm per century since 1850, even in the slightly warmer periods; (5) they did not expect tropical diseases to worsen with some warming; and (6) there is no increase
in incidence of storms or their intensity due to recent mild global warming, a notion with which the British delegation concurred. With no explanation, Russia ratified the Kyoto Treaty in 2005. The obvious explanations, if broadcast, that Europe and Japan would pay Russia billions of dollars to burn more Russian fossil fuels and that Russia would gain entrance to the World Trade Organization were not going to engender better worldwide public relations (p. 224).

To take the Russian Academy of Science's reasons in order, UGW showed that the most blatant attempt to fabricate the Earth's temperatures was that of Michael Mann at the University of Massachusetts, whose infamous graph of global temperatures from 1000–2000 AD became known as "the hockey stick". It showed a general sag in temperature from 1000–1920, leaving out the higher temperatures from 1000–1250 AD than we now have, then a huge leap. Investigation by a pair of Canadian researchers (see Essex & McKitrick, 2002) found that Mann's raw data, obtained by them with great difficulty, was flawed by every imaginable misuse and selection of data (details are given), including faulty use of tree-ring widths, an old standard for weather estimation. Mature trees not only grow faster when they are warmer or have more rain, but also when there is more CO₂, a fact which was ignored by Mann. Studies showed that mature trees at higher altitudes grew in proportion to the CO₂ concentration (pp. 67–70).

UGW cites evidence of many kinds to show that there were two periods hotter than now in the last 2000 years, a main strength of the book. Historical evidence, such as the Icelandic colonization of Greenland, then the failure of the colony; the extent and location of farmland in many locations, including in the Roman Empire and in China, isotopic ratios in fossils, length of the sunspot cycle, and other data was provided. One recent finding since 1979 is satellite data showing that the sun is producing 0.05% more radiation per decade (p. 197). The temperatures used by pseudoclimatologists to scare us show an inexorable rise for any chosen time period and are obviously tainted by urban heat island effects. New York City has warmed up, as have Pasadena and Tokyo; but the warming can hardly be global, since Death Valley, California; McGill, Nevada; and West Point, New York, have cooled, while many places in Europe have not changed much in 200–250 years (Crichton, 2004).

Sea levels around Sweden have been observed for 1000 years with no overall trend. Other evidence on sea levels is given in USW (pp. 47–50).

Malaria outbreaks have been observed as far north as the Arctic Circle. In the cool 1920s, Russia had 16 million infected with malaria, and there were 600,000 deaths (p. 204).

An entire chapter is devoted to debunking the baseless fears that global warming leads to more or more severe storms, and this chapter showed that there were more severe storms and droughts before 1900 than since then (pp. 161–173). The evidence claimed by UGW's title—the 1500-year-long cycle of temperature change—is built up carefully from page one.

The best feature of UGW is its depth of evidence, which is backed up by over 500 solid citations, most to peer-reviewed papers in scientific journals; there is no ranting, and all climate claims are backed by published examples.
Now about that so-called consensus on warming: UGW gives six examples of groups of scientists who were not in agreement with the "warming by CO₂" hypothesis in the 1990s. The largest group was over 17,100 mostly American holders of science or engineering degrees who signed a petition expressing doubt about man-made global warming and opposing the Kyoto Treaty. Of these, fully 2600 had climate science credentials. Perhaps the most important group was the 50 U. S. state climatologists, 90% of whom agreed with the statement: "scientific evidence indicates variations in global temperature are likely to be naturally occurring and cyclical over very long periods of time" (p. 66). Mass media: beware of your credibility—there's no consensus!

One of the strengths of UGW is its civil tone. Here is one of its least civil passages:

The strongest allies of the Greenhouse theory ['warming by CO₂' hypothesis] are:

- [Designers of] computer models that cannot explain past temperatures, let alone accurately forecast future ones, and ... [also enviro-extremists] whose funding depends on the public's fear of radical warming.
- Activists who oppose modern technology, abhor expanded human populations, and especially hate the low-cost energy that alleviates human poverty and misery ....
- European politicians.
- Journalists looking for scary headlines.
- Various national and international bureaucracies and UN-appointed members and staff of the Intergovernmental Panel on Climate Change.

It's not a very impressive lineup. (p. 198)

Another of UGW's strengths is recognizing that "Water vapor is the most important greenhouse gas even during the current [minor] warming [since about 1980]. Water vapor makes up about 60% of the natural greenhouse effect, with CO₂ making up an estimated 20% ...." (p. 40). Not knowing whom to believe in 1997, and to see who was correct about the relative infrared absorption of greenhouse gases, I determined an infrared spectrum of humid air at 40° north latitude. Fully 92% of the absorption was due to water vapor, and 8% to CO₂; no methane or chlorofluorocarbons were detected (Kauffman, 2004). Other strengths of the book are clear writing that allows fast reading, near-perfect editing, a good index and a remarkable glossary.

To show that your intrepid reviewer has not swallowed everything in UGW whole, here are some of the errors that turned up. (1) A change of 4°F does not equal 1.8°C, but rather 2.2°C (p. 112). (2) Britain does not feed its 60 million people from its own fields (p. 119) but still imports 40% of its food (Columbia, 2004). (3) The current Toyota Prius does not get 60 miles per gallon (mpg) city and 43 mpg highway (p. 216). Consumer Reports noted 44 mpg overall for the 2006 model, and Edmunds' survey of owners cited the same overall result (Edmunds, 2006). (4) Intensive farming (p. 121) may not be sustainable unless the fertilizers used are fortified with magnesium (Eades & Eades, 2000) and many
trace elements. Using animal wastes is really sustainable, even if it means using more land per unit of output (Pollan, 2006).

UGW may be the best overall 21st century book on climatology for the educated general public. You should consider having a copy as an antidote to the prevailing unscientific dogma on this subject.

Joel M. Kauffman
kauffman@hslc.org

References


Here I shall review the book about Max Born by Nancy Greenspan. In addition, I shall give some deeper comment about probability in quantum mechanics. I have thought off and on for over 30 years about the matter. So have many other scientists. To be overly brief and blunt, I am currently of the opinion that Born's great contribution to quantum physics, that the evolving wave function $\psi(t)$ of a Schrödinger's equation $\psi'(t) = H\psi(t)$ may be interpreted probabilistically, is more in the way of an artifice and not a reality. Schrödinger himself thought so. On the other hand, that $|\psi(t)|^2 = 1$ gives us the Born probability interpretation has been immensely useful, is irrevocably embedded into the quantum theory in countless books and treatises, and is accepted as equivalent to reality by many practicing and theoretical physicists... so I don't want to be irreverent about it. I just think that we should understand the physical causative mechanisms that may underlie it better. So I will mention some of my thoughts and recent results of my own toward that end.

Generally, quantum theory still contains so many riddles and so much counter-intuitiveness, that I want to state the following right here at the outset: I refuse to be drawn into lots of future arguments with any readers of this review. I don't think my own understandings of quantum mechanics are in any way final. Einstein...
complained bitterly that he worked 200 times as hard on quantum mechanics as on relativity but still did not understand it.

Let us turn first to Greenspan’s book. By chance I also had an old copy of The Born–Einstein Letters, from which Greenspan often quotes. She abbreviates that book as LTRS, and let us use that notation here. Let us also denote the book under review as MBG.

As Greenspan relates, she came to the idea of writing MBG from conversations with Born’s daughter Irene, whose later married name became Newton-John. So the famous British-Australian singer-actress Olivia Newton-John (e.g., Grease, with John Travolta) is actually Max Born’s granddaughter.

MBG carries a flavor as a book written by one rather close to the Born family. Indeed, there is much in the book also about Max Born’s wife Hedwig. Apparently Born had to exercise great patience in dealing with his wife ‘Hedi.’ Of course, Born was also under great pressure from the Nazis in Germany, and eventually he moved to Great Britain, first to Cambridge in 1933, then to Edinburgh in 1936. Born refused to return to Germany in 1945 to take Arnold Sommerfeld’s chair in Munich; also, he at that time refused an offer to return to Gottingen. However, Born retired to Germany in 1953, where he lived until his death in 1970.

Greenspan reports that on Born’s gravestone “is carved \( pq - qp = h/2ni \), … first written down by Born in July 1925.” Of course, this is nowadays usually called (one of the forms of) Heisenberg’s uncertainty principle. A good account as to how this principal evolved is given in pages 120–128 of MBG. Heisenberg, Born, Bohr, Kramers, Slater, Pauli, and Jordan were all talking to each other frequently. One could, in retrospect, call it a team effort, out of which the old quantum mechanics was transformed into a new one, in which one thought in terms of oscillators and discontinuous jumps in energy levels: particles, rather than waves. Heisenberg wrote a first draft with multiplicative rules for transition amplitudes and gave it to Born. Born recast it into matrix language. He found (1925) that the Heisenberg matrix products difference was \( pq - qp = h/2ni \). To quote MBG (p. 126), Born was proud that he was “the first person to write a physical law in terms of non-commuting symbols.”

However, he could not prove that the nondiagonal terms were zero. Pascual Jordan completed that part. The three of them, Born, Heisenberg, and Jordan, then wrote the paper together, which appeared in Zeitschrift für Physik in 1926.

Here I would like to interject three comments. First, many books imply that the uncertainty relation can be used to prove the stability of matter. That is not the case; see, e.g., the treatment in my book. Second, the operator-theoretic mathematics, which is much more general than the quantum physics matrix mechanics for transition rules, is nicely exposited in the little Putnam book. Third, although the operator-theoretic versions of the physical uncertainty principle are of course important (and there are three of those, usually called the Weyl, Schrödinger, or Heisenberg forms; see, e.g., our paper or the Putnam book), when many years ago I first encountered the uncertainty relation, I wanted a really simple proof of its mathematical essence. Here it is. Momentum \( p \) is the operator of differentiation: \( pu(x) = du(x)/dx \). Position \( q \) is the operator of
multiplication: $qu(x) = xu(x)$. All we need is differentiation by parts from calculus, usually written $d(\nu v) = udv + vdu$. Then

$$\frac{d}{dx}(xu[x]) = xu + u,$$

which is $pqu = qpu + u$, which is $pq - qp = \text{Identity operator}$.

Returning to our review of MBG, Born also is remembered for several other contributions to quantum physics. One of them is the Born-Oppenheimer approximation, useful in quantum scattering theory, which in essence determines the shape of a molecule. However, I want to focus on Born's probabilistic interpretation of the Schrodinger equation wave function. As is well known, the Schrodinger-de Broglie wave mechanics is essentially equivalent to the Heisenberg-Born matrix mechanics. The history of Schrodinger's derivation of his partial differential equations of atomic systems is interesting in itself and will not be discussed here. But by 1926 there were the two rival camps, Gottingen and Zurich. Pauli stepped in from the latter and with Schrodinger showed the (unitary) equivalence of these two formulations of quantum dynamics. Born's reaction is treated in pages 138–140 of MBG. He concluded that one could not know the state of an electron after colliding with an atom; one could know only "how probable is the specific outcome of the collision"; Born continued: "I myself am inclined to renounce determinism in the world of atoms." If $\psi(t)$ is an evolving wave function solution of the Schodinger partial differential equation, then the absolute value $|\psi(x,t)(x,t)|^2$ is the probability of finding the scattered particle at $x$. One then renormalizes these probabilities to sum to 1 by setting the $L^2$ integral over the whole configuration space to 1: $\|\psi(t)\|^2_2 = 1$.

I note that one sees in this ansatz (or is it, more correctly, a rationalization?) the influence of Bohr's original views earlier, that one must always take the statistical view of quantum mechanics, viz., as discussed in MBG (p. 122). Then Von Neumann entered the picture and wrote his foundations of quantum mechanics$^5$, which forever put quantum mechanics into a Hilbert Space, and probabilistically so.

So we are still stuck, 80 years later, with wave-particle duality, and as a consequence, Einstein's questioning, from MBG (p. 140), that "I, at any rate, am convinced that He does not play dice." Heisenberg, on the other side, by interpreting his uncertainty principle as meaning that more precision in measuring momentum meant less precision in measuring position, came down on the side of the statistical interpretation. This included Pauli's interpretation that the observer had become part of the problem, an issue that is still with us.

Let us now go to the Born-Einstein book LTRS for some corroboration, which I would like to add here for the convenience of the reader. In LTRS (pp. 83–88) one finds Born's letter to Einstein, dated July 15, 1925, and then Born's later commentary on it. The latter is a bit vague. Born states that "Later on ... the waves represent the spread of probability in the presence of particles. But this is not the place to pursue these matters in detail. Nor ..." On page 87 Born mentioned that he already knew that Heisenberg's manuscript's arguments could be
interpreted as the matrix equation $pq - qp = h/2\pi i$. On p. 91 of LTRS you will find Einstein's famous statement, which I quoted above, about God not playing dice. This is found in a letter to Born on December 4, 1926.

Another interesting clarification of Born's view of his statistical interpretation of quantum mechanics may be found on p. 186 of LTRS. Born tells Einstein that $\psi(t)$ does not really describe a state of a single system. Rather, "For what is really meant is, of course, that you take all individuals of 67 and count the percentage who live for a certain length of time [to say his life expectation is 4.3 more years, the example Born was using]. This has always been my own concept of how to interpret $|\psi|^2$. Instead you [Einstein] propose a system of a large number of identical individuals—a statistical total. It seems to me that the difference is not essential, but merely a matter of language."

I remark that what is going on here is, mathematically, an ergodic hypothesis, that time averages equal space averages. Most physicists today take an ensemble view of the statistical interpretation of quantum mechanics, although recently there are theories and experiments (welcher weg) about single trajectories.

Einstein replies (we are now in 1950, in Einstein's later years) and as usual sticks with his view that quantum mechanics through the wave function $\psi(t)$ is incomplete and is a theory that he hopes will [I skip back to quote directly from LTRS p. 173] "be replaced at some later date by a more complete and direct one." The Heisenberg-Born philosophy is that one must just live with it, and not delve further. Einstein wants deeper knowledge. Schrodinger (LTRS, p. 202) stubbornly believes there are no particles, no jumps, only waves. Born (LTRS, p. 229) takes full credit for the statistical interpretation of the Schrodinger equation wave function $\psi(t)$, which had just brought him the 1954 Nobel Prize. This was 22 years after Heisenberg had received the Nobel Prize principally for the uncertainty principle, which, as I noted above, was substantially a team effort.

One could go on with these retrospectral historical arguments forever. To move on, let me go back to the book under review, MBG, and to be specific, to Greenspan's statement and quote (MBF, p. 139, bottom) that "Born anticipated that some physicists would 'assume that there are other parameters, not given in the [statistical] theory, that determine the individual event'—that is, that something else lay deeper that would reestablish cause and effect." These other parameters have come to be called hidden variables. I remark that to my understanding, Born himself never advocated looking for such parameters. Although for many years after the turbulent quantum years of 1920s and 1930s the Heisenberg-Born-Bohr view that one could not go deeper held sway, in more recent years the belief in possible hidden variables has gained some support.6

As I stated at the beginning of this review, I now want to give some of my own thoughts and results on these matters. As I write this now, I realize that there is no way I can provide all the intermediate background for these thoughts and results: there would just be 50 years of all of the developments in quantum theory to get us nearer to the present. So with apologies, I will just jump to these thoughts and results. The reader may imagine this review to be a short account of a long conversation in which has been recorded only the very beginning and the very end!
Briefly, I want to address some very specific issues concerning (1) the statistical interpretation of quantum mechanics and (2) the question of incompleteness of quantum mechanics. I will call these two issues, respectively, (1) the Zeno Problem and (2) the Bell Problem. For efficient references the reader may consult, respectively, the recent Solvay Congress proceedings and the recent book.

The so-called Zeno’s paradox goes back to Von Neumann’s original formulation of quantum mechanics in Hilbert Space. It has become quite important in quantum computing’s need to control decoherence. Von Neumann saw it as a way to steer one state into another. It started as an attempt to model ‘continuous viewing’ of a quantum state evolution, $\psi(t) = U(t)\psi_0$. It is commonly described as the quantum version of “a watched pot never boils.”

I was actually (in 1974, with B. Misra, here in Boulder, Colorado) one of the first to consider the Zeno Problem. We modelled a continuous quantum observation by the operator limit

$$s = \lim_{n \to \infty} (Ee^{iH/n}E)^n$$

and found, for the interesting case physically (in my opinion), where the projection E does not commute with the Hamiltonian, that we had mathematical difficulties with this limit. Then Misra with Sudarshan published their famous paper in 1977. Probably I should have been coauthor. But I forgave my friend Misra, as he was between positions and the paper certainly helped him secure a position in Bruxelles. And his work with Sudarshan was completely independent of me. If you carefully read their paper you will see that the operator limits, such as the one above, which Misra and I were considering are simply assumed to exist. So the mathematical problems remain. Recently I have recorded my own thoughts on the Zeno Problem in a number of papers.

In particular, I found the following fundamental improvement in the Born probabilistic interpretation of quantum mechanics. Given a quantum evolution $\psi(t) = U(t)\psi_0$, which exponentiates a Schrodinger equation initial value problem in which $\psi_0$ is the initial wave packet which you have prepared, the mathematical theory requires that $\psi_0$ be in the domain $D(H)$ of the governing Hamiltonian if you want evolution equivalence to the Schrodinger partial differential equation. Then my theory says that $U(t)$ must take $D(H)$ one-to-one onto itself, at every second (hour, eon) of the Schrodinger quantum dynamics. The other states $\psi_0$ in the Hilbert Space which are not well enough prepared to properly evolve, i.e., which are not in $D(H)$, will also map one-to-one onto the complement of $D(H)$, but that is less important. The main point is that if we accept the Schrodinger equation and the Born interpretation, then the evolution $U(t)$ cannot lose even one probability distribution $\psi(t)$ from $D(H)$, not even for one instant. Should it lose even one possible prepared ‘probability’ state, uniformity is lost and irreversibility has occurred.

As to the Bell Problem, in recent years I have found intriguing connections between the Bell inequalities of the hidden variable theory and my independent
(since 1967) theory of noncommutative trigonometry. This has been reported in a number of papers. The details are technical, but I may summarize roughly by saying that all of the Bell inequalities may be placed within my more general noncommutative trigonometry. As the latter is essentially a new vector geometry in Hilbert Space, these developments mean that much of the mysticism and physical argumentation about nonlocality, realism, and the like, which surrounds the Bell Problem, may be more precisely seen as mathematical probabilistic-geometric issues. In particular, questions about what quantum probabilities and correlations mean and imply physically may be seen geometrically in terms of my new spin Bell equalities.

My new theory does not solve the hidden variable problem. But it does show, in my opinion, that the Bell inequalities also do not solve the hidden variable problem. The hidden variables problem lies deeper, and to understand it one must bring in, in a comprehensive way, all the Zeno quantum measurement issues, all the Bell geometrical issues, and, in particular, fundamental questions about the Von Neumann Projection Postulate in his quantum measurement theory. Of course other mathematical physicists have wondered about the Projection Postulate, or, if you will, the 'wave function collapse.' There are many, many papers that have been written about this, many of which I certainly have not seen. And this is not the place to try to give a representative view of wave function collapse and the various attempts at explaining it. Perhaps I will eventually finish my study of it, underway some years now. However, let me for the reader's advantage mention here a few treatments.

The physicist Ballentine in his books and papers has addressed the Projection Postulate, e.g., see the cited paper and citations therein. I should mention that prior to Gustafson and Misra, Misra and I tried to develop a model which (quoting from Ballentine) "discards branches of the state vector that are not relevant to the task at hand," a view of collapse that Ballentine advances in his conclusion. But our mathematics showed something quite different. Our so-called decaying states without regeneration did not exist for unstable quantum systems. Instead, they represented regular stationary stochastic processes.

Then there are those who picture quantum probabilities as Brownian motion 'corrections' to the usual Schrodinger dynamics. See, for example, the recent book by Adler and the citations therein. A similar approach is advocated by Percival. One then tries to argue for the Born quantum probability interpretation as an emergent phenomenon. While the mathematics is nice, i.e., becomes a challenging chance to bring the Ito stochastic calculus into the physical model, I cannot buy it as foundational physics. To quote Adler (p. 169): "when ... the operator coefficients of the noise terms are mutually commutative and commute with the state vector reduces on the eigenstate basis that diagonalizes the operator coefficients of the noise terms, with reduction probabilities given by the Born rule ...." I have, I suppose, always had a sort of contrarian reaction whenever I see operators blankety assumed to commute. I have elsewhere set out my objections to what I perceive as hidden assumptions of detailed balance.

To avoid getting deeper into these issues here, let me now just refer the reader
to the book by Penrose\textsuperscript{23}, which in my opinion nicely discusses many aspects of the quantum measurement problem and, moreover, keeps those discussions grounded within the larger context of other key matters in modern physics. In particular, one finds a nice classification of the measurement paradox explanations (a) 'Copenhagen,' (b) many worlds, (c) environmental decoherence, (d) consistent histories, (e) pilot wave, and (f) new theory with objective R. As I am, and always have been, Penrose is also uncomfortable with the density matrix (Heisenberg picture) ansatz.

I need to close this review, so let me rather abruptly say, and with paraphrase, that philosophically, my thoughts on these matters would lead me to currently vote for Einstein's view, that we should seek a deeper understanding, physically, e.g., of the real nature of an electron and a photon in a field, even their own, rather than to settle for the Born-Heisenberg view that to seek such is hopeless.

\textit{Karl Gustafson}

\textit{Department of Mathematics}

\textit{University of Colorado, Boulder}

\textit{Boulder, Colorado 80309-0395}

\textbf{References}


That HIV does not cause AIDS could hardly be stated more cogently, compactly, readably, and even comprehensively than in this book. There are also insights here not sufficiently emphasized in other such contrarian or "dissident" works: the sheer lack of logic in much of what the mainstream says about HIV/AIDS; and that this illustrates a pervasive decline in intellectual standards in science as a whole.

Culshaw is assistant professor of mathematics at the University of Texas at Tyler. Her graduate work, on modeling immunological aspects of HIV theory, led her to recognize deficiencies in HIV/AIDS theory. The book's emphasis on logic may well owe something to the discipline of mathematics: in much of empirical science, paradoxes and anomalies are tolerated so long as something "works"; but being illogical in mathematics is fatal. Perhaps, too, the earlier role of graduate student allowed Culshaw to take a fresh look not easily taken by people immersed in the subject for a couple of decades; though one may hope that HIV/AIDS theory will not (continue to?) exemplify Planck's dictum, 'that the old fogies need to pass away before a mistaken mainstream consensus is corrected.

The Introduction begins with a conundrum pointed out by Duesberg. Official estimates have had about 1 million Americans HIV-positive, steadily, year after year, since 1985. That is incompatible with the chronology of AIDS cases and AIDS deaths: official figures show for AIDS cases a linear increase, and for deaths a rapid increase, peaking in the early 1990s and then decreasing again. Those HIV data are also incompatible, quantitatively as well as qualitatively, with the official estimate of about 10 years from infection to AIDS. Further, the symptoms and epidemiology of AIDS are entirely different in the First and Third Worlds, which is not the case with any other disease, be it sexually transmitted or not. Moreover, AIDS has become so mired in emotion, hysteria, and politics that it is no longer primarily a health issue. . . . [P]ronouncements by powerful government officials and ill-informed celebrities are taken as gospel, and no one even remembers when, a few years later, these pronouncements turn out to be false. (p. 4)

Chapter 1 recounts briefly how Culshaw came to realize that the HIV/AIDS model is wrong. Chapter 2 points out that there is no agreement on, or evidence for, any mechanism by which HIV lulls the cells that are depleted in AIDS.
Culshaw comments, too, on the short life of HIV/AIDS publications, as contradictory claims are published only to be themselves quickly superseded. Here more discussion and specific examples would have been warranted. Culshaw's jaundiced view of the short life of research articles may be a corollary of being a mathematician; in chemistry and physics and other experimental and primarily inductive sciences, most research articles are ephemeral, never cited after a few years (indeed, most are never cited at all). It would make the case here more solidly to illustrate the manner in which later work continually contradicts earlier work in HIV/AIDS; for example, recommendations to treat early rather than late, and to interrupt treatment rather than to avoid any interruptions, have alternated several times with their very opposites, as small unreplicated research studies are published and treated as believable until the next one appears; or, the increasingly far-fetched and divergent notions of how HIV causes CD4+ cells to die off, ignoring articles published in 1992 which found that it is not a virus but an antibiotic-sensitive agent that can kill these cells in culture. In the experimental sciences, rather trivial and ephemeral little items about various details eventually cumulate to a consensus and an understanding; but in HIV/AIDS, there has been no such cumulation, just a succession of anomalies, conundrums, and contradictions. Culshaw notes the overall lack of logic in mainstream pronouncements in this regard:

HIV researchers continually claim that certain papers' results are out of date, yet have absolutely no hesitation in citing the entire body of scientific research on HIV as massive overwhelming evidence in favor of HIV. (p. 12)

Again, Peter Duesberg . . . is often cited as having been discredited despite the fact that there is no record of this "discrediting" anywhere in the scientific literature. (p. 13)

Chapter 2 also mentions a number of competent experts who have pointed to various flaws in HIV/AIDS doctrine and makes an important point that I have not found elsewhere among HIV/AIDS critiques:

HIV science has sold out to the epidemic of low standards that is infecting all of academic scientific research.

... [M]any of the arguments presented in response to the queries of HIV/AIDS skeptics are essentially some form of appeal to the use of low standards, . . . "You don't need a reference that HIV causes AIDS," "The fact that HIV and AIDS are so well correlated indicates that it must be the cause," "HIV is a new virus, and new viruses will meet new standards," "Koch's postulates are outdated and don't apply in this day and age," "We don't need to worry about the actual infectious virus, viral markers should suffice," . . .

... It is this decline in scientific standards that I point to when I am asked how so many scientists and doctors could be so wrong. (pp. 13–14)

Chapter 3 recounts how HIV/AIDS dogma became established, against the evidence even then available. For instance, the mathematical analysis on which HAART (Highly Active AntiRetroviral Treatment) depends was almost imme-
立刻显示是无效的，而这种“鸡尾酒”疗法在常规、标准使用中仍然继续。

第4章评论了在定义"AIDS,"获得性免疫缺陷综合征时的一些不合逻辑之处。通常，当一个新临床综合征—一套症状—被承认时，进一步的工作会细化它的描述，把症状筛选到那些真正特征性且特别源自致病因子的，从而导致对那个因子的识别。与之相反，"AIDS"的定义已经被广泛地扩展，包括越来越多的个体疾病，所有这些疾病在HIV或AIDS出现之前就已为人所知。今天，"AIDS"是完全不同于1980年代早期的一套疾病和症状。再次，血友病患者不应被包括在那些据说缺乏明显原因的免疫缺陷中，因为免疫系统在血友病患者中不能正常工作。此外，CD4+细胞计数作为免疫缺陷的指标在健康人中的差异如此之大，以至于这种"替代标志"的使用是明显的无效——正如许多研究表明的；但，正如HAART，那些直接证据并没有停止这种做法。

许多研究表明，抗逆转录病毒治疗弊大于利，没有结束因治疗而引起的疾病和死亡的不正常现象。正如通常情况，缺乏逻辑：当CD4+细胞计数上升（据称，免疫系统“改善”），患者快速变得更重病；但是，停止这些治疗不是现象的名称，"免疫恢复综合征,"一个假设的，以前未知的，完全不合逻辑的，免疫系统在被治愈后会不断引出它的宿主的假说的性质。在早期的AIDS患者中，患者死于罕见、机会性感染；今天，最常导致AIDS死亡的原因是肝衰竭和心血管事件，这是由药物治疗直接引起的，在AIDS患者在获得抗逆转录病毒治疗之前没有死亡。但是，这种变化被官方统计数据掩盖了："在马萨诸塞州……所有HIV阳性的人的死亡被计为AIDS死亡……（甚至如果死亡结果是）肝衰竭，心脏病发作，自杀，溺水，CMV（巨细胞病毒）感染，或车祸，或其他任何东西”（p. 30）。

所有这一切，使令人震惊的是，疾病控制和预防中心现在建议所有的孕妇都进行HIV检测，因为怀孕是HIV检测的常见原因（p. 31）。

第5章调查HIV检测的错误：一方面，它们从来没有被验证为纯HIV——这可能看起来很惊讶，但事实是纯HIV从来没有被分离出来。所有制造商的测试包声明免责声明：测试不能也不应该用来诊断HIV感染。然而，它们恰恰是这样使用的，尽管是"在标准性、专一性和可重复性方面最差的最糟糕的测试"（p. 36）。这是一个可怕的缺失，媒体没有特别关注到这一点。
since testing HIV-positive has led to incarceration of people found guilty of transmitting HIV and to treatment of many overtly healthy people with toxic drugs. The tests were devised "in an amazing display of circular logic" (p. 38). They involve diluting 400-fold the blood to be tested, because undiluted blood from almost every healthy person tests HIV-positive on these highly non-specific tests. Furthermore, the criteria for "positive" are not clear-cut. They depend on judging the shade of a color (in ELISA tests) or the strength of coloring of a variously defined number of protein-containing bands (on the Western Blot test, which is used to "confirm" a positive initial test on the ELISA). Supposedly there are 10 proteins characteristic of HIV. Depending on the country and the medical authority, detecting anywhere between two and four of these is taken as a "positive" test! But "since none of these proteins is specific to HIV, this would be like saying that since dogs have four legs, are furry, wag their tails, and enjoy eating steak, that any entity that is furry and enjoys steak must be a dog" (p. 41). In logic, if these proteins really were specific to HIV, then finding just one would suffice to demonstrate infection.

In actuality, the "HIV" tests are highly non-specific. People whose health is challenged for any reason at all are more likely to test HIV-positive (as shown by the mass of accumulated HIV data).2 Dozens of medical conditions cross-react with these tests to produce false positives, while on the other hand many AIDS patients never test positive (p. 40). More ill logic is displayed in the mainstream satisfaction because the tests are supposedly 99% specific. Even were that the case, since the numbers of HIV-positives in the United States are well under 1% of the whole population, testing everyone would turn up false positives in much larger numbers than "true" HIV-positives. Ninety-nine percent specific means that 1% of HIV-negative people will test positive as one seeks to identify the less than 1% of the population that really harbors HIV. Thus, testing all Americans would yield more false positives than genuine ones, and that would lead to administering toxic drugs to about 170,000 people who falsely tested positive. As it is, pregnant women and babies are already urged, in some places and cases actually forced, to take these tests and then these drugs.

Chapter 6 is about the lack of evidence that HIV causes AIDS, as admitted by mainstream researchers, for example, in a 2006 article in Nature Medicine: "The pathogenic and physiologic processes leading to AIDS remain a conundrum" (p. 51).

Chapter 7 considers "The sociological implications of AIDS." Culshaw notes the fear that HIV/AIDS orthodoxy has instilled throughout the world and the ad hominem polemics by which the orthodoxy defends itself, owing in some part to the financial interests of drug companies and the vested interests of hordes of researchers. Mainstream beliefs seem to rest to a certain extent on endemic racism and homophobia, given the unsupported speculations in the professional HIV/AIDS literature about the alleged behavior of Africans. Chapter 7 concludes with a brief summary of the persecution of Christine Maggiore, whose child died of
unknown causes, probably an allergic reaction to amoxicillin, but allegedly from AIDS.

The emphasis on social aspects continues in Chapter 8, "Where do we go from here?" (p. 70):

... when the HIV/AIDS hypothesis is finally recognized as wrong, the entire institution of science will lose the public's trust, and science itself will experience fundamental, profound, and long-lasting changes. The 'scientific community' has risked its credibility by standing by the HIV theory for so long. This is why doubting the HIV hypothesis is now tantamount to doubting science itself, and this is why dissidents face excommunication.

And so, given the dereliction of responsibility within the scientific and medical and mass-media communities, it becomes necessary for individual people everywhere to look at the evidence for themselves and reach their own judgments.

On all essential points, this book is right on the mark. I wish for it many future printings and editions, and it is in that spirit that I note a few easily correctable blemishes. The book deserves an index. The Glossary is a good idea, but some entries are superfluous while some other terms should be included. On p. 86, "hysteresis" should be "hormesis." On p. 52, gp41 should be described as a monomer, not an oligomer, of gp160 (a distinction made correctly elsewhere in the book).

I do not usually take much note of a cover design, but the one on this book is highly appropriate to the book's content, and it is striking without being fussy.

If anyone is willing to read just one "dissident" book about HIV/AIDS, this would be the ideal choice among some very worthy contenders and a number of quite useful ones that deal with more limited aspects of the matter; though I hope reading Culshaw's book might also lead readers then to delve into some of the others, and into my own very recent effort.

HENRY H. BAUER
Professor Emeritus of Chemistry & Science Studies
Dean Emeritus of Arts & Sciences
Virginia Polytechnic Institute & State University
hhbauer@vt.edu
www.henryhbauer.homestead.com

Notes

1 Planck's own words have been translated from the German original, paraphrased, and abbreviated in various ways; for example: "New ideas win out not by convincing their opponents but by outliving them"; or, more provocatively, "Science progresses funeral by funeral."

All profits from the book are designated for the Serge Lang Memorial HIV/AIDS Archive. Lang, who died in 2006, was a distinguished mathematician and an outspoken, uncompromising, irreplaceable defender of academic freedom and intellectual standards.


Richard DeGrandpre might be familiar to you as the author of Ritalin Nation. Ritalin comes in for much attention by detailed comparison with cocaine. Both are said to produce the same mental effects, to the point where Ritalin is called "synthetic cocaine". A main theme of this book is that Ritalin is considered an "ethical" drug and an angel in dealing with ADHD, while cocaine is considered a "street" drug and a demon; this artificial difference had nothing to do with the pharmacology of the two drugs, but to the conditions of use and the street or medical dogma on each, called "placebo text". Both are dopamine reuptake inhibitors in the brain.

Methamphetamine as "meth" or "speed" has been called by a federal "drug czar" "the worst drug ever to hit America", and The New York Times wrote that it was "feeding an epidemic of addiction that . . . rivals that of heroin and cocaine over the past few decades". The same drug has been available as Methedrine or Desoxyn for decades and is said to have "all the qualities you could possibly want in an ADHD med—it doesn't cause anxiety, it barely raises heart rate or blood pressure, it totally wipes out depression and fatigue, and it lasts a full twelve hours . . ." (p. 32). Angel or demon?

DeGrandpre also notes that "demon" heroin, introduced by Bayer of Leverkusen around 1885 as a non-addictive form of morphine, was available without a prescription for about 25 years, and is not nearly as addictive as U.S. government officials have propagandized. Doubters should be warned that loose claims are not found in this book, and fully 52 pages are devoted to citations, mostly to medical journals, appendices, and the index. Provision of maintenance doses of heroin and other street drugs in the UK and The Netherlands, paid for by their national health services, was called a far better solution to a violent underground drug supply economy, which is the result of the prohibition in the USA.

Quite a good history of "mind-altering" drugs from Big Pharma is given, including amphetamines, tranquilizers, etc. Prozac from Eli Lilly came in for much attention. Not the first nor the last SSRI, Prozac was at first considered as an anti hypertensive drug. "After the drug succeeded in not killing laboratory animals in initial exploratory studies—although it turned cats from friendly to growling and hissing . . .", Lilly nevertheless responded to competition by launching Prozac as an antidepressant (p. 53).

All subjects were given deep examination from many angles. Prozac is the one example I will treat in depth. DeGrandpre left little doubt that prescription Prozac occasionally led to self-mutilation, suicide and murder (p. 62) by giving 4 case studies as examples. Despite the knowledge that Prozac could be effective in daily doses of just 5 mg (Cohen, 2001), Lilly's recommended starting dose was 20 mg. In each of the 4 cases, more anxiety in the subject caused the physicians to increase the dose to 40 or 60 mg/d. All of these subjects then had even more
anxiety and injured themselves; but so strong was the package (or PDR) message from Lilly that the dose was increased again, with worse results. DeGrandpre then wrote:

Growing reports that Prozac might be unsafe at any dosage [sic] had Lilly running scared. One executive stated in an internal memo in 1990 that if Prozac were taken off the market, the company could "go down the tubes". Responding to concerns expressed by the FDA, Lilly agreed to conduct a study examining the question of whether Prozac induced aggression and suicidal thoughts. The result, known as The Beasley Study, appeared in the 21 September 1991 issue of the *British Medical Journal* [following sensible rejection by the *New England Journal of Medicine*].

The study, which was authored by Lilly employees and included psychiatrist Charles Beasley, looked and sounded like good science. On the surface, it represented the data pooled to date comparing Prozac with either older non-SSRI antidepressants or placebo. In fact, the data had been handpicked to favor the company. The analysis dealt with only 3,065 patients, less than 12 percent of the total data available from clinical studies at the time. Among those left out was the very population most likely to become suicidal—the roughly 5 percent of patients who dropped out of the clinical trials because they experienced unpleasant side effects after taking Prozac. The report also made no mention of the dozen or so suicides that had already occurred in Prozac's clinical trials, a number that, given the population being studied—primary-care outpatients rather than severe depressives—would be expected to be near zero.

Not until the Fentress and Forsyth [court] trials did Lilly's internal documents surface, revealing the depth of the deception [from 1978] . . . .

In contrast to the claims of Lilly executives, [a 1995 study in the *BMJ*] suggests a sixfold increase in suicide risk for Prozac [users] relative to [users of] non-SSRI antidepressants, a number that is close to Lilly's own internal assessment from 1985 . . . . (pp. 41–45) [For confirmation, see Healy, 2003.1

Prozac was used as an example of overpromotion of a drug and its drug class that lasted as long as the patents, then a "newer, better" drug under a new patent would be promoted. Of course, many other recent books with this theme exist (Abramson, 2004; Cohen, 2001; Kauffman, 2006); but The Cult is not primarily a jeremiad against Big Pharma but rather a window into how much the pharmacology of a given drug, including nicotine or alcohol, is combined with the overpromotion, myths or prohibitions of a drug to confuse its supposed benefits and risks.

The lack of effect of nicotine levels on the addiction to cigarette smoking and the failure of alternate nicotine supply treatments to curb addiction more than slightly was quite a shocker. So was the report of a finding in a trial that secretly substituted low-nicotine cigarettes for normal-nicotine cigarettes for a group of existing smokers, many of whom never noticed the difference. The special status of tobacco and alcohol, because they were common farm products in the USA, was brought out. Prohibition of alcohol was a failure partly because alcohol was and is an excellent tranquilizer when not overdosed and is only addictive in a small minority of users.

Gradually the war on street drugs is shown to be similar to the current war on supplements in that Big Pharma wants its most expensive stuff used and has gone
to great lengths with both overt attacks, indirect attacks by entities not identifiable as Big Pharma, and control of government and non-government agencies (Abramson, 2004; Cohen, 2001; Kauffman, 2006). DeGrandpre wrote:

The cult of pharmacology must therefore have served a different purpose than the elimination of dangerous drugs and the sanctioning of psychiatric medications ... during the twentieth century. The competitiveness of the drug market and the fact that one or two successfully approved and marketed compounds could raise a company from rags to riches almost overnight made for an increasingly aggressive and reckless industry. The medicopharmaceutical industrial complex that ... emerged benefited directly from differential prohibition, moreover, in that the demonization of certain natural substances—marijuana, cocaine and opiates—helped set them apart from the "ethical" pharmaceutical compounds, even if the latter had equal or greater toxicity. (p. 241).

Very highly recommended, but with the sole complaint that there was not a single graph, chart, table or photo.

Joel M. Kauffman
Professor of Chemistry Emeritus
University of the Sciences in Philadelphia
kauffman@hslc.org

References


The author's thesis is that the unprecedented range of choice available to people in modern times represents a new phenomenon that we must study to understand its present and future implications. Every choice is a decision, a word with a Latin root meaning "to cut off". Every decision cuts off the many other alternatives that we face. Even if not pathological, the desire to avoid regret following a decision can lead to paralysis, emphasizing "look before you leap" rather "he who hesitates is lost".

There is no doubt that choices in the lives of all of us have expanded. In the fields of recorded music, technological changes that once took many decades now seem to occur almost monthly. Instead of a few television channels, we now have
hundreds, by broadcast, cable, or satellite. We can now own movies and soon we may not even bother.

In reading this book, I was reminded of a Soviet professor friend I knew 25 years ago. When I invited him to lunch at our faculty club, he complained, upon seeing the menu, about the extensive choices, which numbered about 20. I did not appreciate my friend's view until I visited the Soviet Union, where choice was often very limited. The question you faced in a market was not what kind of cheese you should buy but rather was there any cheese to buy. When we enter a modern supermarket, we can choose between hundreds of cheese products by kind, method of packaging, price, etc. Yet few people are paralyzed by the extent of choice in selecting cheese or among the thousands of other items available for sale. We have learned that we need not be intimidated by choice and that we can strike a balance between making familiar choices and occasional experimentation.

The reaction to choice may be related to personality types, such as those classified by Myers-Briggs indicators. Some people prefer closure over possibility by selecting an alternative, while others like to keep present possibilities open. We need a balance between those who seek better designs and those who settle on one so that a product can be produced.

This wide-ranging, well-written book explores the issue of choice in science, religion, philosophy, art, and culture. While generally even-handed, the author betrays a distaste for free markets and privatization that is an exception. He writes,

Starting with the family, if we consider any community of any size (ultimately the global community), there is a natural dichotomy of goods and services into those that we own and those that we share, that is, those that are private and those that are public. And this fact renders the price-mechanistic, free-market, laissez-faire, egoistic system unsuitable as an approach to solve all of our socioeconomic problems.

It is clear that Mr. Rosenthal has not learned much from Milton Friedman. It was once thought that both telephone and postal services must be run by government monopolies. Within the last several decades we have seen how we now have wonderful new choices in these areas by taking most of them out of government's hands. A comparison of how many stamps we buy today with how much of the content of our mailboxes we immediately discard as junk mail shows both the benefits of choice and the consequences of lack of choice. Consider why we do not receive junk mail from package delivery companies.

On the whole, the author fares better when he discusses subjects other than economics. At one point he suggests that insurance could not be provided by a free market economy, perhaps thinking that Lloyd's of London was a British government agency.

My recommendation: read this book as a useful background for your own thoughts on choice and not for the opinions of the author.

RONALD A. HOWARD
rhoward@sdg.com
FURTHER BOOKS OF NOTE


Interesting nuggets are buried in this little book; for example:

- Comparison of Palaeolithic with modern diets (p. 11ff);
- The population of Mesoamerica was almost the same in pre-Columbian times as now (p. 53);
- Of hundreds of potentially useful crops, only a few handfuls are presently being farmed (pp. 8–9). Ancient Mesoamericans had domesticated about 50 plant species, more than in the Old World (p. 61);
- Chia, one of the four main pre-Columbian (Aztec) crops, is nowadays a mere curiosity (p. 10);
- Chia and other native crops were banned by the Spanish for religious reasons (p. 75ff). (Chia dough had been used in a similar fashion as the wafers in Communion, and Aztec priests preached values similar to those embodied in the Ten Commandments.);
- Modern American society has changed with staggering rapidity, from 40% rural in 1900 to less than 2% a century later (p. 11).

Though the book cites research articles pertinent to these matters, non-specialists would prefer a well-digested summary or survey, and that is not provided here. The book could have benefited from a strong editorial hand, to determine at what audience it is aimed, to eliminate repetitions, and to organize the topics more efficiently.

The book is part research report, with a profusion of data that can be meaningful only for other researchers. In another part it is an intended stimulus for bringing back chia as an important crop, complete with suggestions for how to make it commercially viable. That aim seems worthy enough. Chia is extraordinarily rich in the omega-3 fatty acids that are drastically too low in most modern diets. It seems superior in this respect to the fish and flax oils common in supplements nowadays. Aztecs and their descendants made from chia not only solid foods but also beverages, medicinal products, and oils particularly suited for varnishes, including on paintings. However, medicinal applications seem less plausible when an Aztec formula calls for addition of tail of opossum (p. 69) or of a sucking puppy (p. 70).

In yet another part, the book is a tract paying obeisance to the wisdom about diets and food production of the pre-Columbian natives. Thus, "pre-Columbian diets were superior to present diets" (p. 149); and "The Nahua botanical system was not only accurate but also extremely practical and very didactical" (p. 87ff); "Today, use of the name chia to denote different species of the same genus as well as plants from different genera ... could be due to the post-Columbian confusion that occurred when the Nahua botanical classification was abandoned" (p. 89).

Also unfortunate is that the book repeats as proven fact some things that are not, for example, concerning cholesterol. With the aim, apparently, of demonstrating
chia’s superiority over any other sources of omega-3 fatty acids, it is said (p. 122) that in Europe, fish allergies affect about 22% of infants and children (and as high as 39% in Sweden); that seems too high, given that in the United States the estimate is that only 8% of children suffer from any food intolerance, and only 2.3% of adults from allergies to seafood (American Academy of Allergy, Asthma, and Immunology; available at: http://www.aaaai.org/patients/resources/fastfacts/food-allergy.stm).

Sadly, the promise of the nuggets listed at the beginning is not borne out by the bulk of this work.

HENRY H. BAUER
hhbauer@vt.edu
www.henryhbauer.homestead.com

ARTICLES OF INTEREST


It is no secret that parapsychological topics are rarely featured in mainstream discourse. It is also the case that within parapsychology, there is often a reluctance to discuss claims of mediumship and other matters where personal ("subjective") experiences are paramount.

This article has an interesting discussion of some of the relevant factors. The abstract reads:

Although the feminist-critique-of-science literature has paid much attention to epistemological issues, it has generally ignored ontological issues, in particular those having to do with anything that might challenge the axiomatic materialism of mainstream Western science. Although other feminists have embraced various spiritual concerns, feminist critics of science have largely avoided all consideration of gender issues surrounding the exclusion of spiritual entities from mainstream science. An examination of the relevant literature, especially that dealing with ecofeminism, reveals that much hostility to spirit-talk is rooted in the characteristic materialism and anti-religious sentiment of the political left, where spiritual concerns are perceived as being opposed to political engagement. Further opposition comes from a generalized hostility to all perceived dualisms and from a postmodernist dismissal of the issue of the truth (or not) of what passes for knowledge. The widespread feminist privileging of the epistemological role of personal experience offers an opening for a grounded consideration of the need to postulate the existence of spiritual agencies in order to render an intelligible explanation of well attested phenomena. Parallels are drawn between this situation and the epistemological and ontological challenges facing parapsychology.

HENRY H. BAUER
hhbauer@vt.edu
www.henryhbauer.homestead.com
LETTER TO THE BOOK REVIEW EDITOR

What is Potential? Arp responds to Ibison

In JSE 21, no. 1, pages 219–223, Michael Ibison gives a helpfully lucid explanation of the problem of instantaneous communication in General Relativity. The crux of the matter seems to be "... gauge (which relates potentials to forces)." And then "... potentials propagate at light speed but forces remain pointing towards the instantaneous position."

It would seem that, in order for the potential to yield a force which points to an instantaneous position, it would have to receive information instantaneously. But that is forbidden in General Relativity. For me, defining potential as something that gives you the right answer is not satisfactory. What is potential? Is it energy? Matter? Or is it a mathematical symbol when differentiated yields another intangible quantity that is supposed to explain the "pulling toward each other" observed in physical bodies?

Mathematics I think is even more prone than physics to the tendency to give observations a name or symbol and then advertise that a phenomenon has been explained.

HALTON ARp
Max-Planck-Institut fuer Astrophysik
Garching, Germany
arp@mpa-garching.mpg.de