Journal of Arts & Humanities

Volume 10, Issue 08, 2021: 22-36
Article Received: 24-06-2021
Accepted: 29-08-2021
Available Online: 16-09-2021
ISSN: 2167-9045 (Print), 2167-9053 (Online)
DOI: https://doi.org/10.18533/jah.v10i08.2136

Was Hebrew invented?

Marc Egeth[1,2]

ABSTRACT

In this paper, I present a straightforward secular (non-religious, non-theological) theory: the Hebrew in the Torah is an invented language, akin to Klingon or Elvish, that was designed to write the Torah. Under this theory, the letters, words, grammar, and key terms and phrases of the Biblical language were selected, co-determined, and creatively bootstrapped as a sort of language game. The invented language may have been designed, like Esperanto or Basic English, to include elements in common with known natural languages; nevertheless, the details and final form of the language were bespoke and under the control of a creative author. In the invented system, some anagrams were defined to mean related concepts; letters were assigned number values so words and phrases could "add up to" interesting numbers ("Gematria"); and the author planned a text that would include a density of word-games like pangrams (sentences that use all letters of the alphabet), lipograms (sections of the text that are missing one particular letter), and particular numbers of letters and words throughout. However, all signs of seemingly deliberate language games might instead merely be the result of cherry picking of interesting findings. In this paper, I present analysis and new approaches that might help distinguish whether the apparent language games in the Torah imply an intentional invented language or, instead, only result from over-analysis and were not intended by the author to be present in the text.

Keywords: Hebrew, Torah, Gematria, conlang, cognitive anthropology.
JEL codes: C11, C12, C18, C90, C99.

1. Introduction

The book *Sefer Yetzirah*, dated by various sources from between 3000 BCE and 1000 CE, describes the deliberate generation of the Hebrew language lexicon by systematically combining and permuting letters to make words. In this way, according to *Sefer Yetzirah*, anagrams could be intentionally set to represent related meanings, giving the example that the letter string "O-N-G," *oneg*, connotes pleasantness while its anagram "N-G-O," *negah*, connotes unpleasantness (Kaplan 1990, 85). Systematically combining letter sequences to generate a lexicon and assigning meanings to the resulting letter strings is a process not uncommon in the contemporary hobby of inventing languages (see e.g.

1 University of Pennsylvania, Department of Psychology (where work was conceived)
2 Exponent, Inc., Human Factors, Philadelphia, PA (affiliation at time of publication)

Kleffner, 2004), but, of course, this would not be part of natural human language evolution and development.

Along similar lines, Belaga (2013) notes that uniformly-three-letter-long Hebrew word roots seem too "parsimonious" to have developed naturally, and proceeds to speculate that Hebrew's proto-Semitic parent language was an invented language. John Wilkins, working in the 1600s, based his own invented language on Hebrew for a similar reason: "Hebrew is the best Pattern, because that language consists of the fewest Radicals." (Halperin, 2012, p. 5). Whether or not Hebrew represents an invented language, its intrinsic properties have made it at least appear like one to some.

The Talmud represents that letters in the Torah are intended to be interpretable as numerals, thus allowing numerical interpretations of the Biblical Hebrew's words (e.g. Nedarim 32a); this is called Gematria. If so, then the vocabulary of Biblical Hebrew is not merely a veridical phonetic representation of naturally spoken utterances, but, instead, during its authorship, its words were crafted or selected to include numerically-relevant letters.

The sentence- and story-levels of the Torah text have also been proposed to play similar linguistic games. The Torah includes two pangrams, meaning sentences that include every letter of the alphabet. And Ginsburgh (2015) notes that the Torah's verses which, on the surface, describe 13 attributes of G-d themselves include a total of 169 (13x13) letters.

People have invented thousands of languages including for fun, for philosophy, for attempts at human unification, and for science fiction aliens to speak. People also seem to find satisfaction contemplating square numbers, triangle numbers, etc, including contemporary recreational mathematicians as well as in the ancient world (Deza and Deza 2012). These creative activities are recurring, cross-cultural, ancient, and modern human pastimes.

- To develop the 19th century invented language *Solresol*, its author François Sudre systematically combined basic syllable units and determined that anagrams would represent related meanings. (The basic units were notes, so the language could be played musically! The name of the language is a syllable-level musical palindrome.)
- In 1981, Luigi Serafine first published *Codex Seraphianus* written entirely in an artificial language invented for the purpose of writing the book.
- The ancient poet Pingala wrote Sanskrit poetry in which the patterns of syllables correspond to the Fibonacci sequence (Hall, 2005).

Demonstrations such as these help make clear that it would certainly be possible for humans to build a lexicon and write a text in which the number value of words and verses are meaningful, if that is something someone wanted to do. An author who is so inclined could well write about 13 attributes of G-d and deliberately choose to do so using a square number of letters. (I edited this very sentence until it had exactly a square number of letters; it was a good challenge.)

In spite of possibly being essentially correct, *Sefer Yetzirah's* and the Talmud's characterizations seem only to have minimally entered the modern discourse around the study of invented languages. Various modern reviews of invented languages fail to mention *Sefer Yetzirah's* claims, whether to agree, refute, or provide further analysis (e.g. Okrent, 2009; Peterson, 2015). Some modern authors discuss well-known deliberate designs of modern Hebrew vocabulary in the context of invented languages, while not considering the provenance of the Biblical language itself. For example, Halperin (2012) reviews historical perspectives on Biblical Hebrew as a divine language, as well as conscious efforts to revive Hebrew as a modern spoken construction. Higley (2007) describes that Corneilius Agrippa and John Dee working in the 1500s invented languages that were informed by Hebrew and by their reading of *Sefer Yetzirah* specifically (pp. 59-68), but Halperin does not inquire whether the Hebrew, the basis of their inspiration, was itself invented. This leaves us little relevant contemporary literature to review around the proposition that Hebrew represents an invented language.

Therefore, let this paper represent an overdue but standard modern analytic process. The specific form of theory that this paper will put forward is that the Torah was written with a density of word, letter, and number games in its own invented language. I will also describe an alternative theory, which is that the text itself includes no intrinsic linguistic inventiveness, but rather that over-analysis has led to an illusory (i.e., false) appearance of letter games and intentional linguistic design.

Hebrew may be similar in various ways to other ancient near-Eastern languages, but mere similarity to existing languages is not diagnostic of naturalness, because the difference between an invented language and its natural pre-cursors can be arbitrarily small. Star Trek's Klingon is quite different from contemporary human natural languages, but Basic English was designed to be English just without any irregular nouns, plus additional simplifications. Esperanto was also, intentionally, designed to be similar to existing contemporary languages in various aspects.

The best reason to positively expect that the Torah represents a natural language may be simply the argument from base rates: most texts today are written in natural languages, and it is only the rare exception that a text is composed in an invented language. For example, in 1981, there were perhaps one million other books besides *Codex Seraphianus* that were published in whatever languages their authors naturally spoke. Therefore, one might reasonably expect that any particular text, knowing nothing else about it, is probably, though not definitely, written in a natural language.

From this initial position (i.e., 'any particular text, knowing nothing else about it, is probably, though not definitely, written in a natural language,'), further evaluation can proceed by considering specific evidence or reason that might move the needle one way or another, so the field can gain either more certainty that the language was invented or more certainty that the text represents a written version of an oral natural language.

After describing a set of linguistic game-like phenomena that may seem to suggest that the Torah's language was invented, then I will discuss new methods to assess whether the apparent language games are intentional versus accidental.

2. Literature review

2.1 Triliteral roots

As descried above, *Sefer Yetzirah*, John Wilkins, and Belaga (2013) each describe a "parsimony" in the Hebrew roots. Biblical Hebrew grammar has been described as a system of "triliteral roots," or sets of three letters that can be modified to represent verbs in different tenses. For example, to make a verb in the second person singular perfect tense, you take the three-letter root and add a one-letter suffix (the letter "Tav"), or for the first person singular you take the three-letter root and add a one-letter prefix (the letter "Aleph"). The shortest case is the third person masculine singular, in which you add no prefix or suffix and keep just the three root letters. Some analysts argue for occasional roots of other lengths, but e.g. Horowitz (2020) provides a list of verbs in the Torah, and all 609 on the list have roots of exactly three letters.

Most often, a long list of exactly-three-long sets of letters represents formal codes, like file name extensions (.exe, .bat) or airport codes (PHL, BWI). Therefore, a reasonable expectation from base rates might be that if you have a list of three-letter-long strings, you are looking at an artificial formal system, rather than a veridical phonetic transcription of naturally spoken utterances such as *third-person-singular masculine verbs*. This is why Belega (2013) hypothesizes inventedness on the basis of the three-letter roots: "… this unique linguistic phenomenon should arise [sic] today one's scientific curiosity – be it just because of the striking similitude of the abstract perfection and parsimoniousness of such an alphabetical coding of verbs to the way machine codes ... are traditionally represented - by mostly three Latin letters combinations (abbreviations), with a very few codes having two- and four-, or more-letter names."

Perhaps, like Belaga, we might wonder whether it could really be that every third person feminine singular verb that real people spoke aloud in a real human society just happened to require exactly four letters in an alphabet to phonetically transcribe it. To my knowledge, no known spoken languages work that way when moving from spoken to written formats, but instead word lengths are expected to have distributions around some central tendency of length, as do most natural things when length is measured (see e.g., Wimmer, Witkovsky, and Altmann 2010; Grzybek 2007). This apparent discrepancy between how language usually works and how Hebrew works would account for Belega's (2013) apparent surprise at the Torah's Hebrew's appearance of unnatural over-parsimony.

Belaga (2013) also notes "...the pervasiveness of the phenomenon of topologically neighboring (for example, differentiating in only one letter position) verbs having semantically meaningful correlations...." This sounds very much like Sefer Yetzirah's claim of correlated meanings among

anagrams: meanings were assigned to specific letter-strings and related strings were given related meanings.

2.2 Pangrams and lipograms

A pangram is a sentence that includes all the letters of an alphabet.

The Torah includes (at least) three pangrams of note: Exodus 16:16, Deuteronomy 4:34, and the Ten Commandments in Deuteronomy, which technically is a "pangrammatic window" because the Ten Commandments considered as a set are not a single sentence *per se*.

In English, there are no stories in which it just happens to be the case that, after a series of events, "A quick red fox jumps over a lazy brown dog," a well-known pangram. In English, if a story includes a pangram, then the story pretty much needed to have been engineered to be able to use that key sentence in a meaningful way.

Accidental pangrams can happen, but they are not frequent. For example, the wordsmith.org online pangram finder seems to find no pangrams in either the English translation of War and Peace or in the King James Bible (in English). This leads to an inference from base rates again: to a first approximation, if you are looking at a pangram, it was crafted intentionally. The existence of pangrams in a text does not itself require or imply that the language was invented, but might alert you that perhaps you are reading the work of a linguistically creative author, a "Lewis Carroll" or "Martin Gardner" sort (see e.g. Lemos, 2009).

It could be hard to tell if a given pangram is necessarily intentional. Maybe, one might wonder, accidental or "natural" pangrams are more likely in Hebrew than in other languages. If interested, one could examine the frequency of pangrams in contemporary Hebrew literature or newspapers. However, before one gets to counting letters in newspapers, also note that it is has been pointed out as a conspicuous omission that the Hebrew letter "Samech" does not appear in the first chapter of Genesis. A stretch of text with a missing letter is called a "lipogram." To the extent that pangrams become easier to explain due to letter frequency homogeneity, lipograms in the same text become harder to explain. That is to say, if, for whatever reason, all Hebrew letters are each fairly likely to appear regularly, thereby making pangrams relatively unsurprising, then a long stretch of Hebrew text that is missing a particular letter would look correspondingly more like a deliberate letter game.

The sort of person who writes a story that includes a pangram is also the sort of person who writes a story that includes a lipogram. For example, see *Ella Minnow Pea* by Mark Dunn (2001), a novel about pangrams that is written using missing letter games. Pangrams and lipograms should be anti-correlated from the perspective of letter frequency: if you have a language with many of one then you should not have the other. But, the presence of pangrams and lipograms should be positively correlated from the perspective of deliberate authorship, because if an author is "into" one of them, then the author probably has a propensity also to be "into" the other.

2.3 Numbers

The Talmud, mostly compiled by 1500 YA, describes that Hebrew letters can be meaningfully interpreted as numerals, or "Gematria," in which letters have the number-values from 1 to 400 as shown in Table 1.

In Table 1, the "3s" column is highlighted.

Table 1.

Gematria values of Hebrew letters, 1 (upper-right Aleph) to 400 (lower-left Tav).

1s	ט	ח	ז	ו	ה	ד	ג	ב	א
10s	צ	פ	ע	ס	נ	מ	ל	כ	י
100s						ת	ש	ר	ק
-	9	8	7	6	5	4	3	2	1

Maybe I am imagining it, but these highlighted letters, indicating 3, 30, and 300, respectively, each look like a good shape for a numeral representing "threes" - as does the Arabic numeral "3," also highlighted, for that matter. The "2s" as well look like good "twos," and so do some of the "4s." The Arabic "1" also stands out as a (very) good "one." If so, then also Hebrew "Yod," 10, is also good form to

represent 10 in what is a base-10 system, like a "tick mark." If they were intended to be numerals as described in the Talmud, the Hebrew letters do potentially seem to have decent shapes.

Beyond the shapes of the letters/numerals, some Hebrew words may seem to reflect the numeral-letters that comprise them:

- The root meaning "three," שלש, draws on three of the three-looking characters from Table 1.
- The root for "six" is שש.
- The word meaning year, שנה, adds up to 355, which is one of the astronomical options for the number of days in a lunar year.

Design letter forms to also visually represent numbers; select number-words based on the shapes and number-values of letters; select letter strings with relevant Gematria values to represent words (like "year") with relevant number components - these sorts of games would be available to a creative author designing a language.

2.3.1 Fibonacci sequence in the milui of the letter Aleph

Some basic properties or sequences of numbers can draw cross-cultural attention across millennia. The Fibonacci sequence is one of these, which is the set of numbers that results when starting "1, 1, 2, 3, 5, 8, 13…" and continuing to add the last two numbers in the sequence to get the next number. The ratio of succeeding numbers converges on the famous "golden ratio," about 1.681:1.

The first Hebrew letter, Aleph, contains the entire infinite Fibonacci sequence in the recursive spelling of the component letters of its name (see Daf Yomi Review,2020a). This expansion of a word by spelling out the names of its composite letters is called *milui*. The number of letters it takes to spell the names of the letters in the word "Aleph" are Fibonacci numbers. And then, recursively, the number of letters it takes to spell out those letter names are, too – and so on (see Table 2).

Table 2.

How many letters does it take to spell out the names of the letters that spell the name of the first letter, Aleph – and then to continue to spell out the names of those letters?

1	Aleph	1	Lamed	1	Pe
3	Aleph Lamed Pe	3	Lamed Mem Dalet	2	Pe Aleph
8	(Aleph Lamed Pe) (Lamed Mem Dalet) (Pe Aleph)	8	(Lamed Mem Dalet) (Mem Mem) (Dalet Lamed Tav)	5	(Pe Aleph) (Aleph Lamed Pe)
21	[(Aleph Lamed Pe) (Lamed Mem Dalet) (Pe Aleph)] [(Lamed Mem Dalet) (Mem Mem) (Dalet Lamed Tav)] [(Pe Aleph) (Aleph Lamed Pe)]	21	[(Lamed Mem Dalet) (Mem Mem) (Dalet Lamed Tav)] [(Mem Mem) (Mem Mem)] [(Dalet Lamed Tav) (Lamed Mem Dalet) (Tav Aleph Vav)]	13	[(Pe Aleph) (Aleph Lamed Pe)] [(Aleph Lamed Pe) (Lamed Mem Dalet) (Pe Aleph)]
… and so on forever, generating only Fibonacci numbers and all Fibonacci numbers (Daf Yomi review 2020a)					

It looks like the Hebrew alphabet may play this "trick" by naming letters recursively, including making letter-names particular lengths, and making sure each letter includes itself in its spelling and exactly one other letter that includes the original letter:

- Aleph, when spelled, includes itself, as well as including "Pe" which when spelled uses an Aleph.
- Lamed, when spelled, includes itself, as well as including "Dalet" which when spelled uses a Lamed.
- Pe, when spelled, includes itself, as well as including "Aleph," which when spelled uses a Pe.

The word or name "Adam," representing humanity, returns only Fibonacci numbers in a similar fashion (oren evron, 2015). This key word uses only letters that are included within the recursive spelling of Aleph: Aleph, Dalet, and Mem.

Like square numbers and other figurate numbers, the Fibonacci sequence was known before the common era (Leyendekkers and Shannon 2013; Singh 1985). As mentioned previously, Pingala is interpreted as having written Sanskrit poetry in 300 BC with short and long syllables representing numbers in binary and as having noted that the number of different prosody combinations for sentences of different lengths would follow the Fibonacci sequence (Hall 2005). It would be available to the ancient author to include the Fibonacci sequence within the names of letters.

Any modern reader is free to try to replicate such patterns in a model system of their own and get a feel for the sort of thought process implied: creativity and intelligent focused intention are certainly required, but there is no requirement for modern knowledge nor technology.

Further note the following about Aleph-Lamed-Pe:

- The Aleph has the Gematria value of 1.
- The Gematria value of "Aleph-Lamed-Pe" is 111.
- The word Aleph-Lamed-Pe means "1000" in the surface level of the Torah text (pronounced "eleph")

In these ways, the letter Aleph – the first letter in the alphabet – keeps turning up "1s." And as previously mentioned, by the recursive naming of letters described above, the Aleph also follows the Fibonacci sequence to infinity. It does not feel like a stretch to connect this symbolism with the surface-level monotheistic philosophy of the text: one-ness and infinity.

2.3.2 Figurate numbers

Applying Gematria to the text of the Torah appears to result in a theme of *figurate numbers*, or numbers that can be represented by symmetrical geometric shapes. In particular, the numbers that result from that process can be represented by squares/triangles/Jewish stars, both 2D and 3D.

The first sentence (verse) of the Torah includes an early textual instance of this theme. Its Gematria value is 2701, and 2701 points can be arranged as a symmetrical star made of symmetrical stars (Bradford 2008, 71). To aid the visualization, Figure 1 depicts the letters of Bereshit their "Gematria" number of times: for every letter "Bet," there are 2 Bets, for every letter "Tav," there are 400 Tavs, etc.

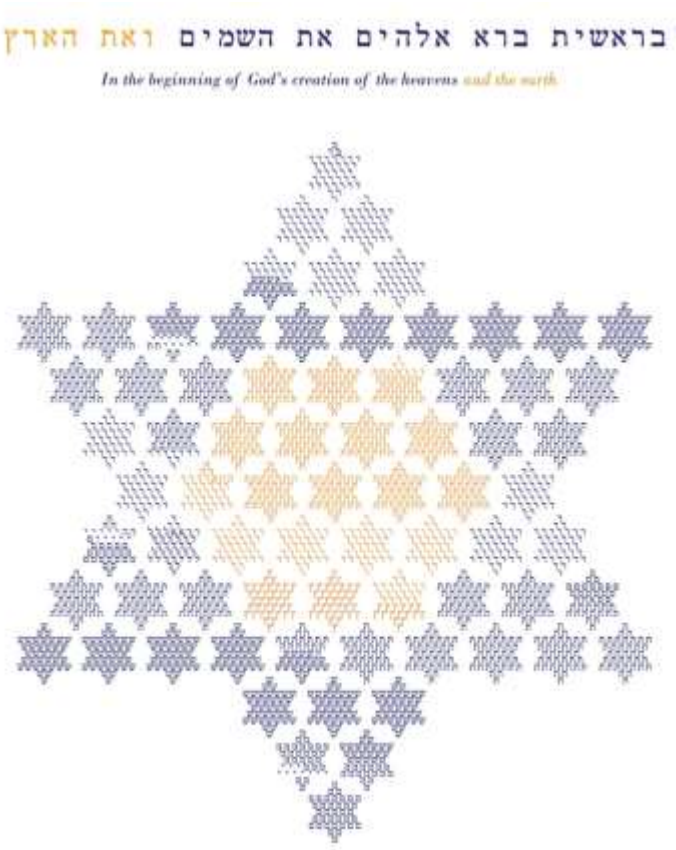

Figure 1. A star-of-stars in Bereshit's Gematria. By Davinica Nemtzow, with permission.

This coded shape reappears repeatedly around the first verse:

- A phrase in the next sentence (Genesis 1:2), "And the spirit of G-d moved upon the face of the waters," has a Gematria value of 1369, a 37X37 also a star-of-stars figurate number.
- Gematria of the *milui* – the spelling out of the letter names – of the first word of the first verse, "Bereshit" has a star-of-stars shape (See Figure 4; Daf Yomi Review 2020b).
- Gematria of the *milui* of the last two words of the first verse (the red letters in Figure 1) is 2701, the same star-of-stars shape as the whole of Figure 1 (Daf Yomi review 2020c)
- Gematria of last word of the Torah, "Israel" – which is proximate to the first verse, in a scroll – is a star shape (see Figure 2).

• Gematria of the *milui* of the last word of Torah is also 2701, the same star-of-stars shape in Figure 1 (Daf Yomi Review 2020d)

Figure 2 further shows how the names of other individuals in Israel's family also suggest similar geometric themes.

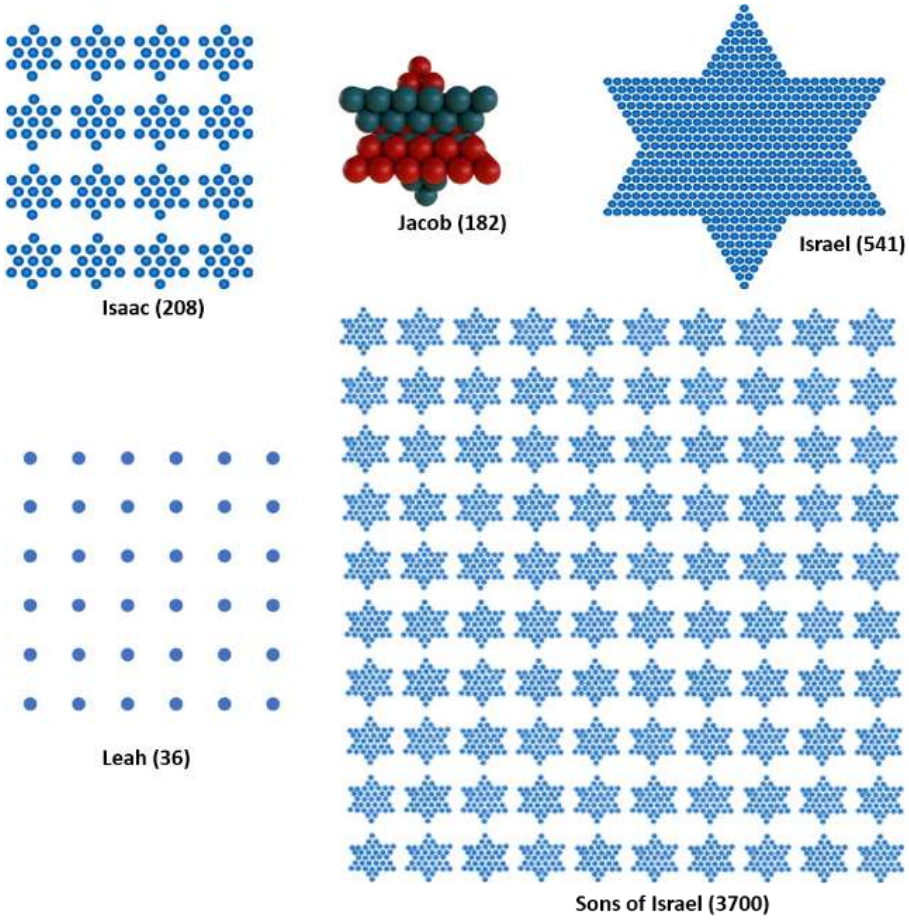

Figure 2. Geometric theme of squares and stars found around Israel. In rows from upper left; all numbers Gematria: Isaac, Israel's father, a square of stars (208). Israel's birth name, Jacob, two intersecting square-base pyramids each of 91 points (182). Israel, a star (541). Leah, Israel's spouse, a square (36). Israel's sons as listed on the breastplate of judgment, a square of stars, 3700; the surface level of the text describes the breastplate as a square.

Key terms in the philosophy and theology of the Torah further embody similar star shapes, including the Gematria of the words for "love" (*ahava*), "one" (*echad*, which begins with an Aleph), and the root of the word *shalom*. As described previously, the list of 13 attributes of G-d's mercy around Exodus 34:6 uses 169 letters (13x13; see Ginsburgh, 2015); while this could be a square, 169 can represent a star of stars equally well, as shown in Figure 3.

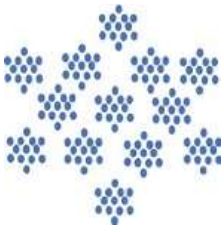

Figure 3. A 13-star of 13-stars, one point for each letter in the verses describing 13 attributes of G-d.

2.3.3 Span of the text

If the start and end of the text are inherently salient textual locations for a creative author to use to play games, then so would be phenomena that span the entirety of the text. Across the entire Torah, the four-letter name of G-d appears 1820 times, a number that makes this same sort of shape, a 3-D star made of 3-D stars, as the figurate shape in "Jacob" listed within Figure 2. See Figure 4.

All known versions of the Torah include this same number of instances of the name of G-d (1820; see Drazin, 2009 for a review of proposed copying errors in Torah transmission).

Finally for this section about figurate numbers, the total number of spaces (i.e., spaces between the letters in the Torah), 304,804, can also be depicted as this same sort of thematic shape, in this case a star made of squares as shown in Figure 5 below.

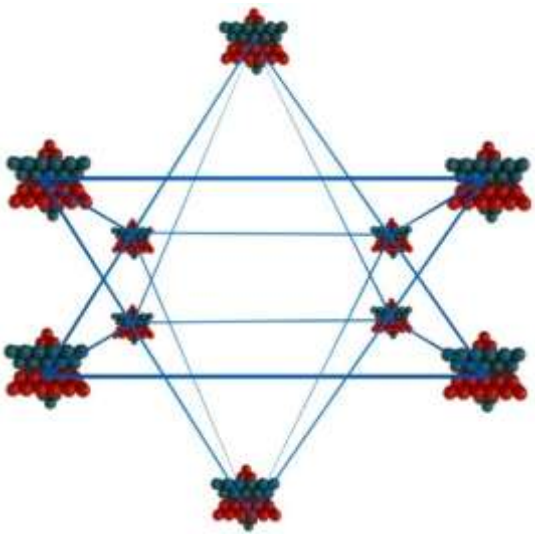

Figure 4. 1820 points as a 3D star of 3D stars, the number of times the four-letter name of G-d appears in the Torah and the Gematria of the milui (letter names) of the first word of Torah, Bereshit.

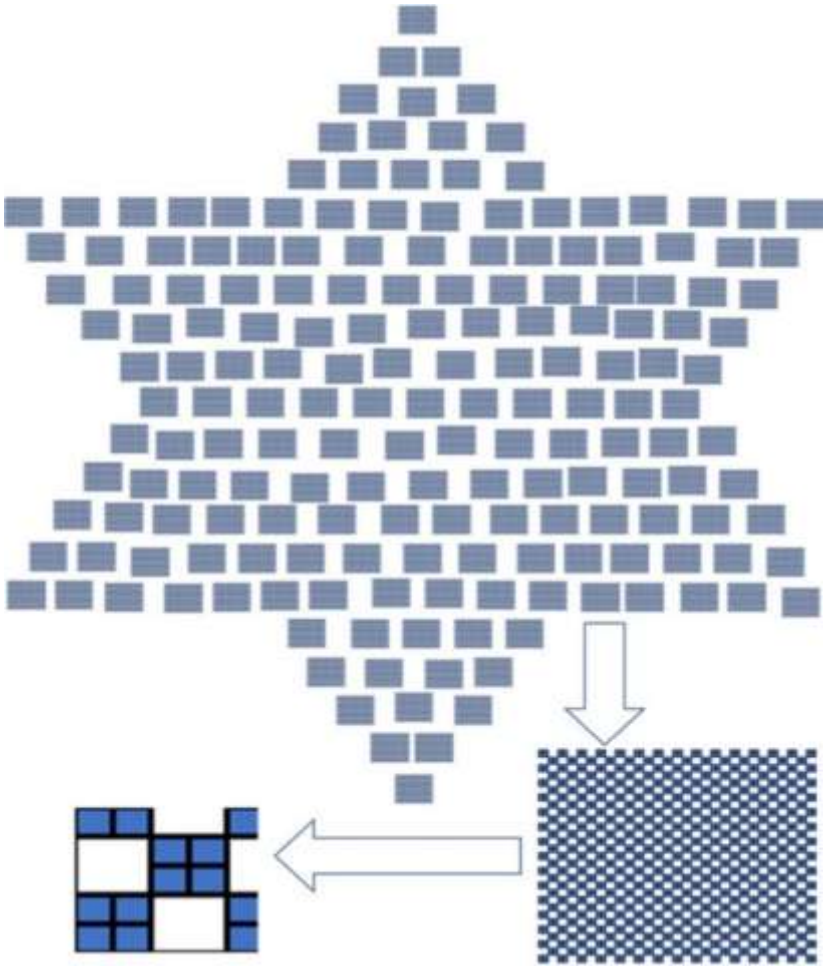

Figure 5. 304,804 spaces between letters in the Torah: a star comprised of 181 squares each comprised of 421 squares each comprised of a square of 4 little blue spaces, one for each space between letters in the text.

Alternatively, this same figurate number phenomenon can be represented for the total number of letters in the text (instead of for spaces between letters), with one letter separate and then the rest of all the letters of the text arrangeable into the same shape as in Figure 5. Artistically, Figure 6 shows a large letter Bet, and the rest of the letters can then be arranged as the "dagesh," or dot, of the Bet. (Torah scrolls are traditionally written with a large initial letter "Bet.")

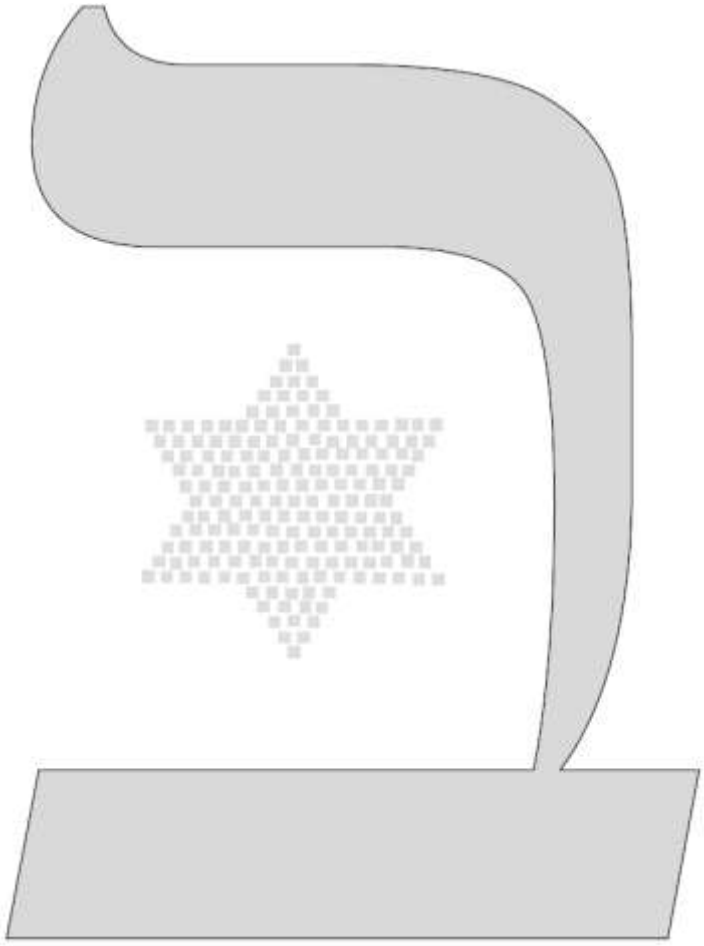

Figure 6. The first letter of the Torah, Bet, "housing" the rest of the letters of the Torah as its dagesh (dot). This diagram represents all 304,805 letters of the Torah. The star is comprised of 181 squares each with 421 sets of 4 letters (a star made of squares made of squares), the same shape shown in Figure 5.

Twitter followers, and people who submit online forms, understand crafting a message to meet a certain number of words or letters. You might personally experience that such constraints are achievable – satisfying to achieve! - and can inspire creativity. What is described above would indeed be a remarkable feat, if it were intentional. Even so, people sometimes pull off remarkable feats. Ancient people arranged millions of actual stone blocks into 3-D pyramid shapes. Arranging thousands or hundreds of thousands of words or letters into triangles, squares, and stars would take some doing, but these are fundamentally shapes and quantities that ancient people otherwise successfully dealt with to produce various large-scale and long-lasting cultural monuments. While one author or contemporaneous group could conceivably do such a thing alone, it is also conceivable that later editors or co-authors could complete a text in this way, ensuring at completion that a particular game is represented throughout the entirety of the completed text.

2.4 Internal references to language construction

In the book of Genesis, Adam names the animal species after contemplating each animal. On the surface, this is a description of the intentional generation of a lexicon – language invention.

In the Tower of Babel story, people are punished by being forced to speak a variety of new languages.

These parts of the text acquire a new implicit meaning if one imagines them being composed themselves in an invented language. The author would be describing something the author has recently been up to, and is in the midst of.

The contemporary Conlang Society for authors of invented languages adopted the Tower of Babel as their symbol (see conlang.org), and translating the Tower of Babel story has become a standard exercise for conlangers (see e.g. CALS 2020). In other words, the plain meaning of the Torah apparently resonates about language invention to language inventors.

2.4.1 Combinatorics

The word "Israel," already described above, may further serve as a particular sort of cryptographic reference to the generation of the language. Two alternate Gematria values have previously been pointed out if the word "Israel" is taken as a rebus:

- "yesh (there are) 231"
- Taking the Aleph to mean 1000, which, as mentioned above, is the surface meaning of the word Aleph-Lamed-Pe, the Gematria value of Israel becomes 1541.

The numbers 231 and 1541 are the number of distinct combinations of two- and three- long letter strings, respectively, from a 22-letter alphabet (the Hebrew alphabet has 22 letters). These are exactly the sort of number values that would be on the mind of an author – or analyst – thinking about inventing a language including permutation and combination of letters, exactly the procedure *Sefer Yetzirah* proposed.

2.5 Alternative explanation

The Torah text is old, but has been copied by assiduous, if human, scribes. Copying errors would interfere with the preservation of some, but not all, letter-level games. Only a handful of potential letter-level differences exist between the canonical Torah version primarily in use today compared to other existing scrolls (Drazin 2009), and most of the few differences would not impact any of the discussion above. Changing one letter for another (a "point mutation") would not affect a count of the number of letters. Adding or subtracting even a word would not change the names or spellings of the letters or the Fibonacci sequence within the letter Aleph. None of the proposed changes during transmission affect the first verse of Genesis, or the spelling of the word Israel. *Etc*. Only one of the games described above (Figures 5 and 6) requires that not a single letter has been added or deleted since the completion of the authorship of the text. As described later, the overall question of inventedness can be informed by converging evidence and does not rely on any one "game" to be intentional.

Alternatively, precisely because the Torah text has been examined so closely and from so many angles over the years, maybe centuries of analysis have reinforced a millennia-long garden path of confirmation bias, data mining, and cherry-picking. In this scenario, any text could be made to seem replete with inventive linguistic features, if only it were to receive as much attention and analysis as the Torah has received over the years. Look at any noise long enough, this critique would suggest, and human perception and cognition are such that, eventually, a (false) signal may appear. Perhaps if any text in any language were scoured as much as the Torah has been then an equally compelling case could be made that it, too, is actually unnatural. See McGough (2014) for a personal narrative of the seduction of cherry picking (followed by his recognition of the error of his ways) and e.g. McKay et. al. (1999) arguing that biased selection of statistical tests led to false findings of "equal-letter-skip" patterns in the Torah specifically.

Ioannidis (2005) has helped show how biases in generating and interpreting data have led to false conclusions throughout the sciences, and McKay et. al. (1999) discuss the particular case of false identification of coded messages in the Torah text through biased statistical interpretation. The underlying potential issue is the same problem Rennie (2004) describes in pharmaceutical industry clinical trial reporting and publishing: a "persistent bias in favor of positive results" (1359). If you only pay

attention to the interesting findings that you want to see, you end up finding patterns that were never there.

Unlike other claims of intentional design, like for a purported fossilized arrowhead that might be intentionally designed or might be just a rock (see e.g. Lacaille 1940), a strong version of the argument from cherry picking prohibits considering existing evidence around natural vs. unnatural language in the Torah in an ordinary fashion because the artifact at hand already and permanently has been over-analyzed through a biased process. While possible, this is also unsatisfying from a scientific perspective, because there would be no path forward for new knowledge – the existing over-analysis automatically gets in the way of any progress evaluating theories of intentional language-level games.

Rennie (2004), Nosek et. al. (2015), and Moonesinghe, Khoury, and Janssens (2007) are among those who point to an empirical way out of the situation of historical data seeming to support an interesting proposition, yet there is also concern over historical biased analysis or reporting in favor of positive results. Their way out is trial registration and replication to support meta-analysis. Replication addresses the issue of biased analysis by setting a new, independent group to work conducting the analysis or experiment. Trial registration addresses the issue of biased reporting of positive results by first collecting a record of intention to conduct an experiment. In this way, a meta-analysis, or analysis of all trials taken together, remains aware of all attempts to replicate a finding, making it hard for negative or null results to disappear. Replication allows a proposed meaningful signal to be distinguished from mere amplified noise whether or not any historical analysis was biased towards positive results.

3. Methodology

Engineers - called Imagineers - at Disney World place circles and ovals reminiscent of Mickey Mouse's head and ears in creative and surprising places. You might see a Hidden Mickey in the shape of a field of solar panels, the topiary of a bush, the arrangement of knobs in a shower stall, or in a collection of leaves swept into that shape by a creative groundskeeper. There are (at least) hundreds of these. You can Google this – and please do, to avoid me having to think about image copyrights and the Walt Disney World company. Hidden Mickeys are compelling in part because they are surprising, showing up at different scales and within a variety of architectural and engineering materials. Hidden Mickeys sometimes require an enormous amount of engineering and planning, all for a seemingly small whimsical payoff that vacationers might or might not ever notice. But when they do notice, they are delighted.

What should scientists do if we wanted to evaluate, with skepticism, whether Hidden Mickeys were intentional or the product of vacationers' over-active imaginations? Let us postulate that interviewing the Imagineers is not an available option.

One option would be to set vacationers (human subjects) to task at different tourist locations by asking them to rate the likelihood that some whimsical theme has been engineered into each tourist destination's infrastructure. These vacationers would visit Disney World, state fairs, local museums, etc. Would naïve, unbiased, but attentive and intelligent vacationers re-discover the Hidden Mickeys – and fail to discover whimsical themes in other destinations? By not drawing the vacationers' attention to "Hidden Mickeys" per se, this sort of experiment could help assess whether to the neutral unbiased intelligent human eye, Disney World contains secret or hidden patterns engineered into itself. This experiment is essentially a replication of the initial "work," whatever it was, of discovering Hidden Mickeys.

What would replication look like in the context of examining whether Biblical Hebrew exhibits signs of intentional design? There are different ways to go; following is one option. Imagine a small group of cryptographers, linguists, theologians, and recreational linguists and mathematicians who have no particular background in Hebrew. Researchers pose the following problem to them:

> *Imagine that we know that at least one of the following 10 texts that we provide to you represents an artificial language and was written to incorporate various coded alphanumeric messages and letter and word games. Your job, as a group, is to evaluate each text and develop a level of confidence as to whether each text is an artificial language with coded messages. You have no time limit - this exercise may take you days, months, or years - but the exercise is not complete until you have ranked the likelihood of each text to represent an invented (rather than natural) language. You should also provide your reasons for your rankings: what about each text led to the ranking you as*

a group give it? No internet searches or other background materials are allowed, you must examine these texts as they present themselves to you.

The control texts provided could be selected to be of similar length to the Torah, but in a language that has never been particularly implicated as including Gematria or having artificial origins, for example a Portuguese novel and a Russian cookbook series. Perhaps some additional background could be given – a dictionary, or a basic guide to grammar, maybe a standard accepted alphabet and pronunciation guide. Or perhaps not, as there is more than one way to run such a study, and more than one such study could be carried out. Maybe researchers would opt to design an artificial language and write a text with coded meanings in it to include as study stimuli, to allow confirmation that such a thing is detectable by the method and the group would identify it as such. Various study design options are available.

Will the Torah's Hebrew emerge with higher-than-average ratings of likelihood to be artificial, compared to agreed-upon exemplars of natural language texts? Would naive code-seekers latch onto Hebrew-letters-as-numerals by taking their shapes as cues? Would word lengths not match expected distributions of word lengths? Or, when given equal attention as other texts, does the Torah provide no particular elevated signal of hidden codes? Would participants in the experiment see some accidental feature of the Portuguese novel as suspiciously code-like, and be equally likely to develop a theory of Portuguese-as-an-invented-language as they would a theory of Hebrew-as-an-invented-language? And, because we do not want to take the chance of only considering results from such experiments that turn out with positive (or negative) results, trial registration (declaration of intent to perform such an experiment) would let a meta-analysis consider all such experimental attempts, taken together.

Levitt (torahcodes.net 2020) describes searching for and failing to find Torah-like codes in other texts ("monkey texts"). To this method I would just add blind analysts and control that each alternate text receives overall as much creative and dedicated analytic attention as the Torah text receives. Tenen describes asking interested code-breakers to approach the first verse of Genesis as if it were a code, but not telling them that the sequence presented to them was from the Torah. One subject proposed decoding in base-3, in which case each letter and its base-3 opposite form an a-a-b-a pattern and suggest symmetrical cube and spiral shapes (see meru.org). This is also similar in spirit to the experiment I am proposing, but again I would add the control condition of asking code-breakers to approach additional texts as well. When the amount of code-breaking attention is formally controlled, does the Torah re-emerge as the letter-level coded work that has been described historically?

4. Results and discussion

This paper factors out the questions of divinity and identify of the author and organizes past observations of seeming linguistic intentionality in Hebrew into a single theory of an invented language. It reviews features of the Torah's text that have been seen as unnatural, or invented, and introduces a few new such features as examples of such findable, seemingly deliberate letter-level games. This paper also describes how over-analysis and "cherry picking" might be responsible for the fact that we can produce such a list. Then, it offers a new empirical approach to help generate data that would avoid the question of cherry-picking: would naïve analysts, given enough time, re-find such features in the Torah compared to control texts? Such an approach is not intended to be decisive, but can contribute to an overall conversation about converging evidence.

If the language were invented, particular historical implications would stem from the mului, the analysis of the supposed "names" of the Hebrew letters used to write strings of text in the Torah. For an author to intentionally construct text with an awareness of milui naturally requires that the author knew the names of the letters. But, the names of the letters per se are not included in the Torah text itself, so it is only by extra-textual analysis and tradition that we have a theory of the names of the letters and that they exist in a particular order. (Aleph, for example - "Aleph Lamed Pe" - does appear in the text, but as a word meaning "thousand," not as a letter name.)

Schoolchildren learn "Aleph-Bet" songs. The recited alphabet provides a list of 22 letter names, so about 70 characters when all are spelled out using that self-same alphabet, and all in a specific sequence. As described above, this particular supposed sequence lets the Fibonacci sequence and figurate numbers appear in words and phrases that do appear in the text. If so, then an oral alphabet

"song" would be a like a mnemonic for a sequence of about 70 characters. This number of characters in sequence is similar to the length of a Bitcoin or ssh "private key," which, like a good password, is practically unguessable, even given trillions or quadrillions of guesses.

It is "easy" to produce a cryptographic key and encode text with it, but decoding needs the key or you are cryptographically stuck. In this proposed key that is the Hebrew alphabet, the name of the first letter is "Aleph" (Aleph Lamed Pe) and not, say, "Aleb" (Aleph Lamed Bet), and it specifically comes first in the sequence and, not, say, ninth. The Hebrew alphabet but with the letter "Aleb" in the ninth position would ruin the Fibonacci game and many of the figurate number star shapes (while other textual games described in this paper would remain intact).

In this scenario, the Hebrew alphabet, meaning letter names in sequence, as is currently taken to be standard, therefore 1) must have been in the mind of the author, and 2) must have been successfully transmitted through to us - and we could say the same for any text string for which we have a sufficiently long key that, when applied to the text string, unlocks a meaningful message. The fact that we can find games using the names and sequence of the letters of the alphabet would in this way carry historical entailments. Note that the author of a key intended to be transmitted orally might encode text using some number of multiple versions of the key based on anticipated likely mis-hearings, errors, or alternate reasonable ways to write down or spell what is heard, all without meaningfully affecting the unguessability of even the approximate correct sequence. Further investigation here might reveal more about the author's expectations for transmissibility and about the historical process of transmission.

5. Conclusion and policy implications

Given that there are many known examples of highly creative invented languages, it seems natural to estimate that that Hebrew is one of them, rather than to expect that every miscellaneous observation that seems to support that perspective is for its own reason misguided or incorrect. No existing well-supported historical or linguistic theories are harmed if Hebrew were invented. If impossible or supernatural mechanisms were entailed, like time travel, or if the theory contradicted established clear facts, then more or stronger evidence would be required. As it stands, invented languages are unusual, but far from impossible, and this particular language (Hebrew) has been described as having features that could be present in invented, but not natural, languages. However, it is also an interesting possibility that over-analysis has made conclusions in this area difficult, because it is difficult to estimate how many analyses have been attempted but uninteresting results not described ("cherry picking"). If so, we might end up with many miscellaneous, but ultimately random and meaningless, examples of seeming unnatural aspects of the text.

To be transparent, I personally think it is reasonable to estimate that Hebrew is invented and that at least some of the apparent letter-level games in the text are intentional. The parsimony of the roots that have struck others, the Fibonacci sequence coming out of the letter Aleph (as well as the name Adam), the multiple Gematria figurate star shapes findable around the first verse, the shapes of the letters to serve as numerals, and then the story-level description of Adam determining names of animals, are all explained neatly if the language and text are a single deliberate linguistic construction. However, this is not a logical, philosophical, or theological conclusion, but a theory that seems plausible, and one that can be adjusted and considered in various ways.

Whichever way one judges the current state of evidence, one's judgement might be influenced, one way or the other, by results of additional new empirical studies in which blind analysts seek evidence of intentional language design in the Torah as well as in control texts. However, new studies are not intended to be decisive, but simply could enter the discourse of reasonable consideration of converging evidence.

We can also speculate about entailments of the theory that Hebrew was invented, even while the theory itself is still also subject to critical examination. If the language was invented, why might someone have done so?

1. If you have elected to write an abstract philosophical guidance touching on ultimate ethics and the origin and purpose of the universe, you might not want to start by being attached to one particular existing human language. Perhaps such a book would be more universal if you make up a language with which to write it.

2. If you invent the language, you can control how words and concepts are linked to each other to succinctly convey linkages between concepts. If you want "peace" and "love" to be implicitly linked, you could make them anagrams of each other, or you could let two words with different letters share a gematria value. Instead of spending time and space writing explicitly or asserting that "peace is related to love," instead you can convey a connection just by including those words anywhere in the text.

3. A "checksum" in computer science is a form of verification that computer code is not corrupted. If there is a clear letter-level code present in a text, you can have some confidence that the text has had no copying errors. If you have two proposed versions of the text different by one letter, and only one would complete a text-level pattern, you might take that into account in trying to reconcile which text might have been copied in error. Given that the Torah contains surface-level descriptions of careful copying of itself, this form of checksum might have been on the mind of the author.

4. To inspire careful copying. Readers and scribes who are aware of the fact of letter-level codes might be especially careful to copy exactly, thus preserving also the surface-level meanings.

5. Just for fun. Like for people today who invent languages, maybe it was in the personality and whim of whoever wanted to expound on ethical monotheism in a written text to do so in the form of an invented language. There are always authors and audiences who want to delight themselves and each other.

This journal seeks policy implications. Research efforts and directions may sometimes end up governed by categories of academic departments, earmarking of funds or grants for the use of particular methods in particular areas, etc. However, the exploration of the theory described in this paper would seem to cut across disciplines. There might be a place for analysis of texts by Rabbinic scholars of past centuries alongside development of pattern detection algorithms and human subjects experiments. Therefore, policy planners might consider ways to support research that are less limited by traditional boundaries of disciplines.

The characterization of non-naturalness or inventedness *per se* is agnostic to theology or the identity of the author(s). Regardless of the identity of the author(s), the process of authorship might be informed by examining the creative cognitive complexity inherent in this artifact. The intentions and creative process of the author(s) might also be illuminated by further analysis of the textual games. And, some might find enjoyment in the way of an impressed audience, like people might receive avant-garde works of recreational mathematics or language games.

Acknowledgements

Thanks to reviews by and discussion with Miriam Steinberg-Egeth, Howard Egeth, Robert Kurzban, Francisco Gil-White, Art Levitt, members of the online conlang community, and members of the Center City Philadelphia Jewish community.

References

Belaga, E.G. (2014). Interpreting Semitic protolanguage as a conlang or constructed language I. US-China Foreign Language. 12(3):183-192.

Bradford, D. (2008). In the Beginning: Building the Temple of Zion. Bloomington: Trafford Publishing.

CALS (Conlang Atlas of Language Structures) (2020) The Tower of Babel. https://cals.info/translation/the-tower-of-babel/ [Accessed July 22, 2020].

Daf Yomi Review (2020a). Fibonacci and Genesis. http://dafyomireview.com/torah_numerology.php?d=14 [Accessed July 22, 2020].

Daf Yomi Review (2020b). Divine Signature - 1820 http://dafyomireview.com/torah_numerology.php?d=13 [Accessed July 22, 2020].

Daf Yomi Review (2020c). Geometric Properties. http://dafyomireview.com/torah_numerology.php?d=10 [Accessed July 22, 2020].

Daf Yomi Review (2020d). Beyond Decimal. http://dafyomireview.com/torah_numerology.php?d=7 [Accessed July 22, 2020].

Deza, E., and Deza, M.M. (2012). Figurate Numbers. Hackensack: World Scientific.

Drazin, I. (2009). Maimonides: Reason Above All. Jerusalem: Gefen Publishing House.

Dunn, M. (2001). Ella Minnow Pea: A Progressively Lipogrammatic Epistolary Fable. San Francisco: MacAdam/Cage Publishing.

Ginsburgh, Y. (2015). 913: The Secret Wisdom of Genesis. Jerusalem: Gal Einai Institute.

Grzybek, P. (2007). "History and methodology of word length studies," in Contributions to the Science of Text and Language (Dordrecht: Springer) pp. 15-90.

Hall, R. (2005). Math for Poets and Drummers. Philadelphia: Saint Joseph's University Press.

Halperin, L. (2012). Modern Hebrew, Esperanto, and the Quest for a Universal Language. *Jewish Social Studies: History, Culture, Society* 19(1): 1-33.

Higley, S. (2007). Hildegard of Bingen's Unknown Language: An Edition, Translation, and Discussion. New York: Palgrave Macmillian.

Horowitz, Y., "Hebrew Word List." PDF, 2020. http://www.rabbihorowitz.com/PYes/BB_Files/HebrewWordList10Page.pdf [Accessed July 22, 2020].

Ioannidis, J.P. (2005). Why most published research findings are false. PLoS Med. 2(8):e124. doi: 10.1371/journal.pmed.0020124.

Kaplan, A. (1990). Sefer Yetzirah: The Book of Creation. York Beach, ME: S. Weiser.

Kleffner, R. (2004). Lexipoeia: A Conlang Wordgen Tool. https://laptrinhx.com/lexipoeia-a-conlang-wordgen-tool-2937966151/ [Accessed November 30, 2020]

Lacaille, A.D. (1940). ASPECTS OF INTENTIONAL FRACTURE (Being notes on the flaking of some rocks other than flint as exemplified by some Scottish artifacts). Transactions of the Glasgow Archaeological Society, 9(4):313-341.

Lemos, M. (2009). Language-games in Alice's adventures in Wonderland or: How language operates in Carroll's text to produce nonsensical meanings in common-sense references. *E-fabulations: e-journal of children's literature 5: 23-34*.

Levitt, A. (2020). The Light of Torah Codes: Essential Introduction - Hidden Successes. http://www.torahcodes.net/new/pages/hidden_successes.html [Accessed July 22, 2020].

Leyendekkers, J., and Shannon, A. (2013). Fibonacci and Lucas primes. Notes on Number Theory and Discrete Mathematics. 19(2):49-59.

McGough, R.A. (2014). Debunking Myself: The Good, the Bad, and the Ugly. https://www.biblewheel.com/Blog/debunking-the-bible-wheel-the-good-the-bad-and-the-ugly/ [Accessed July 22, 2020].

McKay, B., Bar-Natan, D., Bar-Hillel, M., and Kalai, G. (1999). Solving the Bible code puzzle. Statistical Science 14(2):150-173.

Moonesinghe, R., Khoury, M.J., and Janssens, A.C.J.W. (2007). Most published research findings are false - but a little replication goes a long way. PLoS Med. 4(2):218-221.

Nosek, B.A., Aarts, A.A., Anderson, C.J., Anderson, J.E., Kappes, H.B., and Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. Science. 349(6251): 253-267. doi: 10.1126/science.aac4716

Okrent, A. (2009). In the Land of Invented Languages: Esperanto Rock Stars, Klingon Poets, Loglan Lovers, and the Mad Dreamers Who Tried to Build a Perfect Language. New York: Spiegel & Grau.

oren evron. "Fibonacci in the Torah - Genesis Golden Ratio." YouTube Video, September 16, 2015. https://youtu.be/OEitTICcVko

Rennie, D. (2004). Trial registration: a great idea switches from ignored to irresistible. JAMA, 292(11):1359-1362. doi: 10.1001/jama.292.11.1359.

Serafini, Luigi. (1993). Codex Seraphinianus. Milan: Franco Maria Ricci.

Singh, P. (1985). The so-called Fibonacci numbers in ancient and medieval India. Historia Mathematica, 12(3), 229-244. doi: 10.1016/0315-0860(85)90021-7.

Wimmer, G., Witkovský, V., and Altmann, G. (1999). Modification of probability distributions applied to word length research. Journal of Quantitative Linguistics 6(3):257-268. doi: 10.1076/jqul.6.3.257.6163