APPENDIX C

Correspondence between Prof. P. Diaconis & Prof. R. Aumann

5 1 page Stanford University

STANFORD, CALIFORNIA 94305

DEPARTMENT OF STATISTICS
SEQUOIA HALL

September 5, 1990

Professor Robert Auman
Department of Economics
Mail Code 6072
Stanford University
Stanford, CA 94305

Dear Bob:

I am glad to report we are in agreement about the appropriate testing procedure for the paper by Rips et al. A permutation test is to be performed. There are four basic sets of data/test statistics, I will call them additive, multiplicative, with and without Rabbi. For each there is a 32×32 table of distances. It is my understanding that for each such table, one million permutations will be performed. For each permutation $\sum_{i=1}^{32} t_{i\pi(i)}$ will be computed. This gives one million numbers/table. Again for each the number Σt_i will be located. If it is within 1/4000 of the smallest table sums, that test is judged a success. If one of the four tests is successful, the whole experiment is.

In case of ties, the interval of ties will be broken at random. If half the proportion of such breaks ammount to better than 1/4000, that table is successful. Otherwise not.

I hope that the authors agree to make their findings public no matter what the outcomes. Please let me know when you need more input from me.

Sincerely,

Persi Diaconis

Persi Diaconis

PD:kd

[C] 1 page

Robert J. Aumann
Institute of Mathematics
THE HEBREW UNIVERSITY OF JERUSALEM
91904 Jerusalem, ISRAEL

Telephone: 972-2-584327, 584578 Home: 972-2-638264, 639069
Messages: 972-2-660432 Electronic mail: aumann@HUJIVMS.bitnet

September 7, 1990

Professor Persi Diaconis
Department of Statistics
Stanford University
Stanford, CA 94305

Dear Persi,

Thanks for your good letter of September 5, about the paper submitted by Rips et al. to the PNAS.

Since it's important to clarify the precise rules of a statistical test before performing it, allow me to set down here a few points of clarification.

1. The same 1,000,000 permutations may be used for each of the four basic tests. The million will consist of the identity permutation plus 999,999 others. All million will be different from each other.

2. The sample to be examined is that of their "second experiment" (Table 3 of their submission). For each of the four basic tests, the exact same procedures as reported on in their paper (Tables 5 and 7) will be done for each of the 1,000,000 permutations. (Incidentally, "bunching" or "twenty percent" might be a more suggestive name for the test you call "additive.")

3. The precise tie-breaking rule (agreed on by phone today) is this: Out of the million permutations, let there be s that are ranked smaller than the identity, and t with which it is tied (excluding itself). Then the test is successful if and only if $s + (t/2) < 250$.

Again, with many many thanks for all your help on this,

Sincerely,

Bob Aumann

Given to Persi by hand in Sequoia Hall, September 7, 1990, 2:50 PM. He looked it over and approved.

8 1 page

Institute of Mathematics
THE HEBREW UNIVERSITY OF JERUSALEM
91904 Jerusalem, ISRAEL

Telephone: +972-2-584327 Electronic mail: aumann@HUJIVMS.bitnet
Department Fax: +972-2-630702 Home phones: +972-2-638264, 639069

December 6, 1991

Professor Persi Diaconis
Department of Statistics
Harvard University
Science Center
1, Oxford Street
Cambridge, MA 02138

Dear Persi,

Enclosed is the paper of Witztum, Rips, and Rosenberg. The presentation was revised somewhat to make it clearer and take into account the comments of the people to whom I had previously shown it. Needless to say, the test itself was not changed in any way; it is precisely the one to which we agreed in the summer of 1990. NOTE

The delay in getting this to you after I informed you of the result by phone is due to the time it took for rewriting, and for computing the results for the control text R (see Figure 1). Also, one of the authors (Rips) is on reserve duty, and this caused additional delays.

May I ask you to write an official referee report for the Proceedings of the National Academy of Sciences? Enclosed are two sets of forms (one "spare"). If you prefer, you may write your comments on a separate sheet of paper, sign it, and attach it to the form. In any case, please do answer the questions on the form and sign that as well.

If possible, I would appreciate it your sending the report by an express (courier) service. If this is not convenient, please remember to use airmail.

Thanks a lot for your help on this.

With my best wishes for a merry Christmas and a happy New Year,

Yours,

May 7, 1990

Robert Aumann
Institute of Mathematics
Hebrew University
Jerusalem, Israekm 91904

Dear Bob,

I am sending this letter via David Kazhdan and trust it will reach you. I have four points to make on the test proposed by Witzum, Rips, and Rosenberg.

1. For publication of such a fantastic claim I think a signicance level of 1/1000 or better should be required. I arrived at this number after consultation with a variety of colleagues. The authors present the claim that there is a strong effect. If this is the case, there will be no trouble attaining far higher levels.

2. As I understand it, computational considerations are an issue. The following suggestion should make the computations feasable. Let X and Y be spaces and $d(x, y)$ a function from $X \times Y \to [0,1]$. As I understand it, there is a fixed set of matched pairs $(x_1, y_1)(x_2, y_2), \ldots, (x_n, y_n)$ and a statistic $T = d(x_1, y_1) + \ldots + d(x_n, y_n)$. Permuting (say) the x's we get a number $T(\pi)$ for each permutation. The computation of $d(x, y)$ is expensive. NOTE
   If e.g. $n = 30$, and $N = 100$ permutations were selected (which could only offer significance of 1/100 of course) then $nN = 3000$ pair computations of $d(x, y)$ are required.

   Here is my suggestion. Calculate the $n \times n$ matrix $d(x_i, y_i)$ $1 \leq i, j \leq n$. This requires n^2 expensive computations. With this in hand, one can cheaply generate millions of permutations and calculate $T(\pi) = d(x_{\pi(1)}, y_1) + \ldots + d(x_{\pi(n)}, y_n)$ from the matrix.

Indeed, with the matrix at hand, a million samples can be run off in a few seconds. Thus, if formation of the matix is feasable, I recommend carrying this procedure out with N = a million and then seeing what proportion of the $T(\pi)$ are smaller than $T(id)$ as indicated in your letter. A clear provision should be agreed upon in the call of tied values. I do not foresee a problem here.

3. The permutation test can be carried out to satisfactory approximation without performing any permutations. This uses Hoefdings Combinatorial central limit Theorem. This gives a normal approximation to the distrubutions of the sum. The variance depends on the entries of the matrix described above in a simple way. It is this kind of theorem which is lacking in the quantification proposed originally. A recent reference is Bolthausen, E. (1984). An estimate of the remainder in a combinatorial central limit theorem. Z. Wahrscheinlichkeits-theorie , Verw. Gebiete. 66 379-386. This gives references to earlier literature.

4. The authors have changed the basic statistic a few times since we began our correspondence. I think the present version is much clearer: a single number for a list of pairs versus a list of n numbers. Becuase of these changes, I feel a fresh set of 30 names and birthdates should be used. This is standard practice. In matters of this sort, no one will be convinced because of the possibility of some sample permutations having been drawn between our correspondence. Indeed, as we know, a second permutation was drawn and a standard statisical test performed which showed that nothing suprising appears in the sample of 30 names.

I realize that the above is a stingent test but it is the minimum that would satisfy me. I am sorry for the delay, but I really did just receive your material recently.

I will be happy to clear up further details or discuss the suggestions made above. I will be at Stanford, Department of Statistics, from Juen 15-Sept. 30, and then back here.

Best wishes,

Persi

Persi Diaconis

5. The points raised above are completely disjoint from a review of the current version of the paper. This quotes many "probabilities" which have no justification as I understand. Persi

Below are copies of three pages from the original WRR paper sent to Diaconis for approval before the probabilities were computed by use of the permutation test. On this page, the proximity measures are precisely defined in detail. The next page contains a precise description of the permutation test. The third page is that part of the same paper where the resulting probability is given as a question mark, since the calculation had not yet been done.

Due to the restrictions of our scheme (see Section 5), we used only words with at least[15] 5 and at most[16] 8 letters. (That is the reason that some personalities have only two date designations, and Rabbi Menahem Mendel Krochmal (Number 26) has only one.)

7. The Overall Proximity Measures P_1, P_2, P_3, and P_4.

The first graph in Figure 4 (marked G for Genesis) shows the distribution of the distances $c(w, w')$ for the first list[17] of personalities. The other graph shows the corresponding distributions when the control text R replaces G. Text R is obtained by mixing the letters of G using a quasi-random function.

For G, the $c(w, w')$ appear concentrated near zero. This is not so for the control text, where the distribution looks quite random. This might indicate that for G, the words w are indeed closer to the w' than one would expect. To check this, we define two statistics, P_1 and P_2, that quantify the "degree of concentration" of the $c(w, w')$ near zero—and so, indirectly, the "overall proximity" of the words w in the sample to the words w'.

Intuitively, both P_1 and P_2 are based on the following *Uniformity and Independence Assumption* (UIA): that in the absence of the systematic effect we seek to establish, one may expect the $c(w, w')$ to be independent and uniformly distributed (i.e., to take on each possible value with the same probability). The construction of the $c(w, w')$ makes the UIA sound not unreasonable; but we have no proof. So we make no formal use of the UIA; it is used only for motivation.

To define P_1, let n be the total number of word pairs in the sample[18], and S the number of word pairs (w, w') with $0 < c(w, w') \leq 0.2$. (This interval was selected arbitrarily to represent numbers "fairly close" to zero.) We then determine the probability under the UIA that S would be as large as it is; this probability is denoted P_1 (the smaller it is, the higher the concentration). Under the UIA, S is binomially distributed, with $p = 0.2$. The results are given in Tables 3 and 4.

The statistic P_1 ignores all distances $c(w, w')$ greater than 0.2, and gives equal weight to all distances less than 0.2. For a measure that is sensitive to the actual size of the distances, we calculate the product $\Pi c(w, w')$ over all word pairs (w, w') in the sample. To see how this product would be distributed under the UIA, let there be n word pairs. If $x_1, x_2, \ldots, x_n$ are independent

[15] At least 5 letters are needed to apply (x, y, z)-perturbations.

[16] If w has more than 8 letters, there are not enough triples (x, y, z) for which there exist (x, y, z)-perturbated ELS's for w.

[17] With at least 3 columns in Margalioth [2] (see Section 2).

[18] After removing word pairs with $m(w, w') < 10$ (see Section 5).

Note that personality–date pairs (p, p') are not word pairs. The personalities each have several appellations, there are variations in spelling, different ways of designating dates, and so on. Thus each personality–date pair (p, p') corresponds to several word pairs (w, w'). The precise method used to generate a sample of word pairs from a list of personalities is explained in Section 6.

As explained in Section 7, we also used a variant of this method, which generates a smaller sample of word pairs from the same list of personalities. We denote the statistics P_1 and P_2, when applied to this smaller sample, by P_3 and P_4.

Finally, we come to Task (iv), the significance test itself. It is so simple and straightforward that we describe it in full immediately.

The second list consists of 32 personalities. For each of the 32! permutations π of these personalities, we define the statistic P_1^π obtained by permuting the personalities in accordance with π, so that Personality i is matched with the dates of Personality $\pi(i)$. The 32! numbers P_i^π are ordered, with possible ties, according to the usual order of the real numbers. If the phenomenon under study were due to chance, it would be just as likely that P_1 occupies any one of the 32! places in this order as any other. Similarly for P_2, P_3, and P_4. This is our null hypothesis.

To calculate significance levels, we chose 999,999 random permutations π of the 32 personalities, using a standard program for choosing random permutations.[5] The significance level of P_1 is the number of these π for which P_1^π exceeds[6] P_1, divided by 1,000,000. Similarly for P_2, P_3, and P_4.

After calculating the significance levels of P_1 through P_4, we must make an overall decision to accept or reject the research hypothesis. Define ρ_i as the probability, under the null hypothesis, that P_i would be as low as it is; thus $\rho_i = (1 -$ the significance level of $P_i)$. Set $\rho_0 := 4 \min \rho_i$. Then the overall significance level, using all four statistics, is at least[7] $1 - \rho_0$.

[5] The random permutations were chosen in accordance with Algorithm P on p.125 of Knuth[3]. The pseudo-random generator required as input to this algorithm was that provided by Turbo-Pascal 5.0 of Borland Inter Inc.

[6] If P_1 is tied with other P_1^π, half of these others are considered to "exceed" P_1.

[7] One must avoid selecting favorable evidence only. For example, suppose that $\rho_3 = .01$, the other ρ_i being higher. There is then a temptation to consider ρ_3 only and so to reject the null hypothesis at the level of 99%. But this would be a mistake; with enough sufficiently diverse statistics, it is quite likely that just by chance, some one of them will be low. The correct question is, "Under the null hypothesis, what is the probability that at least one of the four ρ_i would be $\leq .01$?" Thus denoting the event "$\rho_i \leq .01$" by E_i, we must find the probability not of E_3, but of "E_1 or E_2 or E_3 or E_4." If the E_i were mutually exclusive, this probability would be .04; overlaps only decrease the total probability, so that it is in any case $\leq .04$. Thus we can reject the null hypothesis at the level of 96%, but not 99%.

3. The Results.

The individual significance levels $1 - \rho_i$ of the four statistics P_i $(i = 1, 2, 3, 4)$, as well as the minimum overall significance level $1 - \rho_0$, are given in Table 1 in the column marked G (for Genesis). The research hypothesis is thus confirmed at the level of ??% at least. NOTE

4. The Distance Between Words. [8]

To define the "distance" between two words w and w', start by letting e and e' be fixed ELS's that spell out w and w', and consider Genesis written out in a fixed two-dimensional array. Set

$$
\begin{aligned}
f &:= \text{the distance}^9 \text{ between consecutive letters of } e,\\
f' &:= \text{the distance between consecutive letters of } e',\\
\ell &:= \text{the minimal distance between a letter of } e \text{ and one of } e',
\end{aligned}
$$

and define $\delta(e, e') := f^2 + f'^2 + \ell^2$. We call $\delta(e, e')$ the *distance* between the ELS's e and e' in the given array; it is small if both fit into a relatively compact area. For example, in Figure 3 we have $f = 1$, $f' = \sqrt{5}$, $\ell = \sqrt{34}$, $\delta = 40$.

Now there are many ways of writing Genesis as a two-dimensional array, depending on the row length h. Denote by $\delta_h(e, e')$ the distance $\delta(e, e')$ in the array determined by h, and set $\mu_h := 1/\delta_h$; the larger $\mu_h(e, e')$ is, the more compact is the configuration consisting of e and e' in the array with row length h. Set $e = (n, d, k)$ (recall that d is the skip) and $e' = (n', d', k')$. Of particular interest are the row lengths $h = h_1, h_2, \ldots$, where h_i is the integer nearest to $|d|/i$. Thus when $h = h_1 = |d|$, then e appears as a column of adjacent letters (as in Figure 1); and when $h = h_2$, then e appears either as a column that skips alternate rows (as in Figure 2), or as a straight line of knight's moves (as in Figure 3). In general, the arrays in which e appears relatively compactly are those with row length h_i with i "not too large."

Define h'_i analogously to h_i. The above discussion indicates that if there is an array in which the configuration (e, e') is unusually compact, it is likely to be among those whose row length is one of the first ten[10] h_i, or one of the first ten h'_i. So setting

$$\sigma(e, e') := \sum_{i=1}^{10} \mu_{h_i}(e, e') + \sum_{i=1}^{10} \mu'_{h_i}(e, e'),$$

[8] This section is rather technical, and may be omitted at a first reading.
[9] Ordinary Euclidean distance in the two-dimensional array.
[10] Ten is an arbitrarily selected "moderate" number.

APPENDIX D

Review of Dr. Randall Ingermanson's "Who Wrote the Bible Code?"

The book "Who Wrote the Bible Code?" by Dr. Randall Ingermanson describes a statistical test done on the Hebrew text of Genesis to detect the presence of Torah codes. He concludes that there cannot be "a Bible code in which encoded ELSs run rampant. However, a believer can postulate a 'sparse Bible code' with a few real codes hidden like golden needles in a vast haystack. My work does not *absolutely* disprove this idea." (Note added in proof.) In addition, Ingermanson's technique cannot detect ELSs with skip distance between the letters less than about 50. Since Ingermanson's results are compatible with the codes that have been discovered and verified so far, this is not a challenge to WRR or the additional experiments described. Thus, we have not included a discussion of this result in the body of this paper. For those who are interested, a review by Professor Robert Haralick is included below that reveals a logical flaw in Dr. Ingermanson's reasoning. At the time of this writing, there is an ongoing dialogue concerning this matter between Prof. Haralick and Dr. Ingermanson.

Review of "Who Wrote the Bible Code?"
By Randall Ingermanson

Reviewed by Robert M. Haralick
Dept. of Electrical Engineering
University of Washington
Seattle, WA 98195

The Ingermanson logic:

(1) "If the Torah contains so much information embedded as ELSs in the text, then the entropy of these ELSs in the Torah must be lower than we would ordinarily expect." p 70.

(2) "If the believer's [of the Torah code hypothesis] are right, then the ELSs in each skip text taken from the Bible will be measurably different from those you'd predict in a random text." p 86.

(3) "If their [the believer's] interpretation is correct, the Torah must be chock-full of ELSs at many different skips. No matter which skip we consider, we ought to see many more meaningful ELSs than random chance predicts. This means that every skip-text must contain many more meaningful words (spelled both backward and forward) than you'd expect to see in a random text.

The digram and trigram frequencies of intentionally encoded words are different from those you'd expect by random chance, and they result in different digram and trigram entropies than those you'd get by random chance." p 86-87.

(4) "If the skeptics [of the Torah code hypothesis] are right, we expect that skip-texts taken from the original will have the same distribution of words, on average, as random skip-texts provided the skip is large enough." p 87.

Ingermanson then makes the entropy calculation for digrams and trigrams of Torah skip texts and finds that for skips greater than around 50 the Torah skip text digrams and trigrams have the same entropy as randomized texts. He concludes that there is no more structure in the Torah skip text ELSs than expected by chance and, therefore, the Torah code hypothesis must be false.

In summary, Ingermanson argues that if the Torah code hypothesis is correct, [this is the premise] there ought to be more ELSs and if there are more ELSs there will be more statistical structure or order in the skip texts and therefore, the entropy of Torah skip texts ought to be lower than the corresponding entropy of randomized Torah skip texts [this is the consequence].

He makes the measurements and finds that the entropy of the Torah skip texts are not lower than the corresponding entropy of randomized Torah skip texts. Having provided evidence that the consequence is not correct, he concludes that the premise is false.

The argument is fallacious because Ingermanson seems not to understand the Torah code hypothesis. The Torah code hypothesis is that there are some domains of logical relationships where if one collects together clusters of key words that are logically related from the domain, then there will be a higher probability that there are more corresponding clusters of ELSs that are more compact (spatially close) in the Torah text we have today than expected in a population of randomized Torah texts.

The Torah code hypothesis does not imply as Ingermanson argues that if the Torah code hypothesis is correct, there ought to be more ELSs. The Torah code hypothesis is completely consistent with a condition that the number and kind of ELSs are exactly what would be expected by chance. The Torah code hypothesis states that the placement of the ELSs in the Torah text is skewed in such a way that there is a higher frequency of ELSs of related key words that appear closer together than expected by chance.

So basically, what Ingermanson has done is to restate the Torah code hypothesis in a way that is not equivalent to the true Torah code hypothesis, and then he provided evidence that his restatement of the Torah code hypothesis must be false. His evidence has no bearing on the correctness or incorrectness of the true Torah code hypothesis.