

Torah Codes: Redundant Encoding

Robert M. Haralick
Computer Science
Graduate Center
City University of New York
365 Fifth Avenue
New York, NY 10016
Revision Date: 10/29/03

Key words: Torah codes, Bible codes, Hypothesis testing, Redundant Encoding, Multiple Encoding

Abstract

With the continuing publishing of books on Torah codes, the recent History channel video production and the upcoming BBC Science division video production on Torah codes, it is appropriate to re-examine the controversy sparked by the publication of the first a priori experiment of Witztum, Rips, and Rosenberg (WRR) in *Statistical Science* in 1994. That publication led to an academic controversy of whether or not what was reported in WRR was in fact real. In the end most of the controversy boiled down to whether or not people hold the naturalist hypothesis. However, there were some technical questions about whether the WRR results could have been achieved by setting up a non-a priori experiment and then evaluating the probabilities as if it were an a priori experiment. In a *Statistical Science* article in 1999 McKay et. al. showed that this was indeed possible. This demonstration gave a possible explanation that satisfied those who hold the naturalist hypothesis. From the broader perspective, the McKay et. al. article served to shift the scientific emphasis on whether or not it was possible to distinguish between what are thought to be the a priori Torah codes as in the WRR experiment in *Genesis* from the non a priori codes of the McKay et. al. demonstration in a Hebrew translation of *War and Peace*.

In this paper, we show that in fact there is a such distinction. The WRR a priori codes are redundant. Much of what is encoded is multiply encoded. This happens in the WRR codes and not in the McKay et. al. cooked codes. We show this by developing a better statistical detector. We develop a single compactness measure and a method of combining compactness values to form a test statistic consistent with testing the Null hypothesis of no Torah code effect against the alternative that there is a Torah code effect and that the Torah codes are multiply encoded. We show that with this test statistic WRR rabbis list 1 has a p-value (the probability of observing a result this significant or more by chance) of less than 1/2000; WRR rabbis list 2 and the combined WRR rabbis list 2 with the cooked and non a priori McKay rabbis list 2 have a p-value of less than 1/40,000 while the McKay rabbis list 2 has a p-value of greater than .05 in the Hebrew translation of *War and Peace* text.

After developing the statistical detector, we repeated the Gans cities experiment. This experiment is a priori with respect to the new detector. The p-value of the Gans cities experiment was .01 indicating that there is also multiple encoding in the appellation and city pairings.

Introduction

Unexpectedly compact formations of equidistant letter sequences (ELSs) of logically historically related Hebrew key word sets when they appear in the Torah text are called Torah codes. To properly determine how unexpectedly compact a formation is requires an experiment that follows a protocol. All the elements of the protocol must be chosen *a priori*, meaning they must be chosen before any experiment or snooping around is done.

There are five major parts to the protocol:

- (1) the choice of the logically historically related Hebrew key word sets;
- (2) the choice of search parameters defining how ELSs will be found;
- (3) the choice of a compactness measure for a set of ELSs;
- (4) the choice of how the compactnesses over the different ELS formations and over the various key word sets get combined to form a score;
- (5) the choice of a control population of texts that will be randomly sampled so that the score of the key word sets in the Torah text can be compared to the score of the sampled texts in the control population in order to estimate the p-value (the probability that as good as a result might occur by chance).

The unexpectedness of the compactness associated with the ELSs of a group of Hebrew key word sets is measured by the fraction of randomly sampled texts from the control population whose score is as good as or better than the score of the Torah text. This fraction is the p-value of the experiment. P-values smaller than the chosen significance level of the experiment lead to a rejection of the Null hypothesis of no Torah code effect in favor of the alternative hypothesis.

After the first *a priori* experiment was done and published by Witztum, Rips, and Rosenberg (WRR) (1994), an experiment that showed that the names of great rabbis and their dates of death or birth formed unexpectedly compact formations in the *Genesis* text, there began to be a flurry of activity repeating the experiment in an attempt to explore whether or not the WRR experiment was valid. If the experiment were valid, then it would imply that the *Genesis* text, a text dating over 3300 years old, had in it information about events that occurred thousands of years later. Certainly this would not be a natural expected occurrence. It would be, in fact, a miracle, a totally inexplicable unexpected happening. And for those who believe the naturalist hypothesis, it is not possible that such an experiment could have obtained the result it did.

There were a number of people who did independent replications of the experiment using the same Hebrew key word sets of the great rabbi appellation and dates. These all produced basically the same result as the original experiment.

The only possible explanation, consistent with the naturalist hypothesis, for the successful replication is that the logically historically related Hebrew key word sets were not chosen *a priori*, but were obtained by snooping and selecting. Snooping and selecting means that the researchers tested in their back room a variety of different appellations and spellings of the appellations associated with the great rabbis and weeded out from among them a significant fraction of bad performers. Then with the selected set of appellation dates they performed the experiment. In such

a case, it would be perfectly expected that the fraction of control population texts having a score as good as or better than the Torah text would be very small.

McKay et. al. (1999) performed an experiment doing just that. They made up a list including as many as they could find of the appellations of the great rabbis and of the possible spellings of these appellations, including some spelling stretches, (Witztum,2000) threw away a significant fraction of the bad performers in a text of a Hebrew translation of *War and Peace*, and then did the experiment. They showed that there was a very small fraction of control texts that had as good or better score than the *War and Peace* text did for their non *a priori* list of appellation and dates. Their conclusion was that the same inexplicable event that happens with the *Genesis* text also happens with the *War and Peace* text! In other words, by an appropriate non *a priori* selection of the key word sets it is possible to force a good experimental result.

Thus arises the Torah code controversy. The results obtained by Witztum, Rips, and Rosenberg can be explained by a non *a priori* selection of the appellation and dates. Witztum et. al. claim that the protocol they followed used an independent expert to select and define the appellation and dates. This expert did no Torah code experiments and did not have a back room. Therefore, WRR argue the selection was in fact *a priori* and their experiment was a valid experiment.

Who was in what back room doing non *a priori* things is not publicly observable. Arguments of those who hold the naturalist hypothesis against those who do not hold the naturalist hypothesis are, by themselves, not academically interesting and we do not review them here. See McKay et. al (1999). The question which we address in this paper is whether there is another characteristic that the so called real Torah codes have by which they can be distinguished from the ELS formations created by snooping and selecting in a text such as the *War and Peace* text, a text which everyone agrees has all its compact formations occurring as chance occurrences.

In this paper we show that in fact there is such a characteristic: redundancy. We provide evidence showing that the *Genesis* text has more multiple instances of unexpectedly compact formations of ELSs of the WRR great rabbi appellation and date list than does the list of McKay's great rabbi appellations and dates in the *War and Peace* text. When this multiple encoding or redundancy is properly taken into account into the compactness combining method to determine a test statistic, McKay's great rabbi appellations and dates in the *War and Peace* text lose their statistical significance and the WRR great rabbi appellation and dates in the *Genesis* text retain their statistical significance. Furthermore, when the Gans cities experiment, an experiment that associates the great rabbi appellations and the cities they were born in or died in, is repeated using the new compactness combining method, the result is statistically significant. There is an unexpectedly small fraction (the p-value) of control texts that have as good as or better score for the Gans cities list with the *Genesis* text than expected by chance.

Experimental Errors

We begin our presentation by recalling that different compactness measures and different compactness combining methods yield different test statistics. Different test statistics have different statistical efficiency and power. Even the good ones will balance type (1) and type (2) errors differently. The type (1) error is the error of calling a result significant when in fact it is not

significant and the type (2) the error of calling a result not significant when it is significant. Some measures may be better, even uniformly better than some others.

In the case of Torah code experiments, the p-value of the experiment is the error of type (1) under the conditions that the key word lists of the experiment are *a priori*. Torah code experiments with non- *a priori* snooped and selected key word lists have been done as in McKay et. al. (1999). In these cases the p-value of the experiment is only a lower bound on the true p-value. This gives rise to the concept that there is an internal p-value to the experiment and an external p-value to the experiment. The external p-value is the p-value that results when performing the same type of snooping and selecting on each text of the control text population or, nearly equivalent, performing the same type of snooping or selecting on non-Torah texts from a natural language text population. Each snooped text is put into an experiment which results in an internal p-value. These internal p-values are not true p-values and must just be considered scores. Snooping N texts and performing N experiments results in N scores. The external p-value is the p-value of the score of the selected and snooped key word list. When the experiment is *a priori*, both the internal p-value and external p-value are essentially close in value and have the meaning of the type (1) error. When the experiment is not *a priori* then only the external p-value has the meaning of the type (1) error.

The problem is that snooping is done interactively, requiring good fluency in Hebrew, and does not follow an algorithmic protocol. Therefore, it cannot be so easily incorporated into an experimental protocol. However, not all combining protocols perform equally well for type (1) errors. If Torah codes are real and if there is a structural difference between chance formations and real codes, then it may be possible to design a testing protocol in which the internal p-value, the lower bound computed by experiment, is higher for the control texts. If this were possible it would help distinguish between real codes and cooked codes. It is exactly this that we accomplish in this paper.

Background

Consciously or unconsciously Torah code proponents have wanted to make sure that the p-value of their experiment in the Torah text will be small so that they do not commit an error of type (2), missing an encoding that is there. Up until this point there has been no systematic examination of what happens for the error of type (1), calling a result significant (because of a small p-value -- the McKay list in *War and Peace*, for example) when in fact it is not significant (there are no real encodings in *War and Peace*, only chance or random formations). And because of this -- only working to find *a priori* encodings in the Torah text -- and not checking if this methodology can be purposefully misused to also find apparent encodings where there are no real encodings, the Torah code opponents have had an easy time of cooking up snooping and selecting examples where the resulting experiment yields a small p-value is. Obviously, if the same methodology permits one to say there are encodings in the *War and Peace* text, where there is none, then the methodology is worthless when applied to the *Genesis* text, which is believed by some to have encodings. It is this methodological defect/limitation which McKay et. al. have exploited, arguments of *a priori* and non *a priori* lists put aside.

As an aside, by presenting best most compact tables, Torah code lecturers set themselves up for problems because one can easily snoop cooked up tables with apparently good compactness in non-

Torah texts. Their only weapon of reply is that the Torah table was done in an *a priori* manner where as the cooked up table was done by snooping. If the difference between something that is real and something that is fake depends on whether the researcher did it *a priori* or whether the researcher did it by snooping and selecting, then the public can well question the methodology. The value of a scientific experiment cannot depend on whether a person snooped in his back room, since whether or not he/she snooped is not so easily publicly observable. A scientific experiment must be entirely observable in an easy way. This is the point to the Barry Simon wiggle room argument: a researcher can wiggle in a non *a priori* manner in his back room and nobody outside will know. Although it is possible to observe a methodology as *a priori*, it is not easy to set up the controls to properly do so.

The structural characteristic of ELS formations we will explore in this paper is that of multiple encoding or redundancy. Since the first “white preprint” paper of Witztum and Rips, it has been noticed informally and anecdotally that some Torah codes are redundant. More than one compact table of the same key words can be found. The same key word pair has multiple ELS pairs many of which are more compact than expected by chance. On the basis of this observation, we now make a formal a change in the alternative hypothesis to the Null hypothesis of no Torah code effect. Our alternative hypothesis changes from: “logically historically related key words tend to have significantly more compact meetings more often than expected by chance” to “logically historically related key words tend to have significantly more compact meetings more often than expected by chance and the pairwise meetings are more multiply encoded than expected by chance.” When we make this change in the alternative hypothesis we are logically compelled to design a combining methodology, a test statistic, that takes multiple encoding into account. To do this we must understand what happens to our test statistic in the case that the alternative hypothesis is true and in the case that the alternative hypothesis is false. In the case that the alternative hypothesis is false it may be false because there are no good meetings or because there is a good best meeting but not enough good near best meetings, to make multiple encodings. What we want to do is to design a test statistic (a compactness combining methodology) which under the condition that the alternative hypothesis is false produces a p-value that is not statistically significant (large) and when the alternative hypothesis is true, produces a p-value that is statistically significant (small).

In this paper we succeed in designing such a test statistic. Our experiments show that the McKay cooked encodings in the *War and Peace* text are as good for its best ELS pairs as are the WRR encodings in the *Genesis* text. But the WRR encodings are better than the cooked encodings for the near best encodings as well. We will prove that it is this structural factor that distinguishes the difference between the McKay cooked encodings in *War and Peace* and the WRR real encodings in *Genesis*.

Experimental Protocol

We begin with the definition of the H_3 compactness measure, as discussed in (Haralick,2003). This measure is the one which for each ELS pair selects the best cylinder size. But instead of taking the harmonic mean across ELS pairs of each key word pair and the geometric mean across key word

pairs and rabbis (as does WRR), we index the compactness scores of the best 20 ELS pairs for each key word pair for each rabbi. We will take products of the rank normalized compactness values, keeping like indexed terms together. Our protocol will determine a best way to define a test statistic by linearly combining the logarithms of the resulting 20 terms with a simple monotonically decreasing weight.

Let R designate the set of rabbis, $r \in R$ designate a rabbi and let $W(r)$ designate the set of appellation date key word pairs for the rabbi. Let an appellation date key word pair be $(w_1, w_2) \in W(r)$. Let $T(r)$ designate the set of texts for rabbi r and let $t \in T$. For each key word w , let $E(w, t)$ be the set of ELSs of w in text t . (The maximum skip for each ELS is set so that the number of ELSs that would be expected to occur in skips 1 through the set maximum skip is just 10 or more.) For any pair (e_1, e_2) of ELSs let $C(e_1, e_2)$ be the set of cylinder sizes that resonate with e_1 or e_2 . Here we define a cylinder size to resonate with an ELS if the row skip of the ELS on the cylinder is 10 or less the cylinder size differs no more than one column skip from the natural column skip of the ELS on the cylinder.¹ Let $d_{min}(e_1, e_2, c, t)$ be the closest squared distance on the cylinder of size c between the letters of ELSs e_1 and e_2 in text t . For any ELS e and cylinder size c and text t let $s(e, c, t)$ be the sum of the squared row skip and squared column skip of ELS e on a cylinder of size c in text t .

For each word pair (w_1, w_2) and text t , we select the smallest 20 terms from

$$\{ \min_c d_{min}(e_1, e_2, c, t) [s(e_1, c, t) + s(e_2, c, t)] \mid e_1 \text{ in } E(w_1, t), e_2 \text{ in } E(w_2, t) \}$$

We designate these smallest 20 terms by $z_1(w_1, w_2, t), \dots, z_{20}(w_1, w_2, t)$. Now instead of taking the harmonic mean over all appellation date ELS pairs and then the geometric mean over all appellation date word pairs and rabbis as done in WRR and our earlier experiments, we just store the H_3 type compactness values for the 20 best ELS pairs of each appellation date pair. The data record for each text then consists of 20 times the total number of appellation date pairs, having ELSs, taken across all rabbis.

We do this for the real text and for each monkey text. The monkey texts are generated from the ELS random placement model. The resulting data file is then rank normalized. Each entry is replaced with the number of entries in its column that is smaller in value plus one half the number of entries equal to its value, the sum divided by K , the number of texts (the number of records in the file). The smallest possible value in a column is then $.5/K$ and the largest possible value in a column is $(K-.5)/K$. Then the product is taken over all comparably indexed rank normalized compactness values. We designate the rank normalized values for word pair (w_1, w_2) and text t by $z_{(1)}(w_1, w_2, t), \dots, z_{(20)}(w_1, w_2, t)$. The product is taken over comparably indexed terms over all word pairs over all rabbis. We designate the resulting product by $q_1(t), \dots, q_{20}(t)$

¹ The natural cylinder size C for an ELS of skip s satisfies $s = mC + c$, where $|c| < C/2$ and m is the row skip of the ELS on the cylinder.

$$q_k(t) = \prod_{r \in R} \prod_{(w_1, w_2, t) \in W(r)} z_{(k)}(w_1, w_2, t)$$

and the normalized ranks, which are the term p-values, by $q_{(1)}(t), \dots, q_{(20)}(t)$.

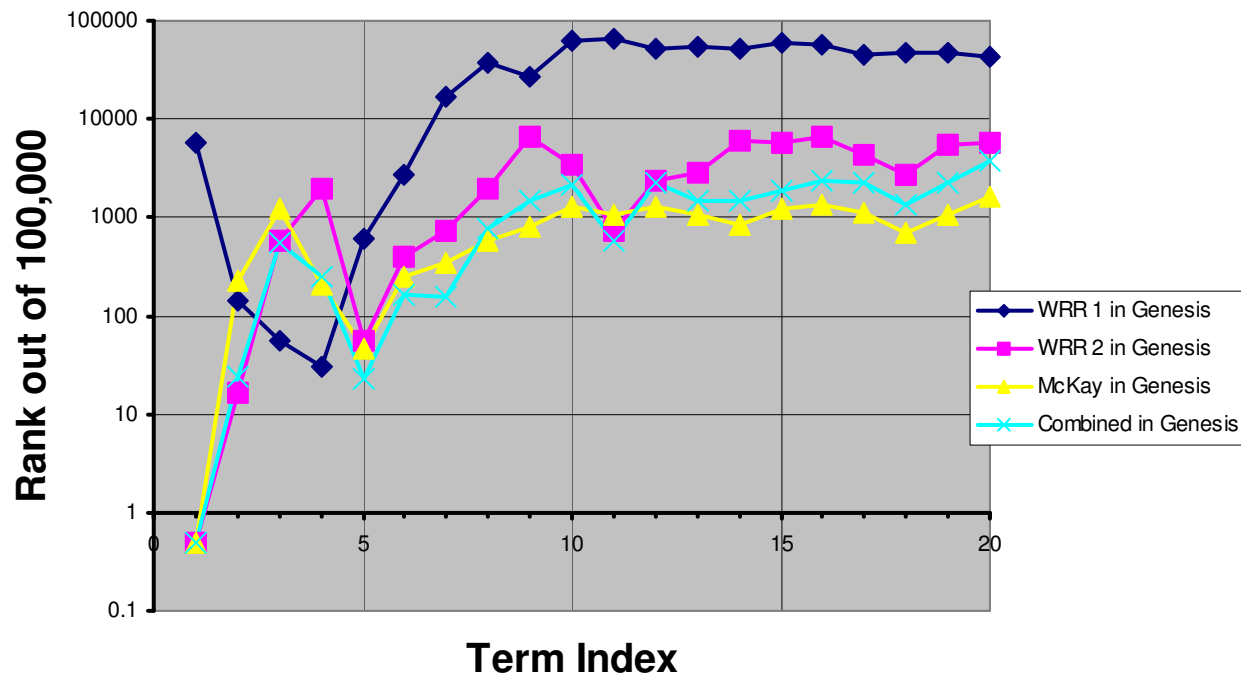
We wish to know whether there is any evidence to support the redundancy hypothesis for encodings in the *Genesis* text compared to encodings in the *War and Peace* text. To do this we will examine the p-values of the record generated by the *Genesis* text and the p-value of the record generated by the *War and Peace* text. Then we will optimize a simple monotonically decreasing weighting scheme to combine together the terms within each record so that we will have one p-value for each text. The optimization attempts to make the p-values of the *Genesis* text with the WRR list small and statistically significant and the p-value of the *War and Peace* text with the McKay list large and statistically insignificant.

Results

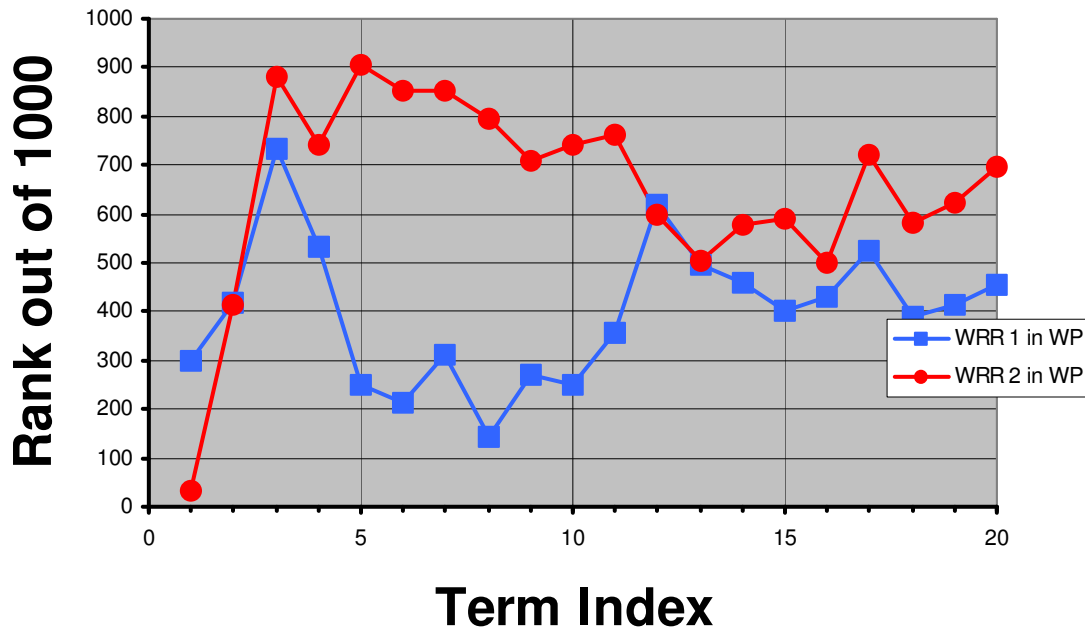
When we examine the p-values of each of these resulting 20 columns for the first data record (the real text), we see that the first few columns tend to have a very small p-value (for each WRR list in *Genesis*) and that the other terms, sometimes up to term 20, also have a smaller p-value than expected by chance. Generally, as the term index increases, the p-value tends to increase to chance value. We showed this with the tools we had in hand at the Torah Code conference in Jerusalem in June 2003. At that time we estimated p-values by the using the Fisher statistic across the rabbis, a method that assumes statistical independence of appellations date pairs to appellation date pairs. The Table below shows these p-values directly observed from the Monte Carlo experiment. In these tables there are no statistical independence assumptions associated with the observed values. The table has four appellation date pair lists: `rabbis1` designates WRR rabbis list 1, `rabbis2` designates WRR rabbis list 2, `McKay` designates the McKay cooked up rabbis list 2, and `rabbis2c` designates the combined list of WRR rabbis list 2 and the McKay list.

Each entry has a numerator and denominator. The numerator is the number of texts having smaller value than the real text plus one half the number of texts having equal value to the first text. The denominator is one plus the number of sampled control texts (the total number of texts in the experiment).

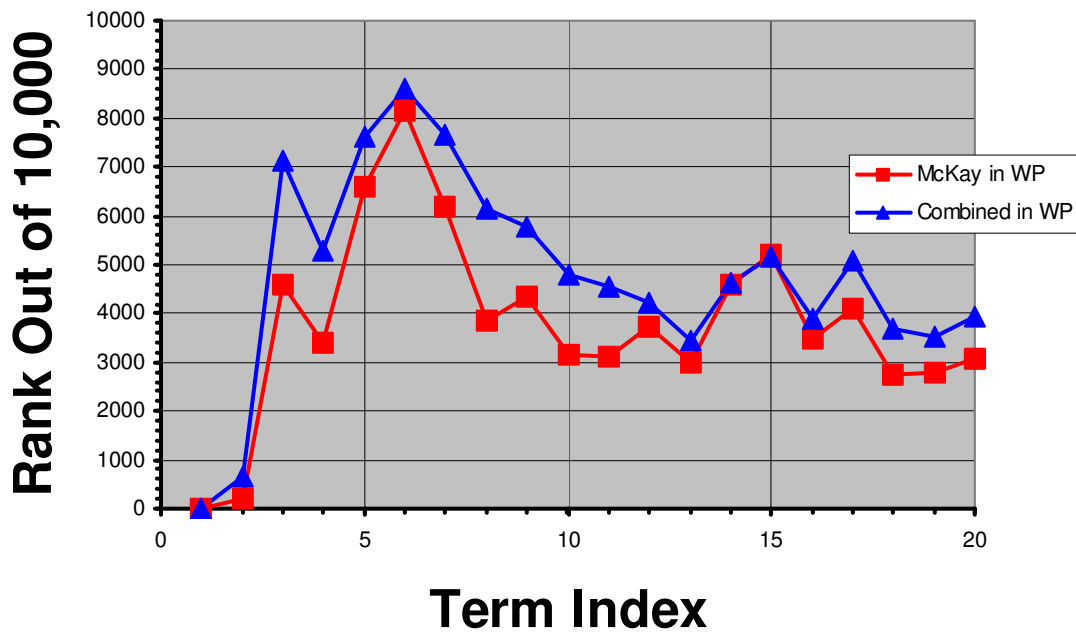
Redundancy P-values in Genesis



Redundancy P-values in War and Peace



Redundancy in War and Peace



Let us examine this table for the experiments done with the *Genesis* text first. Rabbis list 1 has its first two terms with p-values greater than .001. This undoubtedly is the reason why the rabbis list 1 with the *Genesis* text does not do as well with the H_3 measure as with the H_4 measure² with previous protocols (Haralick,2003). It is not until we get to terms three and four that we observe a pair of p-values less than .001. Terms five through nine have p-values significantly smaller than expected by chance although much larger than .001. Terms ten through sixteen are right in the middle of what is expected by chance and the last four terms are slightly smaller than expected by chance.

Rabbis list 2 and list 2c have their first two terms with p-values smaller than .001. And all their remaining terms significantly smaller than expected by chance. None of the terms for the McKay list in *Genesis* have a p-value smaller than .001. But all of the terms are somewhat smaller than expected by chance.

All terms for Rabbis list 1 and list 2 in the *War and Peace* text have large p-values as expected. The first term for rabbis list 2c has a p-value less than .001. Its second term is significantly smaller than expected by chance. Terms three through nine are completely not significant. All but two terms of the terms ten through twenty have p-values less than .5. The first term for the McKay list in *War and Peace* has a p-value less than .001. The second term has a p-value smaller than expected by chance. The remaining terms are not significant, although the last few are slightly better than expected by chance.

The pattern that we observe in this table of p-values is clear. Where encodings are expected, either because they are believed to be real or because they are known to be cooked, the p-values are small for the first term or near the first term. The difference between the cooked encodings in *War and Peace* and the encodings in *Genesis* is that after the best encoding, the near best encodings in *Genesis* have significantly smaller p-values compared to the near best encodings in *War and Peace*. This is due to the redundancy of the encodings in *Genesis*.

The table of p-values just discussed provides evidence of the redundancy of the encodings in *Genesis*, but it is only a partial result. It is partial because there was no experiment that provides a methodology that combines these twenty terms together so that we may associate one p-value with each text and list. Our issue then is how to combine the values across the 20 columns. Since the column 1 entry generally has the best combined compactness value, one would intuitively want it to have the highest weight in the combining. The column 2 entry should have less weight than the column 1 entry and so on. The weights should be monotonically decreasing with term index. Therefore, we explored a class of weighting done by weighting the logs of the entries in the columns by a monotonically decreasing weight of the form $a-i$, where i is the index of the column. We manually ran a few trial values and settled on a value of $a=10.5$. Column indexes go from 0 to 19 and the respective weights go as 10.5,9.5,.....,-8.5. This means the log of the first product is multiplied by 10.5, the log of the second product is multiplied by 9.5,....., the log of the 20th product

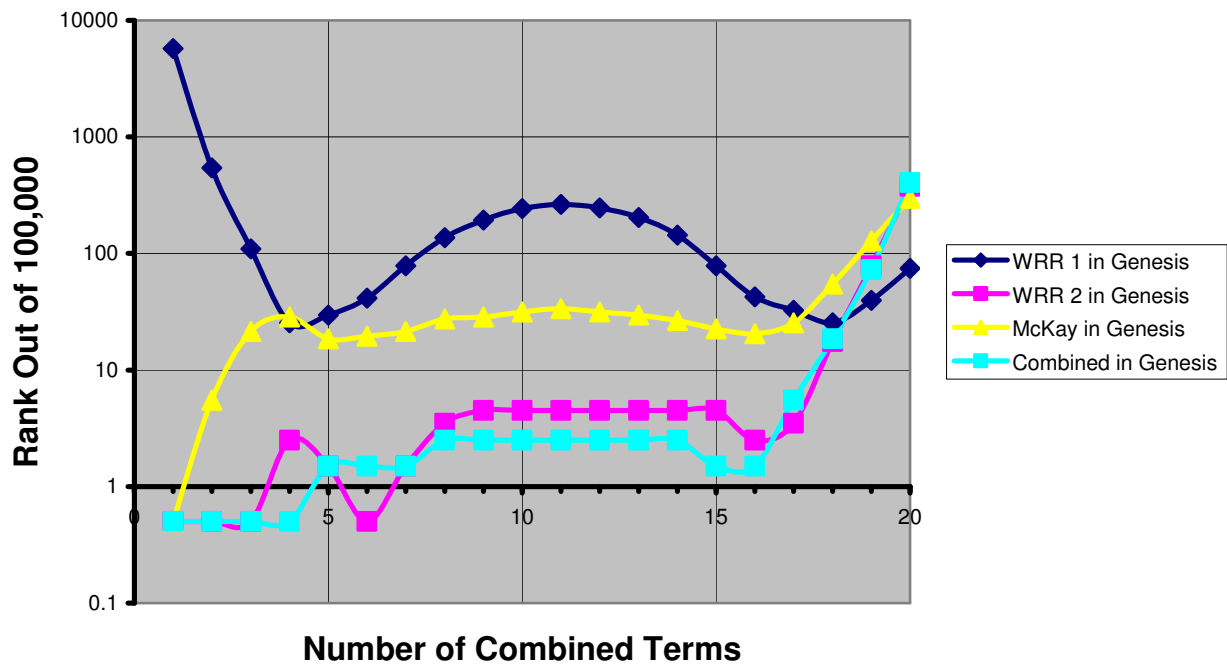
² The H_4 measure is based on

$$\{ \min_{c \text{ in } C(e_1, e_2)} d_{max}(e_1, e_2, c, t) [s(e_1, c, t) + s(e_2, c, t)] \mid e_1 \text{ in } E(w_1, t), e_2 \text{ in } E(w_2, t) \}$$

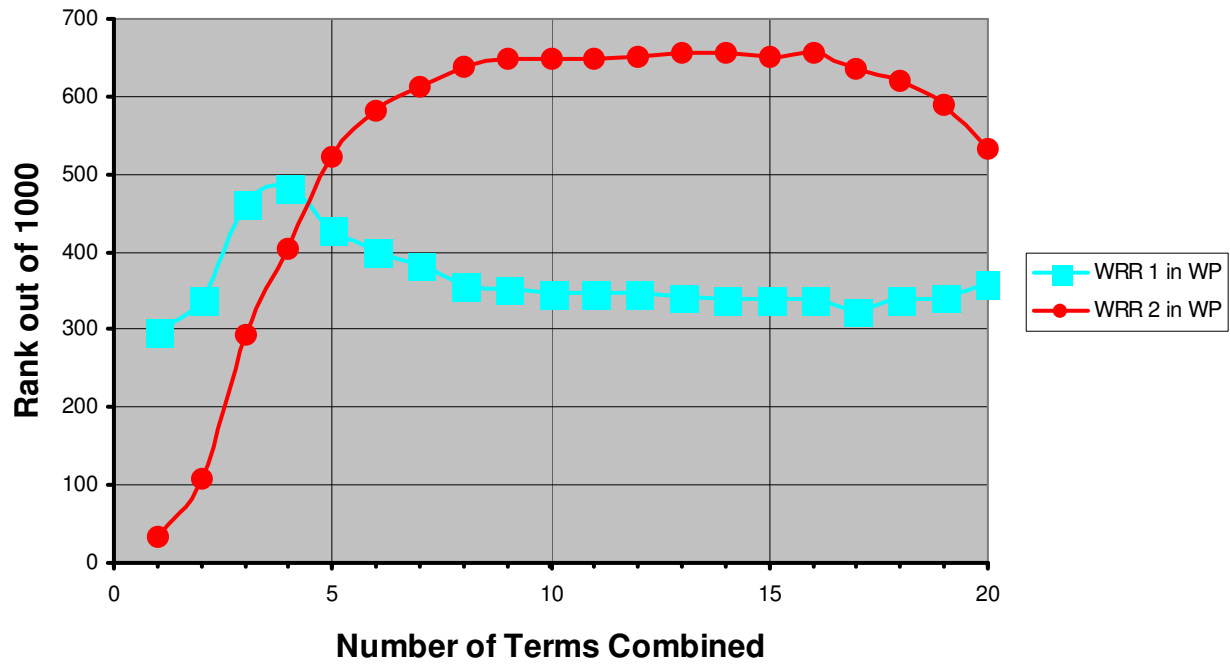
is multiplied by -8.5 . The combined value is the sum of the weighted logs; this is the test statistic used to test the Null hypothesis.

The issue we have in combining is how many of the 20 terms should be combined. We must combine only up to those terms that yield a small p-value result for the two WRR rabbis lists and a large p-value for all lists on the *War and Peace* text. We can combine the first 2 out of the 20, we can combine the first 3 out of the 20,..., we can combine the 20 out of the 20. After combining, and rank normalization, each text has one rank normalized value. The following graphs summarize what we have observed for the rank normalized value of the real texts (first row) as a function of how many terms are combined.

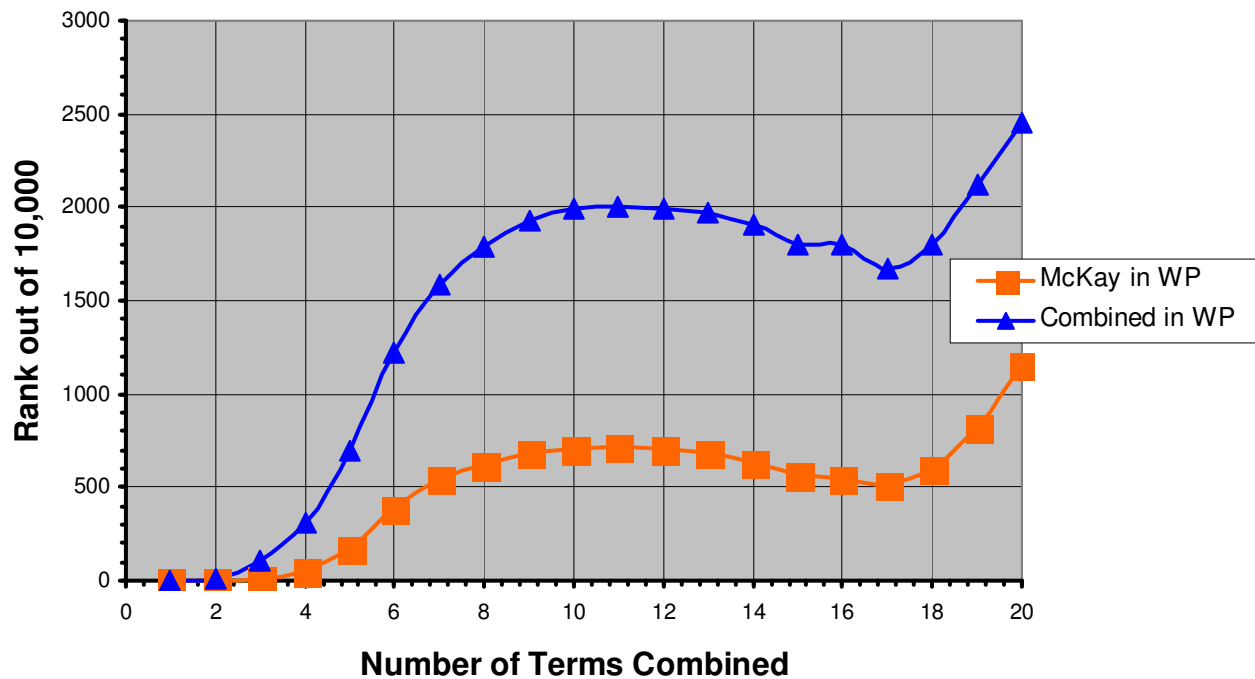
Redundancy P-values in Genesis



Redundancy P-Values in War and Peace



Redundancy in War and Peace



Let us examine the different lists with the *Genesis* text. We observe again that the best term for rabbis list 1 in *Genesis* has a p-value of .0578. The best term is not that good. And this is the reason that rabbis list 1 does not perform well with the H_3 compactness measure in the protocol that takes the harmonic mean. The harmonic mean can be thought of as a soft way to compute with the minimum value. The character of the compactness values for rabbis list 1 in *Genesis* is that the best, and next to best compactness values are not that good. However, because of the redundant encoding, of rabbis list 1 in *Genesis*, the compactness values with indexes between 3 through 20 have small p-values. Examine the normalized ranks of the combined values when terms 1 through 3, 1 through 4, 1 through 5, 1 through 6, 1 through 7, 1 through 15, 1 through 16, 1 through 17, 1 through 18, 1 through 19, and 1 through 20 are combined. These p-values are all below .001. Rabbis list 2 and rabbis list 2c have small p-values for combining terms from one term through eighteen terms. The McKay list with the *Genesis* text yields a p-value less than .001 for only using the best term and best two terms. However combining the best three through the best eighteen results p-values of less than .01 for the McKay list.

Now we turn our attention to the *War and Peace* text. Rabbis list 1 and 2 in the *War and Peace* text have, as expected, p-values much greater than .001 for all combinations. The McKay list in *War and Peace* and consequently the Rabbis 2c list in *War and Peace* have small p-values when the best and next to best terms are combined. But by the time five or more terms are combined the p-values of these two lists in the *War and Peace* text grow to greater than .05.

What we learn from these experiments is that the McKay list, which was cooked up for *War and Peace*, performs well in a protocol heavily weighting the best terms because its best terms were good. However, its terms of index greater than 3 are not so good. The WRR lists with the *Genesis* text, on the other hand, maintain p-values less than .001 for protocols that combine the best three terms, best four terms, best five terms, best six terms, best seven terms, best fifteen, best sixteen, best seventeen, and best eighteen terms. Notice that the optimization does not have to balance the WRR rabbis list in *Genesis* against the McKay rabbis list in *War and Peace*. Starting from terms of index 5 or higher, the McKay list in *War and Peace* does not have a statistically significant level (smaller than .001). The balance the optimization must achieve is a balance of WRR rabbis list 1 against WRR rabbis list 2.

One methodology that produces small p-values for the *Genesis* text with the WRR rabbi list is one where the first four, five, or six terms must be combined. In this case the p-values are all less than .001 and the McKay list with the *War and Peace* text has a p-value greater than .001, much greater for combining the first five or six terms. Combining eight terms weakens the result for rabbis list 1 in *Genesis* to just barely less than .001. The second methodology of combining that enables the distinction between real and cooked encodings is combining the first fifteen, sixteen, seventeen, and eighteen terms. Combining nineteen or twenty terms weakens the result of rabbis list 2. Combining fourteen terms weakens the results on rabbis list 1. The middle point seems in combining fifteen or sixteen terms. This results in Rabbis list 2 and the combined Rabbis list 2 having p-values of less than .000025, Rabbis list 1 having a p-value of .00045 and the McKay rabbis list 2 having a p-value of .00205 while the McKay list on *War and Peace* has a p-value of .05435.

Discussion

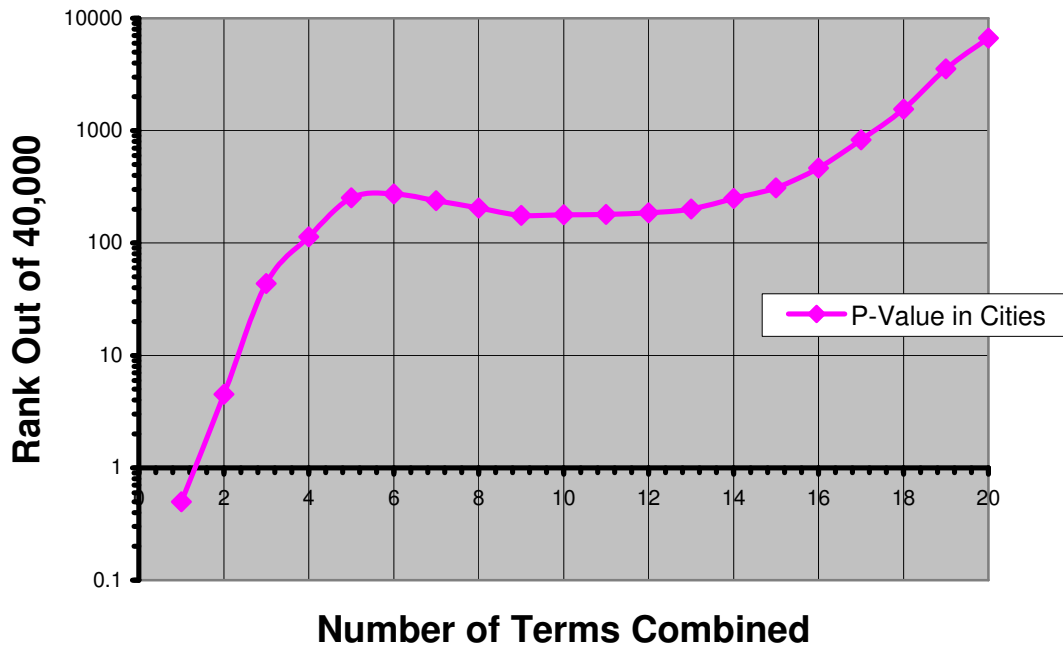
The difference in the behavior between the WRR rabbis list 1 and WRR rabbis list 2 on the H_3 and H_4 measures that was reported on previously (Haralick, 2003) is now resolved. It was observed that WRR rabbis list 1 produces a small p-value in *Genesis* for the H_4 but not the H_3 measure when the combining method across ELS pairs of the same key word pair was harmonic mean. It was also observed that WRR rabbis list 2 produces a small p-value in *Genesis* for the H_3 but not the H_4 measure. This was puzzling. Now we understand that because the harmonic mean too heavily weighs the best encodings, and because the WRR rabbis list 1 has smaller than expected best encodings by the H_3 measure but not so good that they have a p-value of less than .001, WRR rabbis list 1 does not produce a small p-value by the H_3 measure when the test statistic has a harmonic mean combining component. With our new test statistic, which does not use a harmonic mean combining scheme and which more heavily weighs the contribution of the more near best terms, WRR rabbis list 1 and WRR rabbis list 2 do produce a small p-value with the single H_3 measure.

The systematic great rabbi experiments reported here are not to be considered *a priori* experiments to establish the significance of the p-value for WRR rabbis list 1 or WRR rabbis list 2. This was already done by both the WRR experiments and our prior experiments (Haralick, 2003). In this study, we clearly did, in some informal way, an optimization to develop a better test statistic by which future *a priori* experiments can be done in a way that makes it more difficult for cooked up non-redundant encodings to be discovered to be significant in a non Torah text. With our new methodology we have a statistical technique that utilizes the structural redundancy in encodings in the Torah text to distinguish between WRR Torah encodings from false or chance formations in a non Torah text. Our technique therefore, has better false alarm rate statistics than the previous WRR experiment or the previous Haralick experiments. In the next section we discuss *a priori* results on the Gans cities list using the new methodology that takes into account redundancy.

The Gans Cities Experiment

The Gans cities experiment pairs the appellations of each of the great rabbis of the WRR lists 1 and 2 with the cities in which the great rabbis were born or died. The names of the cities were spelled in an algorithmic way based on the spellings that would be acceptable in a *get*, the Jewish divorce document. Below is the table showing the combined compactness value as a function of the number of terms used in combining from the 20 terms available. This experiment was run with 40,000 trials.

Redundancy in Gans City Experiment



There are two interesting facts shown by this table. First the table shows that the p-value using the best term only is less than $1/40000$. Second the table shows that with combining from one term through the 15 successively best terms, the p-value of the experiment is less than .01.

We conclude that the cities experiment shows redundant encoding at the .01 significance level. This redundant encoding is not as strong as what we have measured in the rabbis list 2 or rabbis list 2c experiments. Nevertheless, the evidence is certainly there for redundant encoding through 15 multiple encodings.

Conclusion

In conclusion, our experiments have explored a structural difference between real codes and cooked codes. Our experiments have found one difference (it may not be the only difference): redundancy. Real codes are redundantly encoded. Cooked codes not so. We optimized a statistical methodology to distinguish this difference while maintaining small p-values for the WRR appellation date pairs in the *Genesis* text that are believed to be encoded. Weighing the first sixteen log terms with weights of decreasing magnitude from 10.5 to -5.5 made the distinction possible. This weighting yielded a small p-value, smaller than .00025, for the WRR rabbis list 2 in the *Genesis* text and a large p-value, .0543, (more than 217 times larger than .00025) for the McKay list in the *War and Peace* text. Clearly, when the test statistic is tilted too heavily to weigh the compactness values of the best ELS pairs, such as the WRR methodology which uses a harmonic mean, it does not

properly weigh the near best encodings, which is the distinguishing factor between the WRR encodings in *Genesis* and the McKay encodings in *War and Peace*.

After developing the redundancy detection methodology on the great rabbis lists, we then tested this detection methodology on the Gans cities list. The test on the cities list clearly represents an a priori test. We observed redundancy there statistically significant at the .01 significance level. We conclude that the Witztum et. al. great rabbis list is redundantly encoded and the Gans cities list is redundantly encoded in the Genesis Torah text.

References

- Robert Haralick, *Testing the Torah Code Hypothesis: The Torah Code Effect Is Real*, 2003, submitted for review and publication
- Brendan McKay, Dror Bar-Natan, Maya Bar-Hillel, and Gil Kalai, "Solving the Bible Code Puzzle", *Statistical Science*, Vol 14, No 5, May 1999, pp. 150-173 .
- Doron Witztum, Eliyahu Rips, and Yoav Rosenberg, "Equidistant Letter Sequences in the Book of Genesis," *Statistical Science*, Vol. 9, No. 3, 1994, pp. 429-438.
- Doron Witztum, *Of Science and Parody: A Complete Refutation of MBBK's Central Claim*, http://www.torahcodes.col.il.paro_hb.htm, (2000).