The Bible Code – Data Mining for God

David E. Thomas
New Mexicans for Science and Reason
P.O. Box 1017, Peralta, NM 87042 USA

Abstract

The "Bible Code" is claimed by believers to provide compelling mathematical and statistical evidence of a secret code engineered by God into the Bible, thus proving God's existence itself. Israeli mathematicians pioneered the Bible Code, which led to popular expositions of the Code by authors Michael Drosnin and Ed Sherman. While Bible Code messages appear compelling because of their low probability, the "miracles" are actually obtained by using the computer to exploit the billions and billions of possibilities. Indeed, impressive hidden messages can be derived from any given text. Two successful predictions of sporting events based on hidden codes in Tolstoy's book War and Peace are presented. Because "stunning" results can be obtained even when the search text is scrambled randomly, it becomes clear that the Bible Code is just an arcane method for harvesting the lush fields of chance.

Introduction

In recent decades, there has been a significant increase in the use of advanced methods from mathematics and computer science for various scientific-sounding "proofs" of the existence of the Judeo-Christian God, Allah, or various other deities. The "Bible Code" is perhaps the most famous of the new "scienticized" apologetics. These are the hallmarks of such statistical apologetics:

• The proof of the existence of the Deity is said to depend on a hidden or secret Code or Message implanted by the Deity, usually in the sacred text for the religion.
• This Code or Message, once explained, is claimed to be obvious and compelling, even though it's been hidden and unknown all these years.
• The evidence for the Code or Message is claimed to be statistically overwhelming, and beyond chance.
• The Code or Message is claimed to be too complex to have been created by mere humans, and thus proves the existence of a Higher Power.
• The Code or Message is said not to exist in mundane, non-religious works.
• Other Codes are dismissed as flawed, weak, or wrong. There can only be One True Code.
• Scientific or technical criticisms of the Codes are often dismissed by code apologists with personal attacks on the critic's integrity, or with accusations that the critic is mocking God (or Allah, or whoever).
• Presentations of Codes found in mundane texts are often dismissed as invalid for one reason or another.

The most famous recent example of mathematical apologetics is the "Bible Code." Several professional mathematicians have claimed that the Hebrew Torah contains a hidden code, verifiable with advanced statistical methods. Bible Code proponents point with pride to an article entitled "Equidistant Letter Sequences in the Book of Genesis," published in the respected journal Statistical Science in 1994 (Witztum 1994). The paper is said to provide compelling proof that details of modern rabbis and events are indeed encoded in the ancient symbols of the Torah. In June of 1997, a sensational book entitled The Bible Code, by journalist Michael Drosnin, brought the Bible Code into the mainstream (Drosnin 1997). The book did well on the New York Times bestseller list, and Oprah Winfrey gave it a solid endorsement on her television show.

I spent the summer of 1997 studying the Bible Code, and learned how to implement the algorithm, and apply it to various texts. My analyses on the Code appeared in the November/December 1997 and March/April 1998 issues of the Skeptical Inquirer (Thomas, 1997, 1998). I found that Bible Code puzzles can be found in any text, for any desired message.

What Is a "Bible Code"?

"Bible Codes" are hidden messages in the Bible composed of letters separated by the same distances between them (skips). These messages are called "Equidistant Letter Sequences," or ELS. Indeed, every one who has put the word "generalization" to print has secretly and unintentionally encoded the word "NAZI," as shown in Figure 1. The smaller the skip, the better, as far as code "quality" goes, so the skip of three for this match is truly spectacular by code standards. Normally, codes can extend across hundreds of thousands of letters.

Figure 1: An Equidistant Letter Sequence for "NAZI" occurs inside the word "GENERALIZATION."

The Bible Code is said to work only in the Hebrew Bible (Torah), and not in any other books or texts. However,

McKay et. al. demolished the results of the 1994 paper with a strong rebuttal in the 1999 Statistical Science (McKay 1999), showing the presence of "codes" in the Hebrew translation of War and Peace. I have looked for codes primarily in English texts, as this makes the secrets of "The Code" much more transparent. Additionally, the lack of vowels in Hebrew, along with its striking linguistic flexibilities, make it much easier to find Codes in Hebrew texts than in English ones. Nevertheless, stunning codes can be found in any book of choice.

The Secret of the "Code"

The chances of matching a given code word by pulling out that number of letters randomly from the text are small indeed, usually on the order of winning a multi-million dollar lottery. For an N-Letter Word, the Expected Number of ELS Matches from a Text is the Product of (A) the Probability of Matching a Word with N Random Letters and (B) the Number of Possible ELS Sequences. The Probability of Matching a Word, (A), is the product of the individual letter frequencies. If the given text has frequencies of 0.1%, 7.5%, 6.5%, and 7.5% for the letters J, O, H, and N respectively, the product of these frequencies yields the chances of picking "JOHN" by luck – about 4 in 10 million. Clearly, the chances are reduced by the presence of rare letters like J, or by the addition of more letters in the search word. For example, JOHNSON is harder to match than JOHN.
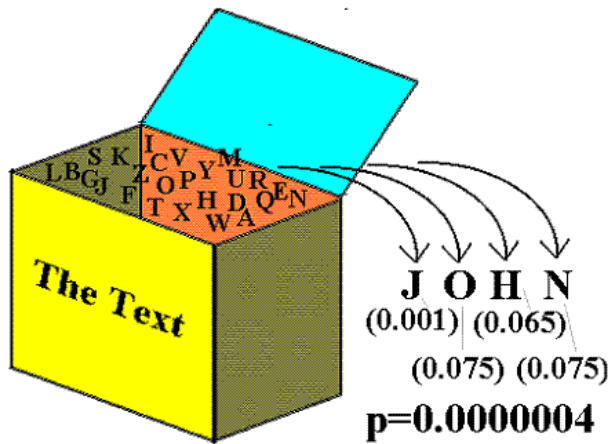

Figure 2: For N letters pulled at random, the chances of matching a word equal the product of the frequencies of the word's letters in the text.

The "secret" of the Code is the huge number of possible ELS locations which can be exploited. This number goes as the *square* of text length, as indicated in Figure 3, and is often quite huge, typically several billions or more. While the chances of any given ELS being a match are small – lottery sized – the computer gives you, for a few seconds' wait, literally billions and billions of chances to win. It's like getting billions of free, different lottery tickets – the question is no longer if you'll win, but rather how many times you'll succeed. And that number is just the product of

the small chance of matching an ELS with the huge number of possible positions.

$$\# \text{ Seq's} \cong \frac{\text{Ntext}^2}{(\text{Nword - 1})}$$

Figure 3: The Approximate Number of ELS for a Word of Nword letters in Length, and for a Text of Ntext letters in length.

Ranking Codes: "Minimality"

Sometimes too many "hits" are found to work with in practice. Therefore, Bible Code researchers developed the concept of "Minimality," essentially favoring the matches with the shortest skips. Only matches that have some area of the text for which they are the shortest-skip ELS can be considered as valid. While Drosnin gives lip service to minimality, he in fact ignored it while writing his book (Drosnin 1997), and even chose a match for "Clinton" that was invalid three times over, as displayed in Figure 4.
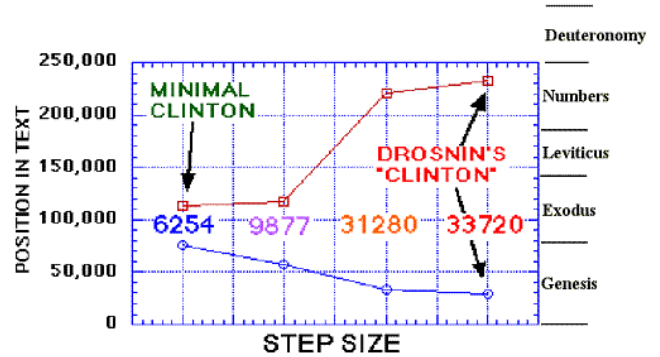

Figure 4: Of the Four CLINTON's in the Torah, Drosnin chose the Worst Possible, Made Invalid ("Non-Minimal") by the Presence of Three Inner Matches.

Drosnin's match for Yitzhak Rabin is the only one in the entire Torah, and is thus automatically Minimal. Yet, so is the GREGHINES code of War and Peace (Books 1 & 2). These two matches, each unique and minimal over the complete text, are compared in Table 1.

Table 1: Comparing Matches for "Yitzhak Rabin" in the Torah to "Greg Hines" in War and Peace

	"Yitzhak Rabin"	"Greg Hines"
Probability of One Match	7.5×10^{-12}	8.77×10^{-12}
Number of Possible ELS	13.3 Billion	5.6 Billion
Expected Number of Matches	0.0991	0.0495

When the expected number of matches is more than one, the interpretation is easy. For Table 1, however, the one-in-ten for Rabin means that one could match it on one of ten typical Torahs; likewise, the one-in-twenty number for Greg Hines means that rare match is found in only one of twenty versions of War and Peace, on average. While not a certainty, getting matches for such names is not that hard, even at one in twenty chances.

Linked Messages

Drosnin also claimed that it was impossible to find *linked* pairs of words in mundane texts, and specifically claimed that one could not find "Hitler" and "Nazi" linked closely in Tolstoy's <u>War and Peace</u>. Yet I found a single 209-word-long passage in Tolstoy's book that held a stunning "Hitler/Nazi" Bible Code display (Figure 5).

Figure 5: HITLER/NAZI Paired in War and Peace.

I found "Roswell" inside just one verse of the King James version of Genesis, accompanied by a "UFO." This particular find was outstanding because (a) the English text could not possibly contain the Hebrew Code, and (b) the step of just 4 letters for the Roswell match is spectacularly "minimal" by Bible Code standards (Figure 6). The match appears completely inside Genesis 31:28: "And hast not suffered me to kiss my sons and my daughters? thou hast now done foolishly in so doing."

Figure 6: ROSWELL/UFO Paired in Genesis (King James).

Perhaps my most elaborate puzzle to date was this one, also from War and Peace: "Guilty Lee Oswald shot Kennedy, Both Died."

Predicting the Future

One of Drosnin's strongest claims in support of the Bible Code was that his code-based prediction of the assassination of Israeli prime minister Yitzhak Rabin was made a year before the event took place. So, almost two months before the 1998 NBA basketball playoffs, I searched for NBA teams in War and Peace, and found an excellent Bible Code message for CHICAGO, BULLS and JORDAN. I rushed my prediction to several reporters and scientists, but didn't warn Karl "the Mailman" Malone of the Utah Jazz, preferring to let history run its course. Chicago almost lost it all, but Tolstoy's Bulls clung to their destiny. And on June 14th, 1998, my prediction came to pass.

As luck would have it, The Amazing Meeting 4 (TAM4) was held just one week prior to Super Bowl XL. Therefore, prior to the NFL playoffs, I searched for the four finalist teams in Books 1 and 2 of <u>War and Peace</u> (about 400,000 characters). The results for expected and actual matches are given in Table 2. The key indicator is the "spread" (expected matches minus actual matches).

Table 2. Superbowl XL War & Peace ELS Matches

	Steelers	Broncos	Panthers	Seahawks
Expected	33	13	6	1
Actual	46	17	4	1
Spread	-13	-4	+2	0
	Pittsburgh	Denver	Carolina	Seattle
Expected	0	1034	3	602
Actual	0	1009	2	609
Spread	0	+25	+1	-7

On the day of my talk, January 29th, 2006, I explained that the spread had favored the Steelers over the Broncos, and the Seahawks over the Panthers. In a week hence, I told the TAM4 attendees, the Steelers would prevail over the Seahawks. And on February 5th, Tolstoy's Steelers beat the spread, becoming the second successful "Tolstoy Code" prediction of a future event.

Recent Developments

I was in a televised 10-minute "debate" with code proponent Ed Sherman of "Bible Code Digest" (BCD), aired June 4, 2005 on PAX TV's "Faith Under Fire." When I examined Sherman's website, biblecodedigest.com, in the weeks before the debate, I learned that he has two main arguments in favor of intentional hidden codes in the Bible.

One of Sherman's arguments is that only the Bible has large clusters of related words coded in small sections of the actual text, and the other is that improbable codes dozens and dozens of letters long are found only in the Bible. The latter claim is easily dismissed. When I had two respected Hebraic scholars look at Sherman's long, tortured "sentences," they were appalled at their quality as prose. Even Sherman himself grades his long codes as Good, Acceptable and Marginal. That is, some of Sherman's "codes" are so poorly slung together that only a God willing

to publish "Marginal" work would release them to the public. What makes these long passages even marginally possible is the incredible flexibility of Hebrew. Without vowels, the same few letters can have a wide variety of meanings. Sherman himself admits, deep in the bowels of a posting he's sure most website visitors will never read, that "significant variation in the 'translation' of any given letter string between different Hebrew experts should be expected." (Sherman 2003b) In his latest response, however, Sherman has written that the experts I consulted (scientist Shlomo Karni and Rabbi Joseph Black) probably only know 20% of the Hebrew that his expert, Dr. Nathan Jacobi, knows. (Sherman 2005)

But the main bones of contention are Sherman's claims of large "clusters" of hidden matches. When I was preparing for the televised debate, I came across this challenge in Sherman's article "Comparing Apples and Oranges": "If the code has six or fewer letters, it is virtually certain that you will find it somewhere in the Bible, or even any Hebrew book. But then, if the code has 15 or more letters, the odds of finding it in any randomly selected, **two-page-long section** in the Hebrew Bible are virtually zero." (Sherman 2003a) I have since found out, from Sherman himself, that he wasn't really talking about searching for codes in a two-page-long section of text. What Sherman does, it turns out, is quite different - he searches the entire Bible for code matches, and then selects those which happen to have at least one letter passing through a two-page-long section of the Bible. Sherman also searches for extensions to given codes, which is not that hard given the flexibility of Hebrew.

By taking Sherman at his word - by restricting my search to just a two-page section of Tolstoy's War and Peace, and to matches of more than six letters, I gave myself a huge handicap compared to Sherman's approach. At the time, I thought the challenge was actually to find messages within just a two-page section - in my case, an 8,968-letter snippet that I already knew contained a code for "Harrison." Even so, I found 253 words of seven or more letters each, including several clearly related to the late Beatle, George Harrison: ARTISTE, CHEERED, COMPOSED, DREAMER, ELEANOR (Rigby), GOATEED, ONENESS, TSUNAMI, WITTIEST, and many more. (I thought Sherman might argue about TSUNAMI, but the rock benefit concerts put on to aid victims of the 2004 tsunami in Indonesia continued a trend started by George Harrison, with the 1971 Concert for Bangladesh.)

A few months after the televised debate, Sherman posted an article titled "Want Loads of Fish? Drain the Lake!" on his website (Sherman 2005). In this article, Sherman claims that what I found in War and Peace is like "draining the lake" (the text of the book) to find a few 7-to-9-inch long minnows (e.g. 7-to-9 letter matches), while what he does is like throwing a fishing line into the lake, and pulling out a 3-foot-long sturgeon. In other words, skeptics have to drain an entire text for a few small matches, whereas the Bible gives you long matches almost every time you dip your line (perform a search).

But something wasn't right. I started to investigate some of the matches Sherman claimed for his extensive "Ezekiel cluster." And I realized that many of these codes were NOT contained in the short section of Ezekiel 36-38, but rather spanned tens or hundreds of thousands of letters of the Bible, with only one or a few letters actually passing by the Ezekiel text. In one shocking "million-letter-match" case, I investigated one 7-letter match for "Saddam Hussein" that Sherman claimed for his "Ezekiel cluster." However, this match, with a skip of 150,864 letters, spans over 900 thousand letters of the actual Bible - almost a million letters!

When I employ the BCD method as it is actually used by Sherman, finding codes is *easy*. It's like playing T-Ball after trying to get hits from major league pitchers.

Here are some of the truly amazing new "Harrison Cluster" codes I've recently found in War and Peace. These were found by searching two books of War and Peace (just under 400,000 letters worth), and counting those that *touch down* in the two-page section for the "Harrison Cluster" discussed previously.

> Let It Be,
> Fab Four,
> Beatles,
> Sweet Lord,
> Yoko Ono,
> Sixties,
> Satsang,
> I Me Mine,
> Soloist,
> Rock God,
> Mop Tops,
> Magical,
> Clapton,
> Shankar,
> Piggies,

and many more directly relevant to the late Beatle.

Drosnin's Sequel

Michael Drosnin's sequel to the Bible Code, called the Bible Code II: The Countdown (Drosnin 2002), was released in November 2002. Amazon.com kindly placed an excerpt from the new book on the internet, which I immediately searched for hidden messages (Thomas 2003).

And when I applied the "Bible Code" technique to this excerpt (a mere 6,966 characters, for 1,617 words), I found this AMAZING secret message in Drosnin's text:

"THE BIBLE CODE IS A SILLY, DUMB, FAKE, FALSE, EVIL, NASTY, DISMAL FRAUD AND SNAKE-OIL HOAX.".

Conclusion

The Bible Code is simply an arcane technique that allows one to harvest any desired message, from any given text.

The messages have no "author," other than the code enthusiast who can find them.

The Code works best with random letter distributions.

In other words - "Seek, and ye shall find." (Matthew 7)

Acknowledgments

References

Drosnin, M., *The Bible Code* (1997), Simon & Schuster, NY.

Drosnin, M., *The Bible Code II: The Countdown* (2002), Viking Adult, NY.

Mckay, B. & Bar-Natan, D. & Bar-Hillel, M. & Kalai, G. (1999) *Solving The Bible Code Puzzle*. *Statistical Science*14-2 150-173.

Thomas, D. E. (1997) *Hidden Messages and the Bible Code*, *Skeptical Inquirer*, Nov./Dec. 1997, 21(6)30-36

Thomas, D. E. (1998) *Bible-Code Developments*, *Skeptical Inquirer,* March/April 1998, 22(2)57-58

Thomas, D. E. (2003) *It's Ba-a-ack! The Bible Code II*, , *Skeptical Inquirer,* March/April 2003, 27(2) 59-60.

Sherman, Richard E (2003a), "*Comparing Apples and Oranges,*" *Isaac Newton Bible Code Research Society*, http://www.biblecodedigest.com/page.php/102.

Sherman, Richard E (2003b), "*Non-Random Equidistant Letter Sequence Extensions in Ezekiel,*" *Isaac Newton Bible Code Research Society*, July 15, 2003 http://www.biblecodedigest.com/page.php/188.

Sherman, Richard E (2005), "*Want Loads of Fish? Drain the Lake!,*" *Isaac Newton Bible Code Research Society*, http://www.biblecodedigest.com/page.php/332.

Witztum, D., Rips. E. and Rosenberg, Y. (1994) *Equidistant Letter Sequences in the Book of Genesis*. *Statistical Science*, 9, No 3, 429 - 438.