

# Basic Concepts For Testing The Torah Code Hypothesis

Robert M. Haralick  
Computer Science, Graduate Center  
City University of New York  
365 Fifth Avenue  
New York, NY 10016

## Abstract

*This is the first part of a tutorial discussing the major strategies and methodologies by which a test of the Null hypothesis of no Torah effect can be done. The basic concepts of equidistant letter sequence, skip specification, resonance specification, and compactness features are discussed here.*

## 1. Introduction

The application of pattern recognition methodologies are not new to religious areas. For example Ikeuchi's use of a 3D laser scanner to obtain range data on religious objects such as the Great Buddha of Kamakura is one such project[2]. Ikeuchi was able to process the range data and create a complex 3D mesh surface model of the Buddha and then give it the gold leaf appearance as it was when it was originally built.

In this paper we discuss another kind of application of pattern recognition and statistical methods to a religious area. We discuss the Torah codes, a topic that has had considerable popular interest with three documentary video productions. Also it has been involved in a great academic controversy ever since the publication of the first formal study by Witztum, Rips, and Rosenberg[4] and the subsequent claim that any apparently successful experiment result must be due to a non a priori data selection[3].

The Torah codes center on the Hebrew Torah text, the five books of Moses. There is some evidence that key words which are historically/logically related have their equidistant letter sequences in a more compact geometric arrangement in the Torah text than expected by chance. This phenomena, if it is really there, is surprising. Here, we do not review the various experiments that have been done. Nor do we discuss the controversy itself. That discussion can be found in Haralick, Rips, Glazerson[1]. Rather, we lay out in a tutorial way, the basic concepts and kinds of pattern

features that are being used in the Torah code investigation and discuss two experiments that have been done with this methodology.

## 2. Definitions

A word  $w$  of length  $K$  is a sequence of  $K$  characters  $w = \langle w_1, \dots, w_K \rangle$ . A text  $T$  is the character string of the text with spaces, punctuation marks and all symbols other than the letters of the alphabet removed. A text is just a very long word. Let  $T = \langle t_1, \dots, t_Z \rangle$  be a given text. The *letter frequency* of alphabet letter  $\alpha$  is just the number of the times the letter  $\alpha$  occurs in the text. It is given by

$$f(\alpha) = \#\{z \mid t_z = \alpha\}$$

The *probability of occurrence* of letter  $\alpha$  is given by

$$p(\alpha) = \frac{f(\alpha)}{Z}$$

An *equidistant letter sequence*, called ELS for short, is a sequence of equally spaced letters in the text not counting spaces and punctuation marks. The sequence of the letter positions form an arithmetic progression. Several properties associated with an ELS  $e$  are:

- $B(e)$ : the beginning position of ELS  $e$ ,
- $E(e)$ : the ending position of ELS  $e$ ,
- $L(e)$ : the number of characters in ELS  $e$ ,
- $S(e)$ : the skip of ELS  $e$ , and
- $W(e)$ : the character string  $\langle W(e)_1, \dots, W(e)_{L(e)} \rangle$  of ELS  $e$

These properties have two constraints:  $B(e) < L(e)$  and the relation binding the end position to the beginning position.  $E(e) = B(e) + (L(e) - 1)|S(e)|$ .

The *positions* determined by the ELS  $e$  are given by

$$B(e), B(e) + |S(e)|, \dots, B(e) + (L(e) - 1)|S(e)|$$

Character  $W(e)_i$  of ELS  $E$  is associated with position  $B(e) + (i - 1)|S(e)|, i = 1, \dots, L(e)$ . The *span* of an ELS  $e$  is given by  $E(e) - B(e) + 1 = 1 + (L(e) - 1)|S(e)|$ . ELS  $e$  is said to be an *ELS of key word*  $w$  when  $w = W(e)$ . The skip  $S(e)$  can be positive or negative depending on whether the ELS positions match in a forwards or backwards order. We call the first kind of ELS a positive skip ELS and the second kind of ELS a negative skip ELS. ELS  $e$  is said to be a *positive skip ELS* of a word  $w$  whose respective characters are  $w_1, \dots, w_{L_w}$  if and only if  $L_w = L(e)$  and  $w_i = W(e)_i, i = 1, \dots, L_w$ . ELS  $e$  is said to be a *negative skip ELS* of a word  $w$  whose respective characters are  $\langle w_1, \dots, w_{L_w} \rangle$  if and only if  $L_w = L(e)$  and  $w_i = W(e)_{L(e)+1-i}, i = 1, \dots, L_w$ .

An ELS  $e$  is said to be an *ELS of a word*  $w$  in a text  $T$  if and only if it is an ELS of word  $w$  and

$$T_{B(e)+i|S(e)|} = \begin{cases} w_{i+1}, i = 0, \dots, L(e) - 1 \\ \text{when } S(e) > 0 \\ w_{L_w-i}, i = 0, \dots, L(e) - 1 \\ \text{when } S(e) < 0 \end{cases}$$

The *set of all ELSs*  $\mathcal{E}$  associated with a word  $w = \langle w_1, \dots, w_K \rangle$  and text  $T$  is given by

$$\mathcal{E}(w, T) = \{e \mid \begin{array}{l} T_{B(e)+iS(e)} = w_{i+1} = W(e)_{i+1}, \\ i = 0, \dots, K - 1, \text{ when } S(e) > 0; \\ T_{B(e)+i|S(e)|} = w_{K-i} = W(e)_{K-i}, \\ i = 0, \dots, K - 1, \text{ when } S(e) < 0 \end{array}\}$$

If we want to name the set of ELSs for a key word  $w$  in a text  $T$  with respect to a general skip specification  $\sigma$ , we will write  $\mathcal{E}(w, T, \sigma)$ .

### 2.1. Number of Placements

The *number of possible placements* for an ELS  $e$  of skip  $S(e)$  in a text of length  $Z$  is  $Z - (L(e) - 1)|S(e)|$ . So the number  $N$  of possible placements for ELSs of absolute skip from smallest skip  $S_{min}$  to largest skip  $S_{max}$  is

$$N = \frac{(S_{max} - S_{min} + 1)}{2} * (2Z - (L(e) - 1) * (S_{max} + S_{min}))$$

If matching is allowed both in the forward direction and reverse direction, then the number of possible placements is exactly double the expression above, providing the key word is not symmetric (spelled the same way forward and backwards).

## 3. Number of ELSs

Given a text  $T$  of  $Z$  characters, there is a corresponding text population of  $Z!$  texts corresponding to all the  $Z!$  letter permutations of the text  $T$ . In the letter permuted text population, the probability  $p$  that any given placement of the letters of the key word  $w = \langle w_1, \dots, w_K \rangle$ , will match the letters in the placement position is given by

$$p = \prod_{k=1}^K p(w_k)$$

The probability for observing a given number of ELSs depends on the control text population and the minimum and maximum skip ELS that is searched for.

In the case of a letter permuted text population, having placement match probability  $p$  for a given key word, the probability that  $K$  ELSs will be found for a key word in a search of  $N$  placements is given by the binomial probability

$$Prob(K \mid p, N) = \frac{N!}{K!(N - K)!} p^K (1 - p)^{N - K}$$

### 3.1. Expected Number of ELSs

Given a key word  $w$ , a minimum absolute skip  $S_{min}$  and a maximum absolute skip  $S_{max}$ , we associate with each text  $T'$  in the letter permuted population the set  $\mathcal{E}(w, T', S_{min}, S_{max})$ . This set is the set of all ELSs of word  $w$  in the text  $T'$  that have absolute skips in the interval  $[S_{min}, S_{max}]$ . This set has a size: the number of ELSs it contains. The arithmetic average of the sizes of the ELS sets taken over all the texts of the population is defined as the expected number of ELSs.

In a population of letter permuted texts, each of length  $Z$ , the expected number of ELSs of a key word  $w = \langle w_1, \dots, w_K \rangle$  is given by  $pN$  where

$$p = \prod_{k=1}^K p(w_k)$$

$$N = \frac{(S_{max} - S_{min} + 1)}{2} * (2Z - (L - 1) * (S_{max} + S_{min}))$$

### 3.2. Poisson Probability Approximation

In the case when  $p$  is small and  $N$  is large, the binomial probability can be approximated by the Poisson probability

$$Prob(K \mid p, N) = \frac{e^{-pN} (pN)^K}{K!}$$

since

$$\sum_{k=1}^N k \text{Pr}ob(k | p, N) = pN$$

## 4. Skip Specification

One of three criteria may be used to specify the skip ELS set of a key word: Expected Number of ELSs, Largest Skip, and Largest Possible Skip. Each criterion has as well a specified smallest skip  $S_{min}$ . The most convincing Torah codes are often found using the expected number of ELS search criterion.  $S_{min}$  is typically set to 1 or 2.

### 4.1. Expected Number

The expected number criterion sets the largest skip to be searched for to be the smallest skip  $S_{max}$  making the expected number of ELSs in a randomly sampled text from a letter permuted population be just larger than a given  $N$  when the smallest absolute skip is 2. This is the protocol followed by WRR. WRR sets  $N$  to be 10 for the Genesis text of 78,064 letters. (Even though the expected number is computed for a minimum skip of 2, the skip specification is free to choose  $S_{min} = 1$  or  $S_{min} = 2$ ). We let  $\sigma$  denote the skip specification and  $\mathcal{E}(w, T, \sigma)$  the set of all ELSs of word  $w$  in text  $T$  satisfying the skip specification  $\sigma$ .

### 4.2. Fixed Maximum Skip

The fixed largest skip criterion, just specifies a given value for  $S_{max}$ .

### 4.3. Largest Possible Skip

The largest possible criterion sets  $S_{max}$  as large as possible while keeping the span of the ELS within the text, here assumed to have length  $Z$ . Hence  $S_{max}$  satisfies

$$S_{max} = \left\lfloor \frac{Z-1}{L-1} \right\rfloor$$

where  $\lfloor \cdot \rfloor$  designates the floor function. The largest possible criterion makes the skip ELS set for a word  $w$  be the set of all ELSs of the word  $w$  in the text  $T$ .

## 5. ELS Row and Column Skip on the Cylinder

When the text, with no spaces and punctuation characters, is spiraled around a cylinder of  $\gamma$  columns, an ELS of

absolute skip  $s$  will appear on the cylinder with *row skip*  $s_r$  and *column skip*  $s_c$  given by

$$s_r = \begin{cases} \lfloor s/\gamma \rfloor & \text{if } s \bmod \gamma \leq \gamma - s \bmod \gamma \\ \lceil s/\gamma \rceil & \text{otherwise} \end{cases}$$

A skip  $s$  ELS converts to a *column skip*  $s_c$  on a cylinder of size  $\gamma$  where

$$s_c = \begin{cases} s \bmod \gamma & \text{if } s \bmod \gamma \leq \gamma - s \bmod \gamma \\ -(\gamma - s \bmod \gamma) & \text{otherwise} \end{cases}$$

We say that the column skip of an ELS on a cylinder is positive if  $s \bmod \gamma \leq \gamma - s \bmod \gamma$ . This corresponds to the condition when the closest way to reach successive letters of the ELS is by proceeding clockwise around the cylinder. We say that the column skip of an ELS is negative when  $s \bmod \gamma > \gamma - s \bmod \gamma$ . This corresponds to the condition when the closest way to reach successive letters of the ELS is by proceeding counterclockwise around the cylinder.

## 6. Resonance Specification

The Torah code phenomena seems to more heavily involve ELSs and cylinder sizes where the relation is that the row skip  $s_r$  and the column skip  $s_c$  are both sufficiently small. A skip of size  $s$  and a cylinder of size  $\gamma$  are said to *resonate* when  $s_r$  and  $s_c$  are sufficiently small.

There are five simple criteria that have been used in various experiments to make this specification: max skip, area, diagonal, and perimeter. Let  $s = s_r\gamma + s_c$ . Then skip  $s$  resonates with cylinder size  $\gamma$  when

- $s_r \leq s_{rmax}$  and  $s_c \leq s_{cmax}$  for the max skip criterion
- $s_r * s_c \leq a_{max}$  for the area criterion
- $s_r^2 + s_c^2 \leq d_{max}^2$  for the diagonal criterion
- $2(s_r + s_c) \leq per_{max}$  for the perimeter criterion
- $s_r \leq s_{rmax}$  for the WRR criterion

On the basis of the resonance specification  $\phi$ , we may define the resonance relation  $Res$ .

$$Res(\phi) = \{(\gamma, s) \mid \text{skip } s \text{ is } \phi \text{ resonant with cylinder size } \gamma\}$$

Depending on the arguments of  $Res$ , we overload it in accordance with the following definitions.

$$Res(\gamma, \phi) = \{s \mid (\gamma, s) \in Res(\phi)\}$$

$$Res(s, \phi) = \{\gamma \mid (\gamma, s) \in Res(\phi)\}$$

## 7. Distance on the Cylinder

Many varieties of compactness definitions involve the concept of the distance between two positions on the cylinder. From one point on a cylinder to another, there are two distinct paths: proceeding clockwise around the cylinder and proceeding counterclockwise around the cylinder. The distance between two positions is defined as the shorter of these two.

Let  $p_1$  and  $p_2$  be two text positions on a cylinder of size  $\gamma$  columns. Let  $r$  denote the row distance between the two positions and let  $c$  denote the column distance between the two positions. Then

$$r = \begin{cases} \lfloor |p_1 - p_2|/\gamma \rfloor & \text{if } |p_1 - p_2| \bmod \gamma < \\ & \gamma - |p_1 - p_2| \bmod \gamma \\ \lceil |p_1 - p_2|/\gamma \rceil & \text{otherwise} \end{cases}$$

$$c = \min\{|p_1 - p_2| \bmod \gamma, \gamma - |p_1 - p_2| \bmod \gamma\}$$

The Euclidean distance  $\Delta$  between positions  $p_1$  and  $p_2$  on a cylinder of size  $\gamma$  is then defined by

$$\Delta(p_1, p_2; \gamma) = \sqrt{r^2 + c^2}$$

## 8. Pairwise Distance Based ELS Compactness

An ELS on a cylinder of size  $\gamma$  can be regarded as a set of points. From this perspective view, the simplest compactness between two ELSs amounts to defining a distance like function between two sets of points. There are two commonly used definitions between the points of two sets: their minimum distance  $d_1$  and their maximum distance  $d_2$ . Let  $e_1$  and  $e_2$  be two ELSs with respective beginning positions  $B(e_1)$  and  $B(e_2)$ , skips  $S(e_1)$  and  $S(e_2)$ , and length  $L(e_1)$  and  $L(e_2)$ . Then we define three distances between the ELSs on a cylinder of  $\gamma$  columns by the min distance  $d_1$ , the max distance  $d_2$ , and the sum of the min and max distance by  $d_{12}$ .

$$d_1(e_1, e_2; \gamma) = \min_{\substack{i=1, \dots, L(e_1) \\ j=1, \dots, L(e_2)}} \Delta(B(e_1) + (i-1)|S(e_1)|, \\ B(e_2) + (j-1)|S(e_2)|; \gamma)$$

$$d_2(e_1, e_2; \gamma) = \max_{\substack{i=1, \dots, L(e_1) \\ j=1, \dots, L(e_2)}} \Delta(B(e_1) + (i-1)|S(e_1)|, \\ B(e_2) + (j-1)|S(e_2)|; \gamma)$$

$$d_{12}(e_1, e_2; \gamma) = d_1(e_1, e_2; \gamma) + d_2(e_1, e_2; \gamma)$$

WRR used a squared min distance modified by the squared skips as it appears on the cylinder. An ELS skip of  $s$  appears on the cylinder of size  $\gamma$  as a skip with distance  $\Delta(0, s; \gamma)$ . Based on this idea, we can define three WRR-like ELS distances.

$$\omega_1(e_1, e_2, \gamma) = d_1^2(e_1, e_2; \gamma) + \Delta^2(0, S(e_1); \gamma) + \Delta^2(0, S(e_2); \gamma)$$

$$\omega_2(e_1, e_2, \gamma) = d_2^2(e_1, e_2; \gamma) + \Delta^2(0, S(e_1); \gamma) + \Delta^2(0, S(e_2); \gamma)$$

$$\omega_{12}(e_1, e_2, \gamma) = d_{12}^2(e_1, e_2; \gamma) + \Delta^2(0, S(e_1); \gamma) + \Delta^2(0, S(e_2); \gamma)$$

In the above definitions, the three terms are weighted equally. But if the natural weights are different because in some sense the scale of the skip distance is not the same as the scale of the closest distance, it would be better to take a product. To keep the product from being zero in the case of the min distance, we bound the min distance below by a small positive constant  $\epsilon$ .

$$\rho_1(e_1, e_2, \gamma) = (\max\{d_1^2(e_1, e_2; \gamma), \epsilon\}) \times (\Delta^2(0, S(e_1); \gamma) + \Delta^2(0, S(e_2); \gamma))$$

$$\rho_2(e_1, e_2, \gamma) = d_2^2(e_1, e_2; \gamma) \times (\Delta^2(0, S(e_1); \gamma) + \Delta^2(0, S(e_2); \gamma))$$

$$\rho_{12}(e_1, e_2, \gamma) = d_{12}^2(e_1, e_2; \gamma) \times (\Delta^2(0, S(e_1); \gamma) + \Delta^2(0, S(e_2); \gamma))$$

### 8.1. ELS Set Distance Based Compactness Measures

The ELS pairwise distance based measures can be easily generalized to ELS set based measures. Every pair of distinct ELSs in a given set has a compactness. One of the pairs has a largest compactness. If we are given a set of ELSs under the hypothesis that all are compactly related, this largest compactness is a reasonable measure of compactness for the set.

Let  $E$  be a set of ELSs. Using capital letters for the set based distance, we can define the corresponding set based compactness measures for a given cylinder size  $\gamma$ .

$$D_1(E; \gamma) = \max_{\substack{e \in E \\ f \in E}} d_1(e, f; \gamma)$$

$$D_2(E; \gamma) = \max_{\substack{e \in E \\ f \in E}} d_2(e, f; \gamma)$$

$$D_{12}(E; \gamma) = \max_{\substack{e \in E \\ f \in E}} d_{12}(e, f; \gamma)$$

$$\Omega_1(E; \gamma) = \max_{\substack{e \in E \\ f \in E}} \omega_1(e, f; \gamma)$$

$$\Omega_2(E; \gamma) = \max_{\substack{e \in E \\ f \in E}} \omega_2(e, f; \gamma)$$

$$\Omega_{12}(E; \gamma) = \max_{\substack{e \in E \\ f \in E}} \omega_{12}(e, f; \gamma)$$

$$\begin{aligned}
R_1(E; \gamma) &= \max_{\substack{e \in E \\ f \in E}} \rho_1(e, f; \gamma) \\
R_2(E; \gamma) &= \max_{\substack{e \in E \\ f \in E}} \rho_2(e, f; \gamma) \\
R_{12}(E; \gamma) &= \max_{\substack{e \in E \\ f \in E}} \rho_{12}(e, f; \gamma)
\end{aligned}$$

## 9. Table Based Compactness Measures

### 9.1. Table Definition

A *table*  $T$  on a cylinder of size  $\gamma$  is a quadruple  $T = (b, n_r, n_c, \gamma)$ , where  $b$  is the beginning position of the table,  $n_r$  is the number of rows of the table, and  $n_c$  is the number of columns of the table and  $\gamma$  is the cylinder size. The table specified by  $T$  is the set of positions  $P(T) = \{p \mid p = b + r\gamma + c, 0 \leq r < n_r, 0 \leq c < n_c\}$ . A table  $T$  is said to contain the characters of an ELS  $e$  if

$$B(e) + |S(e)| * l \in P(T), 0 \leq l < L(e)$$

### 9.2. Table Formation

For any pair of ELSs  $(e_1, e_2)$ , and cylinder size  $\gamma$ , there is a table  $T(e_1, e_2, \gamma)$ , whose positions contain all the positions of ELSs  $e_1$ . Likewise, for any set  $F$  of ELSs and cylinder size  $\gamma$ , there is a table  $T(F, \gamma)$  whose positions contain all the positions of the ELSs in  $F$  and which has the smallest number of rows and columns. We call  $T(F, \gamma)$  a *table formed* from ELS set  $F$  and cylinder size  $\gamma$ .

### 9.3. Table Compactness

The *compactness*  $C(T)$  of table  $T = (\gamma, n_r, n_c)$  can be specified by one of four possible criteria: max side, area, diagonal, or perimeter.

$$C(T) = \begin{cases} \max\{n_r, n_c\} & \text{for the max criterion} \\ n_r * n_c & \text{for the area criterion} \\ n_r^2 + n_c^2 & \text{for the diagonal criterion} \\ 2(n_r + n_c) & \text{for the perimeter criterion} \end{cases}$$

## 10. Combination Methods Over Resonant Cylinder Sizes

Let  $\zeta$  be an ELS set,  $\delta$  be one of the ELS set compactness measures and  $\phi$  be a resonance specification. We define the best compactness  $\psi_{min}$  over resonant cylinder sizes by

$$\psi_{min}(\zeta; \delta, \phi) = \min\{\delta(\zeta, \gamma) : \gamma \in \bigcap_{e \in \zeta} Res(S(e), \phi)\}$$

Other combination methods over resonant cylinder sizes include taking harmonic mean  $\mu_h$  and geometric mean  $\mu_g$ .

$$\psi_{harm}(\zeta; \delta, \phi) = \mu_h\{\delta(\zeta, \gamma) : \gamma \in \bigcap_{e \in \zeta} Res(S(e), \phi)\}$$

$$\psi_{geom}(\zeta; \delta, \phi) = \mu_g\{\delta(\zeta, \gamma) : \gamma \in \bigcap_{e \in \zeta} Res(S(e), \phi)\}$$

## 11 Collections of ELS Sets Defined From The Key Word Set

Let  $W = \{w_1, \dots, w_K\}$  be a set of  $K$  key words describing an event. Let skip specification  $\sigma$  be given. Associated with each key word  $w$  of  $W$  is a set  $\mathcal{E}(w, T, \sigma)$  of its ELSs in text  $T$  in accordance with skip specification  $\sigma$ . Let  $\mathcal{Z}(W; T, \sigma)$  be the collection of all ELS sets, where each set in the collection contains exactly one ELS from each of the ELS sets  $\mathcal{E}(w, T, \sigma)$ ,  $w \in W$ ,

$$\mathcal{Z}(W; T, \sigma) = \{\{e_1, \dots, e_K\} \mid e_k \in \mathcal{E}(w_k, T, \sigma)\}$$

To evaluate whether the key words of  $W$  are encoded as compactly arranged ELSs in text  $T$ , the statistic we use must be a function defined on  $\mathcal{Z}(W; T, \sigma)$  and involving a compactness measure defined over resonant cylinder sizes.

## 12. Key Word Set Compactness

Let  $\delta$  be one of the ELS set based compactness measures,  $\psi$  be one of the combination methods defined over resonant cylinder sizes,  $\phi$  be a resonance specification, and  $\sigma$  be a skip specification. We define  $\Psi$  to be key word set compactness measure that combines compactnesses over resonant cylinder sizes with respect to the skips of the ELSs and over ELSs of the key word set. Combination methods over ELS sets include taking the minimum, the harmonic mean,  $\mu_h$ , and the geometric mean,  $\mu_g$ .

$$\begin{aligned}
\Psi_{min}(W; \psi, \delta, \phi, \sigma) &= \min\{\psi(\zeta; \delta, \phi) : \zeta \in \mathcal{Z}(W; T, \sigma)\} \\
\Psi_{harm}(W; \psi, \delta, \phi, \sigma) &= \mu_h\{\psi(\zeta; \delta, \phi) : \zeta \in \mathcal{Z}(W; T, \sigma)\} \\
\Psi_{geom}(W; \psi, \delta, \phi, \sigma) &= \mu_g\{\psi(\zeta; \delta, \phi) : \zeta \in \mathcal{Z}(W; T, \sigma)\}
\end{aligned}$$

## 13. Key Word Sets

Any description of an event must have one or more key word sets. For the statistics to be meaningful, the key word sets must be specified *a priori*, before any kind of experiment is done and without peeking at the data. The key

word sets must be correct,<sup>1</sup> in some sense complete, and use Hebrew spelling that is determined by rule rather than by personal preference.<sup>2</sup>

## 14. Monkey Text Population

In order to evaluate whether an effect is occurring in the Torah text different from what would happen by chance in an ordinary text, it is required that a population of ordinary texts be defined. We call such a population of ordinary texts *Monkey Texts* to emphasize that in the Monkey Text population, the Null hypothesis of *No Torah Code Effect* is applicable. For the Torah text to be special with regard to Torah codes, it must mean that the strength of the effect is significantly higher (the ELSs of logically/historically related key words are in a more compact arrangement) in the Torah text than in the texts of the Monkey text population. The way to determine significantly higher is to compare. For any given key word set, there is a compactness value  $v_1$  that is computed for the Torah text. Then  $N - 1$  texts from the Monkey text population can be randomly sampled. Associated with the randomly sampled text  $n$  is a compactness value  $v_n$  for the same key word set. The significance  $s$  of the effect in the Torah text is measured by computing the number of the  $N$  total texts having smaller compactness value plus one half the number of the total texts having equal compactness value, normalized by  $N$  the total number of texts examined.

$$s = \frac{\#\{n \mid v_n < v_1\} + .5\#\{n \mid v_n = v_1\}}{N}$$

Here  $s$  is the normalized rank of the Torah text's compactness value.

There are a variety of different kinds of Monkey text populations that can be defined that in some significant way bear some statistical similarity to the Torah text. Each is created by taking the Torah text or its ELSs and performing some kind of randomly shuffling, making whatever compactness relationships that occur in these texts due to pure chance.

Random Shuffling techniques include letter shuffling, word shuffling, verse shuffling, and chapter shuffling. Letter and word shuffling can be done within verse, within chapter, within book, or globally. Verse shuffling can be done within chapter, within book, or globally. Chapter shuffling can be done within book or globally. In the Constant Word Shuffle, the letters of each of the distinct words of the text is randomly shuffled. Each time a word appears in the text the same letter shuffling of that word appears in the shuffled text. All these shuffling techniques guarantee that the

<sup>1</sup>Is has been documented from the Aumann committee experiment where experts were employed that many dozens of errors were made. Errors of course invalidate an experiment.

<sup>2</sup>Hebrew has alternative ways of spelling due to the degree to which additional letters are used in the spelling to designate vowels.

letter frequency of the text is preserved in all the monkey texts. However, the ELS statistics will change. To keep the ELS statistics the same we favor the ELS random placement monkey text population discussed next.

## ELS Random Placement Text Population

The ELS random placement text population is always with respect to a given text and its set  $\mathcal{E}$  of ELSs of the given set of key words and resonance specification. It does not consist of a text as a long string of letters. Rather, each text of the population is represented as a set of ELSs where each ELS keeps the same skip, length, and characters as it had in the original ELS set  $\mathcal{E}$ . However the beginning (and therefore the ending) positions of each ELS are randomly translated. Each translation that keeps the span of the ELS entirely within the text length has the same probability of occurring. This translation happens independently for each ELS. So if there were  $N$  ELSs and ELS  $n$  had  $X_n$  possible translations, then the number of texts in the ELS random placement text population would in effect be

$$\prod_{n=1}^N X_n$$

If  $\mathcal{E}(u, T; \sigma)$  is the set of ELSs of word  $u$  from text  $T$  according to skip specification  $\sigma$ , we write  $\mathcal{E}(u, T; \sigma, \pi)$  to designate an ELS random placement perturbation of  $\mathcal{E}(u, T; \sigma)$  according to perturbation  $\pi$ .

## 15. Summary

We have discussed some of the basic concept involved in testing the Torah code hypothesis. In part two of this paper we will illustrate the application of these concepts in experiments and discuss the results.

## References

- [1] R. Haralick, E. Rips, and M. Glazerson. *Torah Codes: A Glimpse Into the Infinite*. Marcel Printing, 2005.
- [2] I. Katsuchi, A. Nakazawa, K. Hasegawa, and T. Ohishi. The great buddha project: Modeling cultural heritage for vr systems through observation. *IEEE ISMAR03*, 2003.
- [3] B. McKay, D. Bar-Natan, M. Bar-Hillel, and G. Kalai. Solving the bible codes puzzle. *Statistical Science*, pages 149–173, May 1999.
- [4] D. Witztum, E. Rips, and Y. Rosenberg. Equidistant letter sequences in the book of Genesis. *Statistical Science*, 9(3):429–438, August 1994.