The Hidden Birth Dates of Personalities of Genesis

Doron Witztum

Abstract

Witztum, Rips and Rosenberg [4] have shown that when the Book of Genesis is written as two-dimensional arrays with the topology of a cylinder, equidistant letter sequences spelling words with related meaning appear non randomly in close proximity. Here we adopt the quantitative tools developed in [4] to measure another type of pattern found in the same Book. In the new patterns, equidistant letter sequences spelling out expressions appear in close proximity, on two-dimensional arrays, with conceptually related expressions appearing in the string of letters of the text. We measure the significance of this new type of pattern, using two samples of name-date word pairs, thus similar to the experiment done in [4]. The p-levels obtained were 0.00051 *and* 0.000046 *respectively.*

1. Introduction

Witztum, Rips and Rosenberg [4] (henceforth referred to as WRR) study systematically the assumption that a "hidden text" may exist in the Hebrew text of Genesis (G), consisting of words and phrases expressed in the form of equidistant letter sequences (ELS's) – that is, by selecting sequences of equally spaced letters in the text. Specifically, WRR study the appearance in close proximity of "noteworthy" ELS's spelling out words with related meaning. An ELS in a given text (such as G), denoted by (n, d, k), is a sequence of letters in that text, found at the positions $n, n + d, n + 2d, \ldots, n + (k-1)d$. d is called the skip of the ELS. The "noteworthy" ELS's are those for which the skip $|d|$ is *minimal* over the whole text or over large parts of it; for short they are called *minimal* ELS's.

WRR investigate this hypothesis, by tracing the following patterns: *The appearance of minimal ELS's spelling out conceptually related expressions in close proximity, on two-dimensional arrays.*

A two-dimensional array of letters is formed on the surface of a cylinder by spiraling down the text on the surface as one continuous helix. The distance between the ELS's is defined according to the ordinary two-dimensional Euclidean metric.

Let us call such pattern "Pattern of type A" (PTA). WRR develop a methodology for testing the significance of the PTA's in G according to accepted statistical principles. After choosing certain proximity measures and constructing two samples of word pairs to test, they perform a randomization test and obtain a p-value of 0.00002.

In this paper we conjecture that another type of patterns might also exist in G, in which minimal ELS's spelling out expressions appear in close proximity with conceptually related expressions appearing in the string of letters (SL) of the text, i.e with "skip" ± 1. At first glance it seems that such a pattern may be just a special case of a PTA in which one ELS appears with "skip" ± 1. But bearing in mind that we are looking for encoded information in a "hidden text," there may be profound difference between ELS's and SL's. On the one hand, the vocabulary of the ELS's is quite unlimited. In contrast, the vocabulary of the SL's is limited to that of the text, plus relatively small number of word created by backward reading, by fractures of words and by combining the end of one word in the text with the beginning of the next one. This puts heavy constraints on possible encoding through SL's. Therefore, it is not a priori clear how the SL's are connected to the "hidden text." In particular, the existence of PTA's does not imply the existence of the following patterns: *The appearance of minimal ELS's spelling out expressions, in close proximity, on two-dimensional arrays, with conceptually related expressions appearing as SL's.* We call such pattern "Pattern of type B" (PTB). To study these patterns we follow the methodology developed in [5]. This methodology is essentially that of WRR adopted to the case of PTB's.

Here we replicate the experiment done in [4]. Each PTA studied in [4] is determined by a pair of minimal ELS's in G, where one ELS spells out a name or an appellation of a famous rabbinical personality, and the other ELS - his date of birth or death. But, we cannot use these names, appellations, or dates for the case of SL's. They are not expected to appear as SL's in G. Therefore, we take instead the dates of birth and death of personalities appearing in the plain text of G, and examine the PTB's determined by the names of these personalities (as SL's) and their dates (as ELS's). We apply the same measuring scheme as in [5], and perform similar randomization tests. We obtain very small p-values,

that is, we find the results highly statistically significant.

2. Outline of the procedure

In this section we describe the test in outline. In the Appendix, sufficient details are provided to enable the reader to repeat the computations precisely, and so to verify their correctness.

We test the significance of the PTB's on samples of pairs of related words. To do this we must do the following:

(i) define the notion of "distance" between any two words, so as to lend meaning to the idea of words in "close proximity";

(ii) define statistic that express how close, "on the whole", the words making up the sample pairs are to each other (some kind of average over the whole sample);

(iii) choose a sample of pairs of related words on which to run the test; and

(iv) determine whether the statistic defined in (ii) are "unusually small" for the chosen sample.

Notice that the procedure here described is identical to that of WRR, except for minor changes in the details of task (i), due to the fact that here we consider a "distance" between an ELS and a SL, and not a "distance" between two ELS's as in WRR.

Task (i) is done in detail in Section A.1 of the Appendix. For task (ii), we use two different statistics, P_1 and P_2, which are defined and motivated in Section A.5 of the Appendix of [4]. Intuitively, each measures overall proximity in a different way. In each case, a small value of P_i indicates that the words in the sample pairs are, on the whole, close to each other.

To accomplish task (iii) we composed two samples, S_1 and S_2, of pairs of expressions (w, w'), where w's are words appearing as ELS's, and w''s are words appearing as SL's. Like the samples of WRR, S_1 is built from a list L of personalities, (p'), and the dates (Hebrew day and month), (p), of their birth or death. The personalities are those mentioned in the plain text of G, for whom the relevant dates are found in Talmudic literature. Some of the personalities may have several names, there are variations in spelling and there are different ways of designating dates. Thus each date-personality pair (p, p') correspond to several word pairs (w, w'). S_2 is a subset of S_1. The precise method used to generate L, S_1 and S_2 is explained in the Appendix (Section A.2).

Finally, we come to task (iv), the significance test itself. L consists of 14 personalities. For each of the 14! permutations π of these personalities, we define the statistic $P_j^\pi(j = 1, 2)$ obtained by permuting the personalities with accordance to π, so that Personality i is matched with the dates of Personality $\pi(i)$. The 14! Numbers P_j^π are ordered, with possible ties, according to the usual order of the real numbers. If the PTB's under study were due to chance, it would be just as likely that P_j occupies any one of the 14! places in this order as any other. This is our null hypothesis. To calculate the p-value, we chose 999, 999 random permutations π of the 14 personalities. Each of these permutations π determines a statistic P_j^π; together with P_j we have thus 1, 000, 000 numbers. Define the rank order of P_j among these 1, 000, 000 numbers as the number of P_j^π not exceeding P_j; if P_j is tied with other P_j^π, half of these others are considered to "exceed" P_j. Let r_j be the rank order of P_j, divided by 1, 000, 000; under the null hypothesis, r_j is the probability that P_j would rank as low as it does. This test is identical to that in [4], except the number of personalities to be permuted; in WRR's list were 32 personalities. The details of this test are given in WRR's Appendix (Section A.6). Here we do it for S_1 and S_2. For each sample we get the rank orders r_1 and r_2. The resulting p-value for each sample is (by Bonferroni inequality) $r_0 = 2 \min r_j$.

3. Results and Conclusions

In Table 1 we list the rank order of P_j among the 1, 000, 000 corresponding P_j^π. Using the Bonferroni inequality this yields a p-level of 0.00051 for S_1 and a p-level of 0.000046 for S_2. Note that $S_2 \subset S_1$ so that the two p-levels are not independent.

Table 1. Rank order of P_j among one million P_j^π

Text	Sample S_1		Sample S_2	
	Rank order of P_1	Rank order of P_2	Rank order of P_1	Rank order of P_2
G	255	5, 026	23	1, 113
U	237, 343	545, 695	269, 315	651, 402
V	590, 954	192, 069	542, 410	129, 917
W	135, 494	242, 856	255, 286	326, 463
R	524, 689	432, 265	524, 689	432, 265
T	365, 039	792, 164	365, 039	792, 164
I	104, 022	361, 153	224, 536	526, 639

The same calculations, using the same 999, 999 random permutations, were performed for all of the control texts used in [4] (see there for details). U was obtained from G by permuting randomly words within each verse. V was obtained by permuting the verses of G randomly. W was obtained by permuting the words of G randomly. Note that unlike a control text in which the letters of G are randomly

permuted, these control texts maintain all the personality names in the text. R was obtained by permuting the letters of G randomly. T consists of the initial segment of the Hebrew translation of Tolstoy's *War and Peace* of the same length as G. I is the Book of Isaiah. (Note that R contains only 8 of the personality names as SL's, T only 6, and I 11.) It is obvious from Table 1 that the p-levels for the control texts are well within random expectation.

We conclude that the appearance of ELS's spelling the dates of personalities in the Book of Genesis is in closer proximity to the appearance of their names in the text than can be attributed to chance.

A. Details of the Procedure

Here and in the references indicated in this paper there are sufficient details to enable the reader to repeat the computations precisely. Section A.1 deals with the details of task (i), which are those of WRR with minor changes due to the transition from ELS's to SL's. Section A.2 deals with the construction of list L and Samples S_1 and S_2. The text we used, G, is identical to WRR's.

A.1. Details of Task (i)

We first define the "distance", $\delta_h(e, e')$, between an ELS $e = (n, d, k)$ and a SL $e' = (n', \pm 1, k')$ in G, by writing the text as a single helix of letters spiraling down a cylinder with h vertical columns of letters, and setting $\delta_h(e, e') \equiv f^2 + l^2 + 1$, where f is the usual 2-dimensional Euclidean distance (in columns and rows of letters) between two consecutive letters of e on the surface of the cylinder, 1 is the distance between two consecutive letters of e', and l is the minimal distance between a letter of e and a letter of e'. Then $\mu_h(e, e') \equiv 1/\delta_h(e, e')$ is directly related to the compactness of the configuration of e and e' on the cylinder for given h. As in WRR, we choose $h = h_i =$ the nearest integer to $|d|/i, i = 1, 2, \ldots, 10$. We define $\sigma(e, e') \equiv \sum_{i=1}^{10} \mu_{h_i}(e, e')$. Note that $\sigma(e, e')$ tends to be large provided that there is a relatively compact configuration of e and e' on a cylinder for at least one $h_i, i = 1, 2, \ldots, 10$.

Next, given a word w, we choose those ELS's $e = (n, d, k)$ with $|d| \geq 2$ that spell w with the following minimality constraint on $|d|$. Define the *domain of minimality* of e as the maximal segment T_e of G that includes e and does not include any other ELS $\hat{e} = (\hat{n}, \hat{d}, \hat{k})$ for w with $|\hat{d}| < |d|$. We define $\omega(e) \equiv \lambda(T_e)/\lambda(G)$, where $\lambda(T_e)$ is the length of T_e, and $\lambda(G)$ is the length of G. $\omega(e)$ is the "weight" we assign to e. For any two words w and w', we set $\Omega(w, w') \equiv \sum \omega(e)\sigma(e, e')$, where the sum is over all ELS's e with $|d| \leq D(\omega)$ spelling out w and over all SL's spelling out w'. We set $D(\omega)$ equal to that d for which the

expected number of ELS's for w will be 10 (for its computation see [4]).

$\Omega(w, w')$ is the unnormalized compactness measure of the PTB's for the word pair (w, w'). We normalize $\Omega(w, w')$ by introducing "(x, y, z)-perturbed ELS's" $(n, d, k)^{(x, y, z)}$ obtained by taking the positions (see [4]) $n, n + d, \ldots, n + (k - 4)d, n + (k - 3)d + x, n + (k - 2)d + x + y, n + (k - 2)d + x + y + z$, where x, y, and $z \in \{-2, -1, 0, 1, 2\}$ (thus there are $5 \times 5 \times 5 = 125$ such triples). The distance between $(n, d, k)^{(x, y, z)}$ and the SL $(n', \pm 1, k')$ is defined by the same formulae as in the non-perturbed case, where f is taken to be the distance between the first two letters of (x, y, z)-perturbed e. We may now calculate the "(x, y, z)-proximity" of two words w and w' in a manner analogous to that used for calculating the "ordinary" proximity $\Omega(w, w')$. This yields 125 numbers $\Omega^{(x, y, z)}(w, w')$, of which $\Omega(w, w') = \Omega^{(0,0,0)}(w, w')$ is one. We are interested in only some of these 125 numbers; namely, those corresponding to triples (x, y, z) for which there actually exist some (x, y, z)-perturbed ELS's in G for w (the other $\Omega^{(x, y, z)}(w, w')$ vanish). Denote by $M(w, w')$ the set of all such triples, and by $m(w, w')$ the number of its elements.

Suppose $(0, 0, 0) \in M(w, w')$. Denote by $\nu_1(w, w')$ the number of triples (x, y, z) in $M(w, w')$ for which $\Omega^{(x, y, z)}(w, w') > \Omega(w, w')$, and by $\nu_2(w, w')$ the number of triples (x, y, z) in $M(w, w')$ for which $\Omega^{(x, y, z)}(w, w') = \Omega(w, w')$. Then $\nu(w, w') \equiv \nu_1 + 1/2\nu_2$ is the number of triples (x, y, z) for which $\Omega^{(x, y, z)}(w, w')$ is considered to "exceed" $\Omega(w, w')$. If $m(w, w') \geq 10$, define $c(w, w') \equiv \nu(w, w')/m(w, w')$. If $(0, 0, 0) \notin M(w, w')$, or if $m(w, w') < 10$ (in which case we consider the accuracy of the method as insufficient), we do not define $c(w, w')$.

Thus, the corrected distance $c(w, w')$ is so normalized that the maximum distance is 1. A large $c(w, w')$ means that ELS's representing w are far away from the SL's representing w', on a scale determined by how far the *perturbed* ELS's for w are from the SL's for w'.

A.2. Details of Task (iii)

Here we describe the construction of samples S_1 and S_2 of pairs of expressions (w, w'), where w represent words appearing as ELS's, and w' represent words appearing as SL's. The words w are restricted in length to the range 5-8, as in [4] (this results from using (x, y, z)-perturbed ELS's for the normalization of the distance). However, there is no need to restrict the words w' appearing as SL's.

Dates of birth and death for personalities in Genesis may be found in several Talmudic sources, which occasionally conflict. Following the guidelines set up in [5] for a similar situation, we proceeded as follows:

1. Whenever several Talmudic sources (sometimes conflicting) exist for the desired data, we choose one coherent source, which gives more data than others.

2. We use the critical editions of the chosen source text (emended by comparison with manuscripts, etc.).

Here, too, we followed the same procedure:

1. We conducted a computer survey using a CD of the *Bar Ilan Responsa Database* [1] and scanned all the Talmudic period sources, searching for any statement which specifies explicitly the date of birth (or creation for Adam) or death of any of the personalities of Genesis. In this manner we found dates for several personalities - all of which were dates of birth (no explicit dates of death were given). It turned out that *Yalkut Shim'oni* is the most comprehensive of the sources. (For detailed results of the search see [3].)

2. We then examined a critical edition of *Yalkut Shim'oni* [2], and found dates for 14 personalities: Adam, Isaac, and all 12 of Jacob's sons.

List L consists of these 14 personalities. Sample S_1 is a set of word pairs (w, w'), where w' is the name of one of the personalities, and w is his date of birth. The names for these personalities include all of the variants occurring in the Book of Genesis. Each date is written in the three fixed forms of WRR.

Since we are dealing exclusively with dates of birth, it might be appropriate to examine only word pairs (w, w'), where w''s are the names of the personalities as they are spelled in the story of their birth. This subset of S_1 is sample S_2. Both samples are described in full in [3].

Acknowledgments

References

[1] Responsa project, CD 5. Bar Ilan University, Ramat Gan, 1997.

[2] R. S. Hadarshan. *Yalkut Shim'oni, Shemot*. Mosad Harav Kook, Jerusalem, heiman-lehrer-shiloni edition, 1973. Oxford manuscript with textual variants from manuscripts and early editions.

[3] D. Witztum. The hidden birth dates of personalities of Genesis: The data. http://www.torahcodes.co.il/genesis/personaldata.pdf, 2005.

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

[5] D. Witztum, E. Rips, and Y. Rosenberg. Equidistant letter sequences in the book of Genesis: II. the relationship to the text, preprint. 1995.