# Torah Codes: The American Presidents

Robert M. Haralick
Computer Science Department
Graduate Center
City University of New York
365 Fifth Avenue
New York, NY 10016
haralick@netscape.net

February 18, 2007

**Abstract**

In this paper we discuss the strategies and methodologies by which a proper test of the Null hypothesis of no Torah effect can be done against a complex of alternative hypotheses: that almost all the events in one or two of the their key word sets have ELSs that are in a more compact relation than expected by chance. We apply this test to the list of American presidents.

The American president experiments is driven by a ruled based spelling transliteration scheme of English names into multiple Hebrew alternative spellings. This methodology is more immune to claims of non *a priori* data selection because it is rule based and because the rules generate many spelling alternatives all of which are used in the experiment. The rule based scheme itself becomes incorporated into the alternative hypothesis.

With our new methodology, the Null Hypothesis is tested against the alternative that nearly all events have one or two key word sets that have ELSs in a more compact meeting than expected by chance. The Null Hypothesis had to be rejected. The p-level was below 1.5/100,000.

# 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[5]. 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 at least three documentary video productions that have been aired over cable multiple times. The Torah codes have been involved in a great academic controversy ever since the publication of the first formal study by Witztum, Rips, and Rosenberg[7] (WRR) and the subsequent claim that any apparently successful experiment must be due to non *a priori* data selection[6].

The Torah codes center on the Hebrew Torah text, the five books of Moses. It has been claimed that key words which are historically/logically related have their low skip rank equidistant letter sequences in a more compact geometric arrangment in the Torah text than expected by chance. This phenomena, if it is really there, is surprising. Here, we do not review the prior experiments that have been done. Nor do we discuss the controversy itself. That discussion can be found in Haralick, Rips, Glazerson[4]. 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 in section 16 discuss experiments that have been done with a new and more robust feature set.

In order to have an experiment that is reproducible, there has to be an experimental protocol which describes in sufficiently precise detail all the steps and calculations so that another researcher can independently perform the experiment and expect to get results that are insignificantly different from that of the original experiment. It is this kind of replication that the scientific methodology demands. In this paper we provide exact descriptions of experimental protocols that was used to test the Torah Code hypothesis with respect to the American Presidents.

The paper discusses the basic concepts of equidistant letter sequence, skip specification, resonance specification, compactness features, and experimental protocol issues in testing the Torah Code Hypothesis. Principal concepts of the experimental protocol involve the control population, here called the monkey text population, the skip specification, the resonance specification, and the test statistic which is used to do the actual hypothesis testing of the Null hypothesis against a complex of alternative hypotheses.

Our experiments each involve multiple events and each event has multiple key word sets. The formula for the test statistic is motivated by a probability derivation given in appendices A and the associated statement of the complex alternative hypothesis: that nearly all the events have one or two key word sets whose words have ELSs in a more compact arrangement than expected by chance. The experimental protocol uses the test statistic as a score in Monte Carlo experiment. The p-value of the experiment is the normalized rank of the Torah text test statistic in the sampled monkey text test statistics.

We begin our discussion with the basic definitions and concepts[2].

# 2 Definitions

A *word* $w$ of length $K$ is a sequence of $K$ characters $w =< w_1, \ldots, w_K >$. 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 $T =< t_1, \ldots, t_Z >$ 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 $< W(e)_1, \ldots, W(e)_{L(e)} >$ 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

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

Character $W(e)_i$ of ELS $E$ is associated with position $B(e) + (i-1)|S(e)|, i = 1, \ldots, L(e)$. The *span* of an ELS $e$ is given by

$$E(e) - B(e) + 1 = 1 + (L(e) - 1)|S(e)|$$

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 $w_1, \ldots, w_{L_w}$ if and only if $L_w = L(e)$ and $w_i = W(e)_i,\ i = 1, \ldots, L_w$. ELS $e$ is said to be a *negative skip ELS* of a word $w$ whose respective characters are $< w_1, \ldots, w_{L_w} >$ if and only if $L_w = L(e)$ and $w_i = W(e)_{L(e)+1-i}, i = 1, \ldots, L_w$.

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

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

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

$$\mathcal{E}(w, T) = \{e \quad | \quad e \text{ is an ELS of word } w \text{ in text} T\}$$

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 $w =< \alpha_1, \ldots, \alpha_K >$, will match the letters in the placement position is given by

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

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
$w =< \alpha_1, \ldots, \alpha_K >$ is given by $pN$ where

$$
\begin{aligned}
p &= \prod_{k=1}^{K} p(\alpha_k) \\
N &= \frac{(S_{max} - S_{min} + 1)}{2} * (2Z - (L-1) * (S_{max} + S_{min}))
\end{aligned}
$$

since

$$\sum_{k=0}^{N} k Prob(k \mid p, N) = pN$$

## 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 probabiilty

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

# 4 Skip Specification

The most convincing Torah codes are often found using the expected number of ELS search criterion. The smallest skip $S_{min}$ 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 the given expected number$M$ when the smallest absolute skip is 2. This is the protocol followed by WRR. WRR sets the expected number to be 10 for the *Genesis* text of 78,064 letters.[1] We set the expected number to be 10 for the five books of the Genesis through Deuteronomy.

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 $\sigma$.

# 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 induce on the cylinder a *row skip* $s_r$ and *column skip* $s_c$ given by

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

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

where $\lceil \ \rceil$ designates the ceiling function.

We say that the column skip of an ELS on a cylinder is positive if $s$ mod $\gamma \leq \gamma - s$ mod $\gamma$. 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 $s$ mod $\gamma > \gamma - s$ mod $\gamma$. 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 involves ELSs and cylinder sizes where the induced row skip $s_r$ and column skip $s_c$ are both sufficiently small. A skip of size $s$ and a cylinder of size $\gamma$ are said to *resonate* when the induced $s_r$ and $s_c$ are sufficiently small.

The American President experiment uses the max row column skip criterion. Let $s = s_r\gamma + s_c$. Then skip $s$ resonates with cylinder size $\gamma$ when

[1] Even though by convention of WRR, the expected number is computed for a minimum skip of 2, the skip specification is free to choose $S_{min} = 1$ or $S_{min} = 2$. If $S_{min} = 1$, the maximum skip is computed by expected number with $S_{min}$ = 2.

$s_r \leq s_{rmax}$ and $s_c \leq s_{cmax}$

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

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

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 lengths $L(e_1)$ 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 $d_{12}$.

$$\begin{aligned} d_1(e_1,e_2;\gamma) &= \min_{\substack{i=1,\ldots,L(e_1)\\ j=1,\ldots,L(e_2)}} \Delta(B(e_1)+(i-1)|S(e_1)|, B(e_2)+(j-1)|S(e_2)|;\gamma) \\ d_2(e_1,e_2;\gamma) &= \max_{\substack{i=1,\ldots,L(e_1)\\ j=1,\ldots,L(e_2)}} \Delta(B(e_1)+(i-1)|S(e_1)|, B(e_2)+(j-1)|S(e_2)|;\gamma) \\ d_{12}(e_1,e_2;\gamma) &= \sqrt{d_1^2(e_1,e_2;\gamma)+d_2^2(e_1,e_2;\gamma)} \end{aligned}$$

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.

$$\begin{aligned} \omega_1(e_1,e_2,\gamma) &= d_1^2(e_1,e_2;\gamma)+\Delta^2(0,S(e_1);\gamma)+\Delta^2(0,S(e_2);\gamma) \\ \omega_2(e_1,e_2,\gamma) &= d_2^2(e_1,e_2;\gamma)+\Delta^2(0,S(e_1);\gamma)+\Delta^2(0,S(e_2);\gamma) \\ \omega_{12}(e_1,e_2,\gamma) &= d_{12}^2(e_1,e_2;\gamma)+\Delta^2(0,S(e_1);\gamma)+\Delta^2(0,S(e_2);\gamma) \end{aligned}$$

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 between $d$ and $\Delta$. 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 $\epsilon$.

$$\begin{aligned} \rho_1(e_1,e_2,\gamma) &= (max\{d_1^2(e_1,e_2;\gamma),\epsilon)\})\times(\Delta^2(0,S(e_1);\gamma)+\Delta^2(0,S(e_2);\gamma)) \\ \rho_2(e_1,e_2,\gamma) &= d_2^2(e_1,e_2;\gamma)\times(\Delta^2(0,S(e_1);\gamma)+\Delta^2(0,S(e_2);\gamma)) \\ \rho_{12}(e_1,e_2,\gamma) &= d_{12}^2(e_1,e_2;\gamma)\times(\Delta^2(0,S(e_1);\gamma)+\Delta^2(0,S(e_2);\gamma)) \end{aligned}$$

## 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 $\gamma$.

$$\begin{aligned}
D_1(E;\gamma) &= \max_{\substack{e\in E\\ f\in E}} d_1(e,f;\gamma)\\
D_2(E;\gamma) &= \max_{\substack{e\in E\\ f\in E}} d_2(e,f;\gamma)\\
D_{12}(E;\gamma) &= \max_{\substack{e\in E\\ f\in E}} d_{12}(e,f;\gamma)\\
\Omega_1(E;\gamma) &= \max_{\substack{e\in E\\ f\in E}} \omega_1(e,f;\gamma)\\
\Omega_2(E;\gamma) &= \max_{\substack{e\in E\\ f\in E}} \omega_2(e,f;\gamma)\\
\Omega_{12}(E;\gamma) &= \max_{\substack{e\in E\\ f\in E}} \omega_{12}(e,f;\gamma)\\
R_1(E;\gamma) &= \max_{\substack{e\in E\\ f\in E}} \rho_1(e,f;\gamma)\\
R_2(E;\gamma) &= \max_{\substack{e\in E\\ f\in E}} \rho_2(e,f;\gamma)\\
R_{12}(E;\gamma) &= \max_{\substack{e\in E\\ f\in E}} \rho_{12}(e,f;\gamma)
\end{aligned}$$

# 9 Combination Methods Over Resonant Cylinder Sizes

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

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

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

Let $W = \{w_1, \ldots, w_K\}$ 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 $\mathcal{E}(w, T, \sigma),\ w \in W$,

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

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

## 11 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, $\mu_g$ and simply summing.

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

where $\upsilon$ is a weight function whose value is the fraction of the text that the ELSs of $\zeta$ have minimal absolute skip.

## 12 Event Key Word Sets

For the $p - value$ 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. A significant part of the Torah code controversy is about the *a priori* specification of the key word sets, one side claiming that successful experiments have been done with demonstrably *a priori* key word sets and the other side claiming not. To obtain key word sets that are demonstrably *a priori*, experts and previously published lists of key words have been employed. To make the *a priori* demonstrably strong Hebrew spellings should be used that are determined in some objective way such as by rule rather than by subjective preference.[2] However, *a priori* is not the only condition that is required. The key word sets must be historically correct,[3] and in some sense complete. Complete means, for example, if the class of events is American Presidents, then there must be a collection of key word sets for each American president. No American president should be missing.

The specification of a key word set may seem simple but it is not. Let us suppose for the moment that an event has been encoded. The encoding involves a set of key words. But the experimenter does not know the key words. So the experimenter guesses a key word set. If the experimenter guesses wrong, then even though the event is encoded, the p-value of the experiment may not be not significant. Thus it is not unusual for the experimenter to make some small number of guesses, each guess being one key word set.

In this case, the experiment must be set up as a test of the Null hypothesis against multiple alternatives hypotheses. Each alternative hypothesis is specified by some key word subset of the given total set of key words. For a single event, the formal test of the Null hypothesis is against the alternative hypothesis that one or perhaps more than one of the key word sets have a more compact meeting than expected by chance. For multiple events, the formal test of the Null hypothesis can be against the alternative hypothesis that nearly all the events have one or more of their key word sets having a more compact meeting than expected by chance. Or as in WRR the alternative hypothesis is that the key word sets, taken over all events and over all alternative key word sets of the same event, tend to have more compact meetings than expected by chance.

---

[2]Hebrew has alternative ways of spelling due to the degree to which additional letters are used in the spelling to designate vowels.

[3]It has been documented from the Aumann committee experiment[1] where experts were employed that many dozens of errors were made. Errors of course invalidate an experiment.

# 13 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 an encoding 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 test statistic called the compactness score $v_1$ 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 score $v_n$ 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 score, 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 score.

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 might occur in these texts due to pure chance. In the American Presidents experiment, we use the ELS random placement text population.

## 13.1 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 skip specification. A text of the ELS random placement text population 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 $\pi$.

All the other letter or word shuffling schemes produce Monkey texts in which the number of ELSs for each key word will differ from that in the Torah text. The ELS random placement text population is the only random perturbation scheme that produces exactly the same number of ELSs and at exactly the same skips as produced by the Torah text. This is important.

Suppose that for a given key word set, the Torah text has some statistical advantage over texts in say a letter permuted text population because the Torah text has substantially more low rank skip ELSs of some of the key words than expected by chance. In this case an experiment might succeed mainly due to such an ELS distribution in the Torah text, rather than because of the relationship between ELSs of the key words. The ELS random placement text population can be said to be a conservative one because in this case, each ELS random placement text has the identical statistical advantage as the Torah text, and therefore, no text has any advantage.

# 14 Hypothesis Testing

The formal way in which the significance of an encoding is evaluated is by a test of Hypotheses. The Null hypothesis of No Torah Code Effect is tested against an alternative hypothesis that there is an encoding. The alternative hypothesis may be a complex of alternatives[3].

The statistical computation involved in the test of hypotheses amounts to defining a test statistic measuring the strength of the encoding and determining the fraction of monkey texts that have at least as good an encoding as the Torah text. The normalized rank of the test statistic is called the p-value of the experiment.

To do the test of hypothesis, the p-value of the experiment is compared to a significance level $\alpha_0$. If the p-value is smaller than $\alpha_0$, then the Null hypothesis of No Torah Code effect is rejected in favor of the alternative that the key word set has ELSs in an unusually compact arrangement. If the p-value is larger than the significance level, the Null hypothesis is not rejected. It is usual for the p-value of an experiment to be reported and the rejection or non-rejection of the Null hypothesis to be done by the reader based on his/her selected significance level.

## 14.1 Test of Null Hypothesis Against A Complex Alternative Hypothesis

An experiment about a particular historical event is described by a set of what are considered to be the key words relevant to that event. However, not all of the key words thought about might have corresponding ELSs in a relatively compact arrangement. Hence an hypothesis test of the Null hypothesis against the Alternative hypothesis that all of the key words have ELSs that are in a relatively compact arrangement will most likely not be rejected, even when there is an encoding of most of the key words in the set.

Therefore, the experiment must be set up as a test of the Null hypothesis against multiple alternatives hypotheses. Each alternative hypothesis is specified by some subset of the given total set of key words. For a single event, our formal test of the Null hypothesis is against the alternative hypothesis that one or two of the alternative hypotheses is true.

For example, in the American president experiment, an event is an American president's name paired with the Hebrew key word נשאי, meaning president. However, the spelling of the president's name in Hebrew is not known. There are four reasonable spellings for president Lincoln in Hebrew: לינקולן, לינקלן, לנקולן, לנקלן, depending on which of two vowels are explicitly represented in the spelling. Thus for the basic Lincoln event there are four alternative hypothesis represented by the four key word sets: {נשאי,לינקולן}, {נשאי,לינקלן}, {נשאי,לנקולן}, {נשאי,לנקלן}.

## 14.2 Bonferroni

When K separate experiments are done, each testing the Null hypothesis against a different Alternative hypothesis, yielding p-values $p_1, \ldots, p_K$, the smallest p-value is not the p-value of the complex of the K separate experiments. Indeed, if the experiments are separate, then the exact p-value of the complex of K separate experiments cannot be determined if they are not in fact independent. This is the usual case. However, it can be bounded. The Bonferroni upper bound is $K \min\{p_1, \ldots, p_K\}$. The p-value of the K separate experiments must be smaller than the Bonferroni bound. Therefore, if the Bonferroni bound, which is necessarily higher than the p-value of the complex of experiments, is smaller than the significance level, then it necessarily follows that the p-value of the complex of experiments is also smaller than the significance level. In this case the Null hypothesis can be rejected at the given signficance level.

The problem with the Bonferonni bound is that it is an upper bound and in many instances is much higher than the true p-value of the complex of K separate experiments. This is particularly true when the K Alternative hypothesis are statistically dependent. Using Bonferroni in this situation will make it more likely that the Null hypothesis will not be rejected when it ought to be rejected; that is a true effect will be misdetected. This leads us to examine more statistically efficient ways of doing the hypothesis testing.

## 14.3 Combining Over Key Word Sets

Suppose there are $K$ key word sets, each describing the same historical event. In this case it is expected that every pair of key word sets will have a substantial fraction of its key words in common to both sets. Hence the alternative hypotheses will necessarily have statistical dependence and the Bonferroni bound will be much too high.

There is a statistically economical alternative to using the Bonferroni bound when testing the Null hypothesis against a complex of K Alternative hypotheses in a combined experiment where the trial by trial results are available. The alternative is to use K scoring schemes, one appropriate for each of the K Alternative hypotheses, and then combine the scores together in a suitable way. For the sake of simplicity, the discussion which follows assumes that all scores are with respect to one given compactness measure.

After the first trial, which involves the Torah text, each remaining trial of an $N$ trial experiment randomly samples a monkey text from the monkey text population. In accordance with a specified protocol, on trial $n$, the compactness feature of the ELSs from each of the $K$ key word sets is computed, resulting in $c_{1n}, \ldots, c_{Kn}$. For the $k^{th}$ key word set, the $N$ compactness values $c_{k1}, \ldots, c_{kN}$ are rank normalized to $r_{k1}, \ldots, r_{kN}$.

The p-value associated with test of the Null hypothesis against the Alternative that the $k^{th}$ key word set has its ELSs in a more compact relationship than expected by chance is given by $r_{k1}$. The Bonferroni bound $B$ on the test of the Null hypothesis against the K alternative hypotheses is then $B = K \min\{r_{11}, \ldots, r_{K1}\}$.

The $K$ scores and combine method would define a combining function $F$ acting on the rank normalized values $r_{1n}, \ldots, r_{Kn}$, for the $n^{th}$ trial of the experiment. In this situation, combining functions ought to be symmetric in its arguments. For example, one combining function could be the minimum: $f_n = F(r_{1n}, \ldots, r_{Kn}) = \min\{r_{1n}, \ldots, r_{Kn}\}$. The scores $f_1, \ldots, f_N$ are rank normalized and the rank normalized value, $p_1$, associated with $f_1$ is the p-value of the experiment. For the min combining function, $p_1$ is necessarily smaller than the Bonferonni bound $B$.

However, the min combining function is not necessarily the statistically most optimal. For example, a combining function may be motivated by a probability derivation that has even some unwarranted conditional independence assumptions.[4] One such combining function derived in appendix A is

$$\begin{aligned} F_1(r_{1n}, \ldots, r_{Kn}; \theta) &= \frac{1}{K} \sum_{k=1}^{K} p(r_{kn}; \theta) \\ &= f_{1n} \end{aligned} \tag{1}$$

where $p(r; \theta)$ is the probability under the Alternative hypothesis of observing a normalized relative rank of $r$ in an N trial experiment and $1/K$ is the prior probability of any one of the $K$ alternative hypotheses of being true. We base $p(r; \theta)$ on $-\log$ because for small relative ranks $-\log$ will be large $(\log(2N))$ for an $N$ trial experiment where the smallest and unique relative rank is $1/2N$. For larger relative ranks, $-\log$ will be small and indeed be $0$ for a relative rank of $1$. We define

$$p(r; \theta) = \begin{cases} -\beta(\theta) \log(r) & \text{when } r < \theta \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

This combining function arises (up to a fixed constant of proportionality) when exactly one of the $K$ alternative hypothesis is assumed to be true, and for each trial $n$, and for each $k$, the random variables $r_{1n}, \ldots, r_{Kn}$ are conditionally independent given that Alternative Hypothesis $k$ is true. When the Null hypothesis

[4]The unwarranted assumptions are not used in making any probability calculations for the p-value. The probability derived by using the unwarranted assumptions gives a motivation and a formula for performing a calculation of a score function. It is the score function that is used in a proper Monte Carlo experiment for determining the p-values.

holds for the remaining $K-1$ possibilities, they are assumed to follow the discrete uniform on the normalized relative ranks of an $N$ trial experiment. The threshold $\theta$ specifies that the probability of observing a normalized rank greater than $\theta$ under the alternative hypothesis is $0$. We call this method the *first order combining method*.

If two of the $K$ alternative hypotheses are assumed to be true, with the prior probability for any pair to be true to be $2/(K(K-1))$, then under the same conditions as the previous derivation, the combining function should be

$$F_2(r_{1n}, \ldots, r_{Kn}; \theta) = \frac{2}{K(K-1)} \sum_{\{(j,k)|k>j\}} p(r_{jn}; \theta) p(r_{kn}; \theta)$$
$$= f_{2n} \tag{3}$$

We call this method the *second order combining method.*

If it is assumed that when there is an encoding either one or two of the $K$ alternative hypotheses is true, and the prior probability for exactly one alternative being encoded is $q$ and the prior probabiity for exactly two alternatives being encoded is $1-q$ then under the same conditions as the first probability derivation, the score $s_n$ of the $n^{th}$ trial should be

$$s_n = q f_{1n}/N + (1-q) f_{2n} \tag{4}$$

We call this method the *non-rank normalized* method of combining.

Another possible way of combining $F_1$ with $F_2$ would be to take the $N$ values $f_{11}, \ldots, f_{1N}$ and rank normalize them forming the $N$ normalized ranks $t_{11}, \ldots, t_{1N}$. Also rank normalize the $N$ values $f_{21}, \ldots, f_{2N}$ forming the $N$ normalized ranks $t_{21}, \ldots, t_{2N}$. Define the rank normalized score $s_n$ for the $n^{th}$ trial by the convex combination

$$s_n = q t_{1n} + (1-q) t_{2n} \tag{5}$$

In either combining method, the p-value of the experiment is the normalized relative rank of the score for the first trial.

## 14.4 Composite Experiments

Composite experiments are associated with multiple events. Suppose that there are $M$ events. Each event has associated with it a collection of key word sets. The $m^{th}$ such collection is associated with a test of the Null hypothesis against the $K_m$ alternative hypotheses formed by each one of the $K_m$ key word sets in the collection. In the composite experiment, we are interested in a test of hypotheses at two levels. We wish to test the Null hypothesis against the Alternative that more of the $M$ events have their ELSs in a more compact arrangement than expected by chance.

Hence, we treat each event as an experiment that produces in each trial a score which is the normalized relative rank of the compactness associated with the trial. Thus each trial produces $M$ scores. These scores then combined together in a test statistic appropriate for a test of the Null hypothesis against $M$ Alternative hypotheses, where all or nearly all of the $M$ Alternative hypotheses are assumed to be true. Here we take "all or nearly all" to mean all, or all but one, or all but two, or all but three. Appendix A.3 gives the derivation for the computation of the associated test statistic.

# 15 Our Compactness Features

For any key word set $W$, our compactness measure is composed of three basic compactness components. In the definitions that follow recall that $\sigma$ is the skip specification, $\phi$ is the resonance specification.

- $h_1 = \Psi_{min}(W; \psi_{min}, D_2, \phi_\infty, \sigma)$
  $\phi_\infty$ is that resonance specification that specifies that the only cylinder size any skip resonates to is the cylinder of infinite number of columns and on that cylinder any skip resonates. The cylinder of infinite number of columns essentially treats the text as a long linear string of symbols. On that cylinder the distance measure $D_2$ is the largest span length of any pair of ELSs in the ELS set. $\psi_{min}$ indicates that over all resonant cylinder sizes for an ELS set, we take the cylinder size having the smallest distance. As there is only one cylinder size under consideration here, $\psi_{min}$ is the identity function. $\Psi_{min}$ indicates that over all ELS sets of the key word set, we take the set associated with the smallest span length.

- $h_2 = \Psi_{harm}(W; \psi_{min}, \Omega_2, \phi, \sigma)$
  $\Omega_2$ is the largest distance between a pair of ELSs in an ELS set where the

distance itself is defined as the square of the largest Euclidean distance between the letters of the ELSs on the cylinder plus the squared Euclidean distance between successive letters of the first ELS on the cylinder plus the squared Euclidean distance between successive letters of the second ELS on the cylinder. $\psi_{min}$ indicates that of the different distances, each associated with a different resonant cylinder size, we take that distance corresponding to the best cylinder size. The best cylinder size is the one for which this distance is smallest. $\Psi_{harm}$ indicates that over all the ELS sets of the given key word set, each ELS set having a best cylinder distance, we take the harmonic mean.

- $h_3 = \Psi_{harm}(W; \psi_{min}, R_{12}, \phi, \sigma)$
  $R_{12}$ is the largest distance between a pair of ELSs in an ELS set where the distance itself is defined as the square of the largest Euclidean distance between the letters of the ELSs on the cylinder plus the smallest Euclidean distance between the letters of the ELSs on the cylinder all times the sum of the the squared Euclidean distance between successive letters of the first ELS on the cylinder and the squared Euclidean distance between successive letters of the second ELS on the cylinder. $\psi_{min}$ indicates that of the different distances, each associated with a different resonant cylinder size, we take that distance corresponding to the best cylinder size. The best cylinder size is the one for which this distance is smallest. $\Psi_{harm}$ indicates that over all the ELS sets of the given key word set, each ELS set having a best cylinder distance, we take the harmonic mean.

The key word set compactness measure $h_1$ is the 1D compactness measure. It's value is the length of the smallest length text segment that contains at least one ELS of each of the key words. This measure is the first measure we experimented with over ten years ago. Although there are substantial number of key word sets describing events that this measure detects with associated small p-values, there are many key word sets believed to be encoded that this measure does not detect.

Examining cases where it works and those where it does not work, we realized that when the compactness of a short word (say 4 characters or less) is measured with a longer word (say 6 characters or more) what happens is that the skips of the ELSs of the longer word will tend to be much larger than the skips of the ELSs of the shorter word. Hence the span of the ELSs of the longer word will tend to be much larger than the span of the ELSs of the shorter word. That increases the probability that regardless where the random placement puts the ELSs of the longer word, there will most likely be some random placement of the shorter word ELS that is nearly contained within the span of the longer word ELS. This translates to a p-value for the 1D compactness measure alone that will be around 0.5, making it essentially useless as a measure under these conditions.

The key word set compactness measure $h_2$ is a variation motivated by the Omega measure of WRR. Here the distance used is the sum of the closest and furthest distance between pairs of letters of the two ELSs. Instead of adding to that squared distance the squared distance between sucessive letters of each ELS on the cylinder as done in the Omega measure and $h_3$, we multiply the sum of the closest and furthest squared distance between pairs of letters of the two ELSs with the sum of the squared distance between successive letters of each ELS on the cylinder. Both the $h_2$ measure and the $h_3$ compactness measure were reported on orally at the International Torah Code conference in Jerusalem in 2000 and 2001.

The key word set compactness measure $h_3$ is very much in the spirit of the Omega measure of WRR[7]. However, in the Omega measure of WRR, the distance used is the closest distance between pairs of letters of the two ELSs. In $h_3$, the distance used is the maximum distance between pairs of letters of the two ELSs.

# 16 Our Experimental Protocol

For our skip specification $\sigma$, we set the largest skip permitted for ELSs of a given key word to be such that the expected number of ELSs searching from a minimum skip of 2 would be 10. And we set the minimum skip for ELSs to be 1.

For our resonance specification $\phi$, we require that at least one ELS from each key word in a key word set be resonant on a cylinder size and on the resonant cylinder size the skip of the ELS must be no more than 10 rows and no more than 10 columns. This differs from WRR who only insisted that the row skip on the cylinder be no more than 10 rows.

For our monkey text population we use the ELS random placement population with 100,000 trials. The Monte Carlo is carried out with an independent execution for each event. The random number seed was obtained from the digits of $\pi$. Starting from the first digit after the decimal point, the digits were broken up into strings of seven long. Each successive string of seven $\pi$ digits was used as the random number seed for each successive event Monte Carlo.

We test the Null hypothesis against the complex of alternatives that nearly all the events are encoded and that for each encoded event, one or two key word sets are encoded by either compactness measure $h_3$ or by both compactness measures $h_1$ and $h_2$.

We use both first order (1) and second order (3) methods of combining over the key word sets of each event. We set our threshold $\theta = .2$ in (2), a value used by WRR[7] in a slightly different context, but in the same spirit as we used it.

We use the nearly all combining method to combine over all the events as described in Appendix A.3. We do this for each trial. The prior probability for each of the four cases: all, all but one, all but two and all but three are identical and equal to .25. We define the probabilty $P_2$ by

$$P_2(r) = \frac{e^r}{1 - r}$$

# 17 The American Presidents

In this section we report on an experiment pairing the names of the American presidents with the key word נשיא, meaning *president*. There are 42 people who have served as presidents, some multiple times. To perform the experiment, we must transliterate their names into Hebrew. Due to the various ways non-Hebraic names can be spelled in Hebrew, we devised a rule base system to provide a reasonable set of Hebrew spellings for each president's name (see Appendix B. In addition we use two variations: the last name alone and the first character of the president's first name as a prefix to the spelling of the last name. Presidents with a middle initial we also provided an additional alternative with the first and middle initial as a prefix to the last name (see Appendix C. The total number of spellings was 314, on the average nearly seven and a half spellings per name. The number of spellings having at least one ELS was 261.

The spellings generated by the rules include nearly all the spellings used in Hebrew encyclopedic sources. Those spellings not generated by the rules are spellings that are closer to Yiddish spellings rather than Hebrew spellings.

The rules are simple, without exceptions, and complete. The rules map the English consonants to the Hebrew consonants based on the phonetic sound. They map the long and short English vowels to the Hebrew letters א,וי, and הin a manner consistent with the way Hebrew represents vowels by letters in the system call *mater lectionis* or simply the full spellings. For each possible vowel that can be represented in the system of *mater lectionis*, the rules produce a spelling with and without the *mater lectionis* letter.

There is an issue of what to do with double consonants like the double **r** in Harrison, whose syllables parse as *har-ri-son*. The typical Hebrew spelling will map the double **r** to a single ר. However, since we do not have the pre-knowledge that an encoding must do it that way, we also allowed for a double English consonant to map to a double Hebrew consonant. Although it is rare for Hebrew names or words to have a double consonant with a single phonetic sound it does occur. An example name is the Israelite tribe Issachar which in Hebrew is spelled יששכר, with a double שׂ. Two example words with double ר are סָרַרְנוּ, meaning *we have been disobedient* and צָרַרְנוּ, meaning *we have been hostile*.

The American president experiment tests the Null hypothesis of *No Torah Code Effect* against the complex alternative hypothesis that

1. in accordance with the Hebrew to English translation rules of appendix B
2. and in accordance with the skip specification, and the resonance specification stated in section 16
3. for nearly all the presidents
4. each president has one or two Hebrew spellings of his name
5. that have ELSs which are in a more compact arrangement with ELSs of the Hebrew word נשאי, meaning president
6. in the 5 books of the Chumash
7. by compactness measure $h_1, h_2$ or $h_3$

The compactness measure $h_3$ produced the smallest p-value. In a 100,000 trial Monte Carlo experiment taking nearly three weeks on a dedicated PC, the $h_3$ p-value $.5/100,000$. Of the $42$ presidents, $36$ of them had some spelling paired with נשיא so that on one of the compactness measures the rank normalized value was less than $.10$. $73$ of the $261$ key word pairs has a rank normalized value on the $h_3$ measure of less than $.10$. The probability that under the Null hypothesis $73$ or more out of $261$ key word sets would have a rank normalized value less than $.10$ is $2.88 \times 10^{-16}$. From this we conclude that main result is not due to either a few presidents or a few key word sets. The result is something characteristic to the entire set of presidents and their key word sets.[5]

[5] By the entire set of presidents and their key word sets we do not mean each and every president and each and every key word set. Only way more presidents and way more key word sets than expected by chance.

# 18 Concluding Discussion

We have discussed an experimental protocol by which an experiment can be done that tests the Null hypothesis of No Torah Code Effect against a composite alternative. We used the protocol on the American Presidents. The composite alternative hypothesis is that more of the events have one or two key word sets that have their ELSs in a more compact arrangement by either compactness measure $h_3$ or simultaneously by compactness measure $h_1$ and $h_2$ than expected by chance.

For this purpose we developed a score function based on a probability derivation of what the probability would be if one or if one or two of a fixed number of choices follows a given probability function while the remaining follow a discrete uniform probability function. And by a similar derivation what the probability would be if nearly all (all, all but one, all but two) follow a given given probability function while the remaining follow a discrete uniform probability function.

The protocol employing the score function is direct, statistically motivated, self normalizing, consistent with the nature of the alternative hypothesis, and (in our opinion) aesthetically simple. No part of the protocol has large numbers of variables or parameters whose values can be set to memorize the pattern of the ELS data from the Torah text versus that from the monkey texts. The parameters of the protocol itself were three: maximum skip set so that the expected number of ELSs was about 10; the maximum row and column skip of an ELS on a cylinder was 10. The probability threshold was .2. The rest of the freedom in the protocol came from methodological choices: the monkey text population, the compactness measures, the various rank normalizations, the combining method over key word sets of an event, the combining method over scores of events.

For various reasons, that we did not discuss due to space limitations, our methodology is more conservative than that employed by WRR[7] and less prone to the kind of wiggling done by McKay to make an experiment apparently successful. The best of the three experiments produced a p-value of .5/100,000. So by BonFerroni, the p-value of the combined experiment is bounded above by 1.5/100,000. It is clear that the Null hypothesis of No Torah Code Effect has to be rejected.

For our future work, we will be applying this protocol to new event data sets with a larger number of trials.

# Appendix

## A One Or Two Of

A sample $x_1, \ldots, x_K$ is taken. Each observed value either is sampled from population 1 or population 2. The probability of any observed value is the same constant in population 1; i.e. population 1 values are distributed as a discrete uniform. The probability of any observed value $v$ from population 2 is given by probability function $P_2$.

It is known a priori that either only one or two of the sampled values come from population 2. And the prior probability is equal to $2/(K(K+1))$ for any one of the $K(K+1)/2$ possibilities that either one or two observed values are sampled from population 2.

Let $\Phi$ denote the set of $K(K+1)$ mutually exclusive possibilities.

$$\Phi = \{\phi \subset \{1, \ldots, K\} \mid \#\phi \leq 2\}$$

$\Phi$ can be written as $\Phi_1 \cup \Phi_2$ where

$$\begin{aligned}
\Phi_1 &= \{\phi \subset \{1, \ldots, K\} \mid \#\phi = 1\} \\
\Phi_2 &= \{\phi \subset \{1, \ldots, K\} \mid \#\phi = 2\}
\end{aligned}$$

Once the possibility specified by a $\phi \in \Phi$ is given, the values are independently sampled.

Let $\Delta$ be the constant probability that a sampled value comes from population 1.

### A.1 One Or Two

Now we can compute the probability of the observed sample $x_1, \ldots, x_K$ given that the sampling comes from one of the possibilities of $\Phi$.

$$\begin{aligned}
P(x_1, \ldots, x_K \mid \Phi) &= \frac{P(x_1, \ldots, x_K, \Phi)}{P(\Phi)} \\
&= \sum_{\phi \in \Phi} \frac{P(x_1, \ldots, x_K, \phi)}{1}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\phi\in\Phi} P(x_1,\ldots,x_K \mid \phi)P(\phi) \\
&= \sum_{\phi\in\Phi} P(x_1,\ldots,x_K \mid \phi)\frac{2}{K(K+1)} \\
&= \sum_{\phi\in\Phi} \prod_{j\in\phi^c} \Delta \prod_{j\in\phi} P_2(x_j) \\
&= \frac{2}{K(K+1)}\left\{\sum_{\phi\in\Phi_1} \Delta^{K-1} \prod_{j\in\phi} P_2(x_j) + \sum_{\phi\in\Phi_2} \Delta^{K-2} \prod_{j\in\phi} P_2(x_j)\right\} \\
&= \frac{2\Delta^{K-2}}{K(K+1)}\left\{\sum_{\phi\in\Phi_1} \Delta \prod_{j\in\phi} P_2(x_j) + \sum_{\phi\in\Phi_2} \prod_{j\in\phi} P_2(x_j)\right\} \\
&= \frac{2\Delta^{K-2}}{K(K+1)}\left\{\sum_{k=1}^{K} \Delta P_2(x_k) + \sum_{j=1}^{K} \sum_{k=j+1}^{K} P_2(x_j)P_2(x_k)\right\}
\end{aligned}
$$

## A.2 One of

If it is known a priori that only one of the sampled values comes from population 2, then we can proceed in a similar manner.

$$
\begin{aligned}
P(x_1,\ldots,x_K \mid \Phi_1) &= \frac{P(x_1,\ldots,x_K,\Phi_1)}{P(\Phi_1)} \\
&= \sum_{\phi\in\Phi_1} \frac{P(x_1,\ldots,x_K,\phi)}{1} \\
&= \sum_{\phi\in\Phi_1} P(x_1,\ldots,x_K \mid \phi)P(\phi) \\
&= \sum_{\phi\in\Phi_1} P(x_1,\ldots,x_K \mid \phi)\frac{1}{K} \\
&= \sum_{\phi\in\Phi_1} \prod_{j\in\phi^c} \Delta \prod_{j\in\phi} P_2(x_j) \\
&= \frac{\Delta^{K-1}}{K} \sum_{k=1}^{K} P_2(x_k)
\end{aligned}
$$

When smaller values are more probable, and when $x$ is known to lie in a range that is greater than 0 and less than or equal to 1, then probability distributions that monotonically decrease with increasing value include the harmonic distribution $P_2(x) = \frac{\alpha}{x}$ and the geometric distribution $P_2(x) = -\beta log(x)$

### A.2.1 Testing the Null Hypothesis

Suppose that we are testing the Null Hypothesis that each $x_1 \ldots, x_k$ comes from the discrete uniform distribution against the Alternative Hypothesis that one of the observed values come from population 2 which is distributed by the harmonic distribution. Then the likelihood ratio of the Null hypothesis to the Alternative Hypothesis is given by

$$\begin{aligned} LR &= \frac{\Delta^K}{\frac{\Delta^{K-1}}{K}\sum_{k=1}^K \alpha/x_k} \\ &= \frac{\Delta}{\alpha}\frac{1}{\frac{1}{K}\sum_{k=1}^K \frac{1}{x_k}} \\ &= \frac{\Delta}{\alpha}\text{harmonic mean}(x_1, \ldots, x_K) \end{aligned}$$

This motivates the use of the harmonic mean of $x_1, \ldots, x_K$ as the test statistic. When the harmonic mean is sufficiently small, the Null hypothesis would be rejected.

Suppose that we are testing the Null Hypothesis that each $x_1 \ldots, x_k$ comes from the discrete uniform distribution against the Alternative Hypothesis that one of the observed values come from population 2 which is distributed by the geometric distribution. Then the likelihood ratio of the Null hypothesis to the Alternative Hypothesis is given by

$$\begin{aligned} LR &= \frac{\Delta^K}{\frac{\Delta^{K-1}}{K}\sum_{k=1}^K -\beta\log(x_k)} \\ &= -\frac{\Delta}{\beta}\frac{1}{\frac{1}{K}\sum_{k=1}^K \log(x_k)} \\ &= -\frac{\Delta}{\beta}\frac{1}{\frac{1}{K}\log(\prod_{k=1}^K x_k)} \\ &= -\frac{\Delta}{\beta}\frac{1}{\log\left((\prod_{k=1}^K x_k)^{\frac{1}{K}}\right)} \\ &= -\frac{\Delta}{\beta}\frac{1}{\log\left(\text{geometric mean}(x_1, \ldots, x_K)\right)} \end{aligned}$$

This motivates the use of the geometric mean as the test statistic. When the geometric mean of $x_1, \ldots, x_K$ is sufficiently small, the Null hypothesis would be rejected.

## A.3 All or Nearly All

A sample $x_1, \ldots, x_K$ is taken. Each observed value either is sampled from population 1 or population 2. The probability of any observed value is the same constant in population 1; i.e. population 1 values are distributed as a discrete uniform. The probability of any observed value $v$ from population 2 is given by probability function $P_2$.

It is known a priori that all, all but one, all but two or all but three of the sampled values come from population 2. And the prior probability for the all case be $q_0$, the prior probability for the all but one case be $q_1$, the prior probability for the all but two case be $q_2$, and the prior probability for the all but three case be $q_3$. For the all but one, the prior probability for each of the $K$ cases is $1/K$. For the all but two, the prior probability for each of the $K(K-1)/2$ cases is $2/(K(K-1)$. For the all but three, the prior probability for each of the $K(K-1)(K-2)/6$ cases is $6/(K(K-1)(K-2))$.

Let $Q_0$ be the probability of the observed values given the all case. Let $Q_1$ be the probability of the observed values given the all but one case. Let $Q_2$ be the probability of the observed values given the all but two case. Let $Q_3$ be the probability of the observed values given the all but three case. Let there be $N$ trials. Assuming conditional independence, we can write,

$$\begin{aligned}
Q_0 &= \prod_{k=1}^{K} P_2(x_k) \\
Q_1 &= \frac{1}{N}\frac{1}{K}\sum_{k=1}^{K}\prod_{\substack{i=1\\ i\neq k}}^{K} P_2(x_k) \\
&= \frac{1}{N}\frac{1}{K}\prod_{k=1}^{K} P_2(x_k)\sum_{i=1}^{K}\frac{1}{P_2(x_i)} \\
Q_2 &= \frac{1}{N^2}\frac{2}{K(K-1)}\sum_{j=1}^{K}\sum_{k=j+1}^{K}\prod_{\substack{i=1\\ i\neq j,k}}^{K} P_2(x_i) \\
&= \frac{1}{N^2}\frac{2}{K(K-1)}\prod_{i=1}^{K} P_2(x_i)\sum_{j=1}^{K}\sum_{k=j+1}^{K}\frac{1}{P_2(x_j)P_2(x_k)} \\
Q_3 &= \frac{1}{N^3}\frac{6}{K(K-1)(K-2)}\sum_{i=1}^{K}\sum_{j=i+1}^{K}\sum_{k=j+1}^{K}\prod_{\substack{m=1\\ m\neq i,j,k}}^{K} P_2(x_m)
\end{aligned}$$

| | | | |
|---|---|---|---|
| B | ב | P | פ |
| C (see) | צ | Q | ק |
| C, ck (kay) | ק | R | ר |
| D | ד | S (ess) | ס |
| F | פ | S (zee) | ז, ס |
| G | ג | Sh | שׁ |
| H | ה | T | ט |
| J | ג | Th, Ta | ת |
| K | ק | V (next to long vowel) | ב |
| Kn | נ | V (next to short vowel) | ו |
| L | ל | W | ו |
| M | מ | X | קס |
| N | נ | Z | ז |

**Table of transliteration of English consonants into Hebrew consonants**

$$= \frac{1}{N^3}\frac{6}{K(K-1)(K-2)}\prod_{m=1}^{K} P_2(x_m)\sum_{i=1}^{K}\sum_{j=i+1}^{K}\sum_{k=j+1}^{K}\frac{1}{P_2(x_i)P_2(x_j)P_2(x_k)}$$

Hence the probability $Q$ of observing the values $x_1,\ldots,x_K$ given the nearly all and the conditional independence assumption is

$$Q(x_1,\ldots,x_K) = \sum_{i=0}^{3} q_i Q_i(x_1,\ldots,x_K)$$

# B Transliteration of English Names Into Hebrew

In this appendix we give the principles by which English names can be transliterated into Hebrew in all the possible forms. The consonants are simple to transliterate. The table above shows the letter to letter correspondence.

The next table gives the possible transliteration for the vowel sounds of English. The vowels are the place where there is some variability in the sense of putting the vowels in completely or incompletely or some vowels present and some vowels not present. The only exception to this is the long I vowel which has no way in Hebrew of being shown by nikud. Therefore, whenever the English word has a long I sound, either a double יי must be used or a single י. The logic

| English Vowel | Long Vowel | Hebrew | Short Vowel | Hebrew |
|---|---|---|---|---|
| A | C**a**ke<br>H**a**yes, T**ay**lor<br>R**ea**gen | א<br>יי<br>יי | c**a**t<br>Buch**a**nan<br>**A**dams, **A**rthur<br>T**a**ft, Gr**a**nt, C**a**rter | -<br>א<br>א<br>א |
| E | s**ee**k, b**ea**d<br>f**ie**ld, P**ie**rce | י<br>י | s**e**t, J**e**ff**e**rson | - |
| I | b**i**ke<br>T**y**ler | אי, יי | F**i**llmore, Harr**i**son<br>Mad**i**son, W**i**lson<br>Cl**i**nton, N**i**xon | י |
| O | b**oa**t, r**o**se<br>P**o**lk, R**oo**sevelt | ו | Wils**o**n, Clint**o**n | ו |
| U | Tr**u**man, H**oo**ver | ו | p**u**p, Roos**e**velt | ו |

**Table of transliterations of English vowels into Hebrew**

behind this is that the double יי is the modern Hebrew convention to show the long I sound. The single י is also possible because the English long I vowel is a dipthong. It is really a composition of and Ah sound with and ee sound. The Ah sound can be designated by a patach nikud on the previous consonant and the י designates the **ee** sound.

# C Transliterations of American President Names

| No. | President's Name | Last Name | Last Name with Initials |
|---|---|---|---|
| 1 | George Washington | ושינגטונ | ג ושינגטונ |
| | | ושנגטונ | ג ושנגטונ |
| | | ושינגטנ | ג ושינגטנ |
| | | ושנגטנ | ג ושנגטנ |
| 2 | John Adams | אדמס | ג אדמס |
| 3 | Thomas Jefferson | גפרסון | ת גפרסון |
| | | גפרסן | ת גפרסן |
| | | גפפרסון | ת גפפרסון |
| | | גפפרסון | ת גפפרסון |
| 4 | James Madison | מדיסון | ג מדיסון |
| | | מדיסן | ג מדיסן |
| | | מדסון | ג מדסון |
| | | מדסן | ג מדסן |
| 5 | James Monroe | מונרו | ג מונרו |
| | | מנרו | ג מנרו |
| 6 | John Quincy Adams | אדמס | ג ק אדמס |
| | | | ג אדמס |
| 7 | Andrew Jackson | גקסון | א גקסון |
| | | גקסן | א גקסן |
| 8 | Martin Van Buren | ון ביורן | מ ון ביורן |

| No. | President's Name | Last Name | Last Name with Initials |
|---|---|---|---|
| 9 | William H. Harrison | הריסון | ו ה הריסון |
| | | | ו הריסון |
| | | הריסן | ו ה הריסן |
| | | | ו הריסן |
| | | הרסון | ו ה הרסון |
| | | | ו הרסון |
| | | הרסן | ו ה הרסן |
| | | | ו הרסן |
| | | הרריסון | ו ה הרריסון |
| | | | ו הרריסון |
| | | הרריסן | ו ה הרריסן |
| | | | ו הרריסן |
| | | הררסון | ו ה הררסון |
| | | | ו הררסון |
| | | הררסן | ו ה הררסן |
| | | | ו הררסן |
| 10 | John Tyler | טיילר | ג טיילר |
| | | טילר | ג טילר |
| 11 | James Knox Polk | פולק | ג נ פולק |
| | | | ג פולק |
| | | פלק | ג נ פלק |
| | | | ג פלק |
| 12 | Zachary Taylor | טיילור | ז טיילור |
| | | טיילר | ז טיילר |
| | | טלור | ז טלור |
| | | טלר | ז טלר |
| 13 | Millard Fillmore | פילמור | מ פילמור |
| | | פילמר | מ פילמר |
| | | פלמור | מ פלמור |
| | | פלמר | מ פלמר |
| | | פיללמור | מ פיללמור |
| | | פיללמר | מ פיללמר |
| | | פללמור | מ פללמור |
| | | פללמר | מ פללמר |

| No. | President’s Name | Last Name | Last Name with Initials |
|---|---|---|---|
| 14 | Franklin Pierce | פירס | פ פירס |
| | | פרס | פ פרס |
| 15 | James Buchanan | ביוקאנן | ג ביוקאנן |
| | | ביוקנן | ג ביוקנן |
| 16 | Abraham Lincoln | לינקולן | א לינקולן |
| | | לנקולן | א לנקולן |
| | | לינקלן | א לינקלן |
| | | לנקלן | א לנקלן |
| 17 | Andrew Johnson | גונסון | א גונסון |
| | | גנסון | א גנסון |
| | | גונסן | א גונסן |
| | | גנסן | א גנסן |
| 18 | Ulysses S. Grant | גראנט | י ס גראנט |
| | | גראנט | י גראנט |
| | | גרנט | י ס גרנט |
| | | גרנט | י גרנט |
| 19 | Rutherford B. Hayes | הייז | ר ב הייז |
| | | | ר הייז |
| | | | ר ב הז |
| | | | ר ב הז |
| | | האז | ר ב האז |
| | | | ר האז |
| 20 | James A. Garfield | גרפילד | ג א גרפילד |
| | | | ג גרפילד |
| | | גרפלד | ג א גרפלד |
| | | | ג גרפלד |
| 21 | Chester A. Arthur | ארתור | צ א ארתור |
| | | | צ ארתור |
| | | ארתר | צ א ארתר |
| | | | צ ארתר |
| 22 | Grover Cleveland | קליבלנד | ג קליבלנד |
| | | קלבלנד | ג קלבלנד |

| No. | President's Name | Last Name | Last Name with Initials |
|---|---|---|---|
| 23 | Benjamin Harrison | הריסון | ב הריסון |
| | | הריסן | ב הריסן |
| | | הרסון | ב הרסון |
| | | הרסן | ב הרסן |
| | | הרריסון | ב הרריסון |
| | | הרריסן | ב הרריסן |
| | | הררסון | ב הררסון |
| | | הררסן | ב הררסן |
| 24 | William McKinley | מקינלי | ו מקינלי |
| | | מקנלי | ו מקנלי |
| 25 | Theodore Roosevelt | רוזוולט | ת רוזוולט |
| | | רזוולט | ת רזוולט |
| | | רוזולט | ת רוזולט |
| | | רזולט | ת רזולט |
| 26 | William H. Taft | תאפט | ו ה תאפט |
| | | | ו תאפט |
| | | תפט | ו ה תפט |
| | | | ו תפט |
| 27 | Woodrow Wilson | וילסון | ו וילסון |
| | | ולסון | ו ולסון |
| | | וילסן | ו וילסן |
| | | ולסן | ו ולסן |
| 28 | Warren G. Harding | הרדינג | ו ג הרדינג |
| | | | ו הרדינג |
| | | הרדנג | ו ג הרדנג |
| | | | ו הרדנג |
| 29 | Calvin Coolidge | קולידג | ק קולידג |
| | | קלידג | ק קלידג |
| | | קולדג | ק קולדג |
| | | קלדג | ק קלדג |
| 30 | Herbert C. Hoover | הובר | ה צ הובר |
| | | | ה הובר |
| | | הבר | ה צ הבר |
| | | | ה הבר |

| No. | President's Name | Last Name | Last Name with Initials |
|---|---|---|---|
| 31 | Franklin D. Roosevelt | רוזוולט | פ ד רוזוולט |
| | | | פ רוזוולט |
| | | רזוולט | פ רזוולט |
| | | | פ ד רזוולט |
| | | רוזולט | פ רוזולט |
| | | | פ ד רוזולט |
| | | רזולט | פ ד רזולט |
| | | | פ רזולט |
| 32 | Harry S. Truman | טרומן | ה ס טרומן |
| | | | ה טרומן |
| | | טרמן | ה ס טרמן |
| | | | ה טרמן |
| 33 | Dwight D. Eisenhower | אייזנהוור | ד ד אייזנהוור |
| | | | ד אייזנהוור |
| | | אייזנהור | ד ד אייזנהור |
| | | | ד אייזנהור |
| | | איזנהוור | ד ד איזנהוור |
| | | | ד איזנהוור |
| | | איזנהור | ד איזנהור |
| | | | ד ד איזנהור |
| | | ייזנהוור | ד ד ייזנהוור |
| | | | ד ייזנהוור |
| | | ייזנהור | ד ד ייזנהור |
| | | | ד ייזנהור |
| 34 | John F. Kennedy | קנדי | ג פ קנדי |
| | | | ג קנדי |
| | | קננדי | ג פ קננדי |
| | | | ג פ קננדי |

| No. | President's Name | Last Name | Last Name with Initials |
|---|---|---|---|
| 35 | Lyndon B. Johnson | גונסון | ל ב גונסון |
| | | | ל גונסון |
| | | גנסון | ל ב גנסון |
| | | | ל גנסון |
| | | גונסן | ל ב גונסן |
| | | | ל גונסן |
| | | גנסן | ל ב גנסן |
| | | | ל גנסן |
| 36 | Richard M. Nixon | ניקסון | ר מ ניקסון |
| | | | ר ניקסון |
| | | נקסון | ר מ נקסון |
| | | | ר נקסון |
| | | ניקסן | ר מ ניקסן |
| | | | ר ניקסן |
| | | נקסן | ר מ נקסן |
| | | | ר נקסן |
| 37 | Gerald R. Ford | פורד | גר פורד |
| | | | ג פורד |
| | | פרד | גר פרד |
| | | | ג פרד |
| 38 | James Earl Carter | קארטר | ג א קארטר |
| | | | ג קארטר |
| | | קרטר | ג א קרטר |
| | | | ג א קרטר |
| 39 | Ronald W. Reagan | רייגן | ר ו רייגן |
| | | | ר רייגן |
| | | ראגן | ר ו ראגן |
| | | | ר ראגן |
| | | רגן | ר ו רגן |
| | | | ר רגן |
| 40 | George H. W. Bush | בוש | ג הבוש |
| | | | ג בוש |

| No. | President's Name | Last Name | Last Name with Initials |
|---|---|---|---|
| 41 | William J. Clinton | קלינטון<br>קלנטון<br>קלינטן<br>קלנטן | ו קלינטון<br>ו קלנטון<br>ו קלינטן<br>ו קלנטן |
| 42 | George W. Bush | בוש | ג ו בוש<br>ג בוש |

# References


[1] R.J. Aumann, H. Furstenberg, I. Lapides, and D. Witztum. Analysis of the "Gans" committee report. http://www.Ma.huji.ac.il/ raumann/pdf/dp_365.pdf, pp.6-9.

[2] Robert Haralick. Basic concepts for testing the torah code hypothesis. In *International Conference On Pattern Recognition*, 2006.

[3] Robert Haralick. Testing the torah code hypothesis: The experimental protocol. In *International Conference On Pattern Recognition*, 2006.

[4] Robert Haralick, Eliyahu Rips, and Matityahu Glazerson. *Torah Codes: A Glimpse Into the Infinite*. Marcel Printing, 2005.

[5] Ikeuchi Katsuchi, Atsushi Nakazawa, Kazuhide Hasegawa, and Takeshi Ohishi. The great buddha project: Modeling cultural heritage for vr systems through observation. *IEEE ISMAR03*, 2003.

[6] B. McKay, D. Bar Natan, M. Bar Hillel, and G. Kalai. Solving the bible codes puzzle. *Statistical Science*, pages 149–173, May 1999.

[7] D. Witztum, E. Rips, and Y. Rosenberg. Equidistant letter sequences in the book of Genesis. *Statistical Science*, 9(3):429–438, August 1994.