

The McKay et al. paper

The paper “Solving the Bible Code Puzzle” by Brendan McKay, Dror Bar-Natan, Maya Bar-Hillel, and Gil Kalai (MBBK) was posted on the Internet in June 1999 and was accepted for publication by Statistical Science for the May 1999 issue. This paper addresses two main questions concerning the WRR experiment. In their words, “In precise terms, we ask two questions: Was there enough freedom available in the conduct of the experiment that a small significance level could have been obtained merely by exploiting it? Is there any evidence for that exploitation?” (MBBK, page 151). It is clear that the answer to these questions is “no”, based on the analysis presented in this paper up to this point. We have shown clearly that (1) for the most part, there is no wiggle room, and (2) even in those cases where there theoretically is wiggle room (as for list 1, or for the appellations in lists 1 and 2) it was not exploited. Therefore any “evidence” for such exploitation must necessarily be inconclusive. One cannot have conclusive evidence for something that is not true. Yet, MBBK do present evidence, and it is worthwhile understanding why that evidence is not conclusive. In addition, the MBBK paper does raise a few issues that we have not yet discussed.

The tool used by MBBK to provide evidence that “wiggle room” was exploited to produce apparently significant statistical results in WRR is called “the study of variations”. On page 158 they say, “...there is significant circumstantial evidence that WRR’s data is indeed selectively biased toward a positive result. We will present this evidence without speculating here about the nature of the process which lead to this biasing. Since we have to call this unknown process something, we will call it *tuning*. Our method is to study variations on WRR’s experiment. We consider many choices made by WRR when they did their experiment, most of them seemingly arbitrary (by which we mean that there was no clear reason under WRR’s research hypothesis that they should be made in the particular way they chose to) and see how often these decisions turned out to be favourable to WRR”. In other words, since these choices are “arbitrary”, by chance alone one would expect about half to be favorable to WRR and the other half to be unfavorable. If most of the choices turn out to be unfavorable to WRR, this is evidence that the WRR parameters were chosen because they were known to be favorable, i.e., better than many other choices of parameters. Thus, there must have been many “hidden” experiments used to effect the “tuning” of these apparently arbitrary choices.

The first flaw in MBBK’s case is in their very words: “...there was no clear reason under the research hypothesis that they should be made in the particular way they chose to...”. As an example of the problem, let us apply this argument to the proximity formula (which MBBK do on page 169). The idea that all the choices in the proximity formula (and there are many) were made blindly is absurd. We fully expect that Witztum and Rips made several observations of the phenomenon before constructing a hypothesis to test in an experiment. Thus, we would expect that any decent mathematician – and Rips is world class – would construct a proximity formula that truly measures the Torah codes phenomena that they observed, providing the phenomena is real. If the phenomenon is not real, then the proximity formula is meaningless and it is not expected to work better than other formulas on new data in an experiment. It follows that the formula is expected

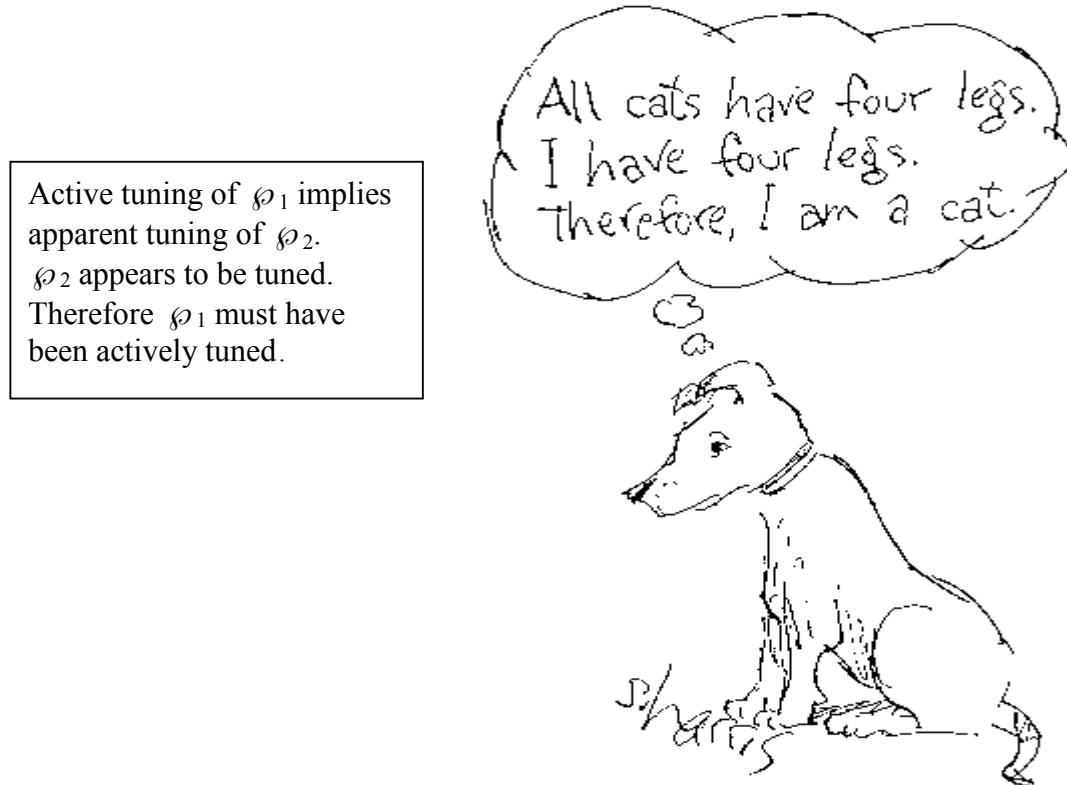
to look “tuned” if the Torah codes phenomenon is real because it was tuned – on earlier observations of different data. This is perfectly honest, and is, in fact, the normative scientific procedure. In scientific research, one makes observations, forms a hypothesis based on those observations, and then tests the hypothesis on data that is disjoint from the original observations. Hence, if one finds evidence of tuning, there are two possibilities. (1) Torah codes exist and Witztum and Rips correctly constructed the mathematical parameters of the WRR experiment based on prior observations of other anecdotal examples. (2) Torah codes do not exist and therefore the mathematical parameters of the WRR experiment should be totally arbitrary. If (2) is true and evidence for tuning exists, this implies that these parameters were tuned to produce an apparently significant result on the WRR experiment. Consequently, if one assumes that there are no Torah codes (case 2, above), then evidence of tuning is evidence of a fraudulent experiment. We see then that the use of evidence of tuning to discredit WRR works only if one assumes first that there are no Torah codes (case 2 as opposed to case 1). This is known as “circular reasoning”. What we have shown here is that even if there is evidence for tuning, this is not evidence that the WRR experiment was not done honestly – unless one first assumes that there are no codes in the Torah. The apparent tuning may simply be the result of Rips and Witztum constructing an appropriate measuring tool to detect the phenomenon based on previously observed examples of codes. In fact, we can go further. Given the evidence that we have presented that tuning could not have taken place, evidence for tuning is actually evidence for the existence of codes. As we pointed out, if no tuning took place and Torah codes do not exist, then there can be no evidence of tuning. This logical argument effectively destroys MBBK’s entire “study of variations” on WRR. We shall see shortly, that in any case the evidence for tuning is non-existent.

At this point, the reader may wonder why MBBK discuss tuning of the proximity measure and the date forms when we have already shown that these components could not possibly have been tuned. They were fixed for the experiment on list 1 and thus could not be changed for list 2. MBBK ask this very question and provide two answers. On page 158 they say, “This naturally raises the question of what insight we could possibly gain by testing the effect of variations which WRR did not actually try. There are two answers. First, if these variations turn out to be overwhelmingly unfavourable to WRR, in the sense that they make WRR’s result weaker, the robustness of WRR’s conclusions is put into question whether or not we are able to discover the mechanism by which this imbalance arose. Second, and more interestingly, the apparent tuning of one experimental parameter may¹⁸ in fact be a side-effect of the active tuning of another parameter or parameters.” The first answer given by MBBK is of no relevance to the question of whether the experiment on list 2 was honestly done or not. If anything, a lack of robustness implies that only a narrow range of parameters uncovers the Torah code phenomenon. This is expected if the phenomenon is real – if one tries to detect the code the wrong way, it is not detectable. Concerning the second answer, MBBK admit that for list 2 almost nothing could have been tuned except the appellations (MBBK, page 159). However, they have no way to prove that the appellations were tuned. Instead, they attempt to bring evidence that the proximity formula and the date forms look tuned and then rely on the possibility that this “may” be a side effect of active tuning elsewhere.

¹⁸ This author’s emphasis.

Nowhere do MBBK prove or provide empirical evidence that this statement is true. Their entire thesis concerning active tuning of the experiment on list 2 is based on a “may be”! Until this presumed link between active tuning of one parameter and apparent tuning of another parameter is proven, the strength of their argument reduces to “may be”; a logical argument is no stronger than its weakest link.

It is essential to note that even if their “may be” turns out to be correct, the implication goes the wrong way. They claim that active tuning of a parameter (i.e., the appellations) may cause an apparent tuning in another parameter (e.g., the proximity formula). That is to say, statement A: “Active tuning of parameter φ_1 (may) imply apparent tuning of parameter φ_2 ”. MBBK then claim to demonstrate that parameter φ_2 appears tuned. From this they wish to conclude that φ_1 must have been actively tuned.



This conclusion, however, is only valid if statement B: “Apparent tuning of parameter φ_2 implies active tuning of parameter φ_1 ” is true. But statement B is the converse of statement A and logically the truth of a statement does not imply the truth of its converse. This is a well-known fallacy in formal logic.

We have refuted the MBBK thesis of tuning thrice over on purely logical grounds. First, we have shown that even if there were evidence for tuning, this would not imply that the experiment on list 2 was a hoax unless one first assumes that the Torah codes do not exist. This is circular reasoning. Second, their evidence rests on an assumption for which they provide no proof or evidence. Third, their argument rests on the false logical assumption that the truth of a statement implies the truth of its converse. Note, too, that the additional experiments of Witztum and Gans could not be tuned at all; MBBK deal with those experiments by ignoring them. There is, however, still another problem with

MBBK's evidence. How can we be certain that the choices of parameters of WRR made by MBBK for documentation of tuning were not themselves tuned? Since tuning is an "unknown process" (MBBK, page 158), perhaps it infects the choices made by MBBK? Clearly, a bias in selecting which parameters to examine for "tuning" will result in a biased conclusion. As in MBBK, we will not speculate on the causes of such tuning except to note that they do say on page 152, "Nothing we have chosen to omit tells a story contrary to the story here" and on page 161, "Our selection of variations was in all cases as objective as we could manage; we did not select variations according to how they behaved". Nevertheless, their study of variation seems to manifest tuning. For example, we showed in the section above on "hidden failures" that changing the number of personalities used on list 1 (to 20) and changing "bunching threshold" in the calculation of P1, whose value can be arbitrarily specified, from 0.2 to 0.5 makes the measure 16 times stronger. For the entire list 1, this single change makes the P1 measure about 7,000 times stronger! Let us examine how MBBK present this fact. On page 171 they state, "Table 10 shows.... The same table shows the effect of changing the cut-off 0.2 used to compute P1 and P3. Values greater than 0.2 have a dramatic effect on P1, reducing it by a large factor (especially for the first list). However, the result of the permutation test on P1 does not improve so much, and for the second list it is never better than for P4". Now here we have definite evidence that WRR did not tune list 1, since at the time that the WRR paper was first submitted for peer review, Diaconis had not yet suggested the permutation test. Nor had list 2 been requested. Hence, the only measure of success at the time was P1. If we next examine MBBK's table 10 numbers for cut-off values greater than 0.2 we find the following entries:

Cut-off defining P1:

0.25	[1 , <u>0.8</u> ; 1 , 1.1]
0.33	[1 , 1.0; 1 , 1.0]
0.4	[1 , 1.0; 1 , 1.0]
0.5	[1 , <u>0.4</u> ; 1 , 1.0]

Values smaller than 1 represent improvement and are underlined, e.g., 0.5 would be twice as strong. The boldface "**1**" means that the variation does not apply so that the value cannot change. Four score changes are presented for each variation. The greatest change is for a P1 cut-off of 0.5, and it is only 0.4 – a bit more than twice as strong. This certainly does not seem to be very "dramatic" and it is a far cry from 7,000 times better! How do we understand this? The reason is quite simple. If we go back to MBBK's page 160 we find that the four values listed in the table are the P2 value for list 1, the least permutation rank for list 1, the value of P4 for list 2, and the least permutation rank for list 2 (each divided by a constant). P1 is not shown in Table 10 even though it purports to examine "Cut-off defining P1"!¹⁹ In this way, the "dramatic" improvement that MBBK mention in the text of their paper is nicely hidden from view when examining their table. Effectively MBBK's notation, which controls what is revealed and what is hidden, has been tuned! Specifically, for list 1 it hides what is essential, P1. We shall see more examples of this phenomenon shortly. Note, too, that in studying different date forms (MBBK, page 156), the date forms used by WRR were not optimized (for list 1!).

¹⁹ Recall that when the experiment on list 1 was performed, P1 and P2 were the only measures known. P3, P4, list 2, and the calculation of permutation rank (probability) were introduced later.

Furthermore, all of the non-standard date forms did poorly – a result that might be expected if the Torah codes are real.

Let us now turn to the presentation of McKay's evidence of tuning. On page 169 they provide a table of different proximity measure variations and how they affect the strength of the WRR proximity measures and probabilities. There are 67 variations considered of which only 20 show at least one of the measures improving. Bear in mind that as a result of their choice of notation, one can only see the improvements that MBBK want revealed. For example, it is conceivable that P1 improves for all 67 variations but their table would not show any trace of this improvement. Consider the following example. Suppose we wish to study 4 variations. We calculate the results of variations 1 through 4, and then add in the results of variation 4 combined with the first 3 variations. We thus produce 7 values for only 4 variations. If variation 4 and each of variations 4 combined with the first three variations produces weaker scores does this imply that 4 variations are weaker, or does it mean that variation 4 is weaker and it does not matter what other score it is combined with? MBBK actually do precisely this! They list 34 variations (33 in the first column and 1 at the top of the second column) and the 34th variation combined with the other 33²⁰. This 34th variation produces very weak scores – on the average it makes the four proximity measures 95 times weaker. When it is combined with the other 33 variations, not a single one makes the WRR measures stronger. On the other hand, if we look at the 34 variations by themselves, 20 of them have at least 1 improved measure²¹. It is also interesting to note that MBBK do not present a table of values for multiple changes which include changing the P1 bunching threshold from 0.2 to 0.5; presumably there would be too many improvements to support their hypothesis! We see, then, that the presentation of the “evidence” for tuning has itself been tuned!

For the WRR experiment, we have shown that tuning is either impossible (the date forms and the proximity formula), or we can present strong evidence that there was no tuning (the appellations and the testimonies of Rav Deutch and Rav Fisher, and the additional experiments of Witztum and Gans). This is not possible for MBBK’s presentation of evidence since we cannot know how many variations not presented would show evidence against their thesis. In fact, their thesis must be taken on faith. Perhaps it is for all these reasons that MBBK state, “...we are not going to attempt a quantitative assessment of our evidence. We merely state our case that the evidence is strong and leave it to the reader to judge” (MBBK, page 159). In other words, all this “evidence” rests on a “may be”, has two fatal logical flaws, shows evidence of itself being tuned, and cannot be quantified. The “evidence” is left to the individual reader’s judgement!

²⁰ It is interesting to note that this introduces dependencies in their table. Dependencies make mathematical analysis difficult. When they cannot be avoided, as in the case of lists 1 and 2, elaborate procedures must be developed to circumvent the problem (if possible). This is the first time I have ever seen dependencies purposefully and unnecessarily put into data!

²¹ It can be shown statistically that the population consisting of the first 34 single variations is different than the population of the 33 double variations, and therefore they should not have been included in a single table. Specifically, the Mann – Whitney Sum of Rank statistic comparing the two populations gives a score of 6.42 sigma, indicating that the probability of the two sets of variations coming from the same underlying distribution is 6.8E-11.

On page 161, MBBK raise another objection to the WRR results. They compare the P2 proximity measures obtained for lists 1 and 2, viz.: 1.29E-9 and 1.15E-9, and note that they are unexpectedly close. The claim is that they are closer than expected even if the WRR hypothesis were true. They conclude that this is evidence that list 2 was tuned to give a proximity measure close to list 1 because of “naive statistical expectation” that the two lists should give close proximity measures. MBBK estimate the probability of the two measures being so close as less significant than 1/100 and therefore admit that one cannot “conclude too much from” this example alone (MBBK, page 161). They next proceed to bring another example to bolster their theory – and herein lies the downfall of their thesis. They note that if one partitions the Gans city list into its list 1 and list 2 components, the probabilities obtained on the two parts are again very close. They estimate the probability of this closeness is less than 1/500 – clearly too small to be ascribed to chance alone²². Before we deal with this “closeness” phenomenon, note a subtle change in the way MBBK report their results. For the WRR experiment, they measure the closeness of proximity measures, while for the Gans experiment they measure the closeness of probabilities. This is like mixing apples and oranges. In looking at two samples that supposedly manifest the same phenomenon, the measures used on both samples should be the same. The fact that MBBK chose to switch measures suggests that several different things were measured, and whatever happened to support their hypothesis better was quoted in each case. Thus, the presentation of the MBBK data once again shows evidence of itself being tuned. Even if we assume that their observations are valid, recall that every single city name in lists 1 and 2 were generated mechanically using the Inbal protocol. There is not one exception. Therefore, it is absolutely impossible that the list 2 component of the cities list could have been tuned to match the list 1 component! We have thus proven that this “closeness phenomenon” between lists 1 and 2 does not imply that list 2 was tuned to match list 1. The “closeness phenomenon”, if it is not simply a creation of MBBK’s tuning, shows that there is something present in these lists which is not random – and not a result of tuning. But then, this observation is consistent with the WRR hypothesis that there is something very non-random in these lists. The specific cause of this “closeness”, if it really exists, remains to be explained.

We have disposed of all the arguments and evidence presented in the MBBK paper save one. We have left what they consider to be their strongest argument for last. In their conclusions, on page 167, they say, “Be that as it may, our most telling evidence against codes is that we cannot find them. All of our many earnest experiments produced results in line with random chance.” Note the use of the word “earnest”. This is important because if their experiments were not earnest, they cannot be expected to succeed, even if there are codes in the Torah. In fact, a necessary (but not necessarily sufficient) condition for an experimental success is that the experiment be carefully designed.

MBBK claim to have performed “many real experiments” (MBBK, pages 163 and 165) of which they document a handful. Let us examine some of the experiments that they do document. On page 164 they report on a “cities” experiment of Barry Simon in

²² In all fairness, shouldn’t the threshold here be 1/1,000 also? Apparently there is a double standard!

²³ If these sigma values were normally distributed, 3.47 would be 3,892,000 times weaker than 6.42.

which he “uses the names of all cities mentioned in each rabbi’s entry in Margalioth’s encyclopedia as places of birth, death, living, working or studying, without any modification of spelling or addition of prefixes.” First, note what has been added to the experiment. Gans only used cities of birth and death, paralleling WRR’s use of dates of birth and death. The fact that 3 new categories of cities were added to the experiment means that the experiment has wiggle room, that is, it could possibly be tuned to fail. Even more wiggle room is introduced by the ambiguity of terms such as “living”. Thus, for example, Simon lists “Jerusalem” and “Hebron” as placing of living for the Rambam even though the Rambam only visited these locations. Does “visit” qualify as “living”?

Let us now examine a more fundamental problem with this experiment. Consider the conclusions to be drawn from a hypothetical experiment in which we use WRR’s list 1 or 2, but use English dates rather than Hebrew dates. The failure of such an experiment would prove nothing. In fact, the success of WRR’s experiment with Hebrew dates provides a strong expectation that an experiment with English dates would fail; if the dates are known to be encoded in Hebrew, why should they also be encoded in English? The same is true for the cities. There are 129 distinct city name/spellings on the list, of which 64 have Jewish names in addition to their secular names. (For the remainder, we assume that if there was no specifically Jewish name for a city, then the Jews must have referred to that city by its secular name. Thus, the Jewish name and the secular name are the same). The Margalioth encyclopedia often uses the secular names of the cities; a comprehensive list of the Jewish names is obtained from the articles on the cities (not the article on the rabbis) and the index in the Encyclopedia Hebraica. By using the entries from Margalioth without any modification of spelling, they have effectively replaced many of the Hebrew names with secular names, or left out many Jewish names. Thus, for example, the birthplace of Rabbi Yehuda HaChasid is given in Margalioth as “arypc”, which is a Jewish name. However, the additional Jewish names/spellings for this city, “arypc”, “arypc”, “ruypc”, and “rypc” are obtained from the Encyclopedia Hebraica. This experiment is designed in such a way that much of the data is expected to fail and is expected to dilute any remaining statistical significance to within the range of random expectation. It is quite clear that this experiment, besides having been constructed so as to include wiggle room, was either very poorly designed, or tuned to fail.

MBBK also mention experiments which pair the rabbis of list 2 with their years of birth and with the names of the books that they wrote. The details of these experiments were first made public in May 1997. Here are a few observations on their data, taken from a letter to “Jewish Action”, vol. 59, No. 2 (page 90) by Doron Witztum: (a) the books of the Vilna Gaon were represented by a single book on geometry. (b) Many books were listed incorrectly. For example, they listed “hqzj dy” instead of “hqzjh dy” and “hbqrhh” instead of “hbprhh”. (c) Of the 66 dates given, at least 11 are incorrect. For example, R’ Avraham the son of the Rambam was assigned a year of death 48 years before he was born!

Let us examine still another example of one of MBBK’s “earnest” experiments. It will be recalled that in WRR’s experiment on list 2, 4 overall proximity measures were calculated. We indicated that the 3rd and 4th measures are minor modifications of the 1st

and 2nd measures respectively. We shall explain what this modification consists of. For technical reasons, the process used to calculate the proximity measures can only do so for words that are between 5 and 8 letters long, inclusive. Many of the appellations provided by Havlin have the prefix “rabbi” (in Hebrew, ybr, a 3-letter word). As a result, there is only 5 letters left for the name. In such a case, only the personal name of the personality was used, not the surname. An appellation without the prefix “ybr” has much more chance of including the surname, and thus being distinct. A simple example will illustrate this principle. The first 3 personalities on list 1 and the first 4 personalities on list 2 all have the same appellation, “Rabbi Avraham” (hrba ybr). Thus, there are 7 personalities with the same appellation. On the other hand, there are 13 remaining appellations for 6 of these 7 personalities, and each is uniquely associated with a single personality. It follows that if one wishes to maximize use of appellations that correspond to a unique personality, one must exclude appellations with the prefix “rabbi”. The 3rd and 4th overall proximity measures are exactly the same as the 1st and 2nd measures respectively, except they are applied to this subset of appellations from the original list. We thus have 3 sets of appellations associated with each list. (1) The entire list. (2) The subset scored by the 3rd and 4th overall proximity measures, in which appellations tend to be uniquely associated with personalities, and (3) the difference between (1) and (2). The Gans city experiment was scored with the 3rd and 4th overall proximity measures only because it is clearly advantageous to have a unique association between personalities and appellations, given the hypothesis of codes. (In fact, there is a down side to scoring set (2) only: the data size is smaller. Perhaps this is why for list 2, the 4th probability is best, but the 2nd is close behind). No one has ever suggested using set (3) above for an experiment because it makes no sense. Why score only non-unique data that is desirable to discard? Yet, this is exactly what MBBK describe on page 165. Not only is the use of this data questionable, the data size has been reduced from a total of 188 appellations to 55. Cutting the data size down to less than a third its original size, and retaining only the “undesirable” data is a sure way of insuring that any statistical significance in the original data will be destroyed. We can appreciate just how serious this reduction of data size is by computing what effect it has on the two proximity measures, even if we ignore the questionable quality of the data that has been retained. Just reducing the data size as indicated is expected to reduce the first proximity measure from a sigma value of 6.42 to 3.47, a dramatic weakening of the result²³. As for the second proximity measure, it is expected to change from 5.2E-9 to 5.7E-3, or 1,100,000 times weaker. We again have an experiment that is either poorly designed or expertly designed to fail.

It is important to note that the objections raised to this experiment are specific to this data set and may not apply to other data sets. Thus, for example, in Witztum’s “replication” experiment, both “*ben name*” and “*ben rabbi name*” were used. However, in that case the number of unique names of personalities is the same for both forms, and correct forms are not excluded. There are also legitimate experiments done on small data sizes, e.g., “personalities in Genesis”. The problem is not in a data size being small *a priori*, (although from a statistical point of view, bigger is usually better) but rather in the *a posteriori* reduction of the size of a set, whose score is known, to a point where the expected value of that score on the reduced data is within random expectation.

If these experiments are typical of MBBK's "many earnest" experiments, then their "most telling" evidence falls, along with all the other "evidence" reported in their paper. Finally, note that if MBBK feel that their inability to perform a successful Torah codes experiment is the strongest argument against codes, then surely the success of Witztum and his associate, Rips, Gans, Rottenberg, Bombach, and Schwartzman in performing new, mathematically verifiable, and wiggle-free codes experiments is the strongest argument for codes.

At this point, the reader may wonder why Witztum and Rips did not submit a rebuttal of the MBBK paper to Statistical Science. The reason is straightforward. On June 23, 1998, Professor Rips sent an email to the editor of Statistical Science stating, "I would appreciate very much if you [could] possibly let me know whether the journal 'Statistical Science' has received papers or comments or remarks or other material related to our paper 'Equidistant Letter Sequences in the Book of Genesis' (by Witztum, Rips and Rosenberg) which appear in the issue of Aug 94. I would be most grateful if you would send such material to us, in order that we will be able to respond to it and explain our point of view." The response from the editor of Statistical Science came the same day: "We have received papers related to your August 1994 Statistical Science article. These papers are currently undergoing review, and until a decision is made whether or not to publish them, we are bound by confidentiality not to disclose their contents or their authors. If a decision is made to publish any of these papers, you will be sent copies and invited to respond."

In May 1999, an Internet posting indicated that Statistical Science had decided to publish the paper by MBBK in that month's issue. Neither Rips nor Witztum had been informed of the publication, nor were they "invited to respond". On May 7, 1999, Witztum sent the following email to the editor of Statistical Science. "It was brought to our attention that you have decided to publish a paper by MBBK related to our Statistical Science article. If I may, I would like to remind you of your letter dated 30 June 1998 to Prof. Rips. If there is any misunderstanding on my behalf, please let me know." The response came the same day: "There is no misunderstanding on your part concerning my letter of June, 1998, but there has been a change of policy based on advice that I have received.... I have decided to publish the MBBK article without commentary of any sort... I regret having to retract my invitation to Professor Rips to respond to the MBBK article. I remind you, and him, that your original article also was not discussed..." No commentary is needed; the above letters speak for themselves.

Conclusions

There is a popular saying, "Where there is smoke, there is fire". One must wonder how Torah codes could be real given that so many arguments and pieces of evidence have been brought against them. Nevertheless, we have systematically refuted all of the critics' arguments. We have detailed new experiments that are "wiggle-proof" and that even MBBK do not attempt to refute. One of these experiments uses a word list provided by McKay and his associates! (This was the Nations prefix sample). We have even seen evidence that the very "tuning" that they accuse WRR of infects the critics' evidence and experiments. We have seen numerous instances where obvious opportunities for tuning

by WRR were passed up. We have documented a plethora of logical flaws in the critics' reasoning.

Still, the MBBK paper passed peer review for a highly respected journal. This is not easily dismissed. It is, perhaps apropos to quote from the directions for submissions to another highly respected journal: ECONOMETRICA. "If you plan to submit a comment on an article that has appeared in ECONOMETRICA, please follow these guidelines: First send a copy to the author and explain that you are considering submitting the comment to ECONOMETRICA. Second, when you submit your comment, please include any response that you have received from the author. ...Authors will be invited to submit for consideration a reply to any accepted comment of this kind."²⁴ It is abundantly clear that if a party refuses to hear one side of a dispute, that refusal is itself evidence of bias. Furthermore, without input from both sides, the peer reviewers cannot be unbiased even if they wish to be. They simply do not have all the information needed to reach a fair conclusion.

This same bias also manifested itself when Witztum submitted his papers "Personalities of Genesis and their Dates of Birth" and "A Replication of the Second Sample of Famous Rabbinical Personalities" to Statistical Science. These two papers report the results of the "additional experiments" described earlier in this paper. In a May 12, 1999 response to this submission, the editor states: "Neither paper offers anything new or interesting in terms of statistical theory or methodology. Given Statistical Science's declared policy of publishing papers with high statistical content, these papers are inappropriate for Statistical Science". This response was given in the same month that the MBBK paper was to be published by Statistical Science. Apparently, this "declared policy" of Statistical Science applies to papers that establish the veracity of the Torah codes, but not to those that challenge it! Incidentally, this letter shows that the editors of Statistical Science were well aware of these two experiments performed by Witztum, yet condoned their omission from the MBBK paper. They also appear not to have had any objection to MBBK's statement that "Nothing we have chosen to omit tells a story contrary to the story here".

In response to this unabashed bias of Statistical Science, four independent mathematicians wrote a letter to the editor of Statistical Science in which they say, "It seems elementary that Witztum et al. should have been asked for their response BEFORE²⁵ McKay et al. was sent to referees". They go on to say, "The secrecy under which the whole process took place up to now is not worthy of a top flight journal like Statistical Science". These mathematicians summarize their position as follows: "Allow us to emphasize very strongly that we are not taking any position on the substance of the accusations. We also agree that accusations of fraud may in principle be published, and indeed should be published – if correct. All we say is that if you do publish them, you should be certain that they are indeed correct AND²⁵ that you have followed equitable and reasonable procedures". This letter was dated July 5, 1999, about two months before the May issue of Statistical Science containing the MBBK paper appeared in print. The

²⁴ This quote is taken from a letter written by four mathematicians, which will be referenced shortly.

²⁵ Their emphasis.

letter made no difference at all!

The critics have raised many objections to the Torah codes, yet, one irrefutable argument would have sufficed. It is reminiscent of Hamlet's remark that "The lady doth protest too much". Hopefully, this paper will aid the truth seeker in seeing through the smoke screen and obfuscation to the truth.

Acknowledgements

I would like to thank Rabbi Moshe Zeldman, as well as those who wish to remain anonymous, who read the text of this paper in detail and offered many useful suggestions. I am grateful to Doron Witztum and Professor Eliyahu Rips for providing me with the documents that are included in the appendix. My special thanks go to Professor Robert Haralick for writing the foreword, offering insightful suggestions and for his permission to include his review of the book by Dr. Ingemanson as an appendix. I would also like to thank Rabbi Dovid Lichtman for permission to include his article as an appendix. Finally, I most grateful to Rabbi Moshe Heinemann who has given me his valuable time, moral support and encouragement for my work on the codes in the Torah and has urged me to write this primer.

APPENDIX A

THE ACCURACY OF OUR WRITTEN TORAH

Rabbi Dovid Lichtman

Our Torah scroll is perhaps our most revered physical possession today. The honor and respect with which we handle our Torah in synagogue results from our knowledge that it contains the words of Hashem as dictated to Moshe over 3300 years ago. Meticulous care has been taken to insure the proper transmission of the Torah. There are many factors which collectively contribute to the wholeness of the Torah, but perhaps the single most important factor is the orthography, or proper spelling of each word. In fact, the orthography of the Torah is considered so important that the scribe is instructed to "be careful with your task, for it is sacred work; if you add or subtract even a single letter, [it is as if] you have destroyed the entire world!" (Eruvin 13a). The Rambam writes (Hil. Sefer Torah 7:11) that if one letter is added to or missing from a Torah, it is invalidated and is not conferred the sanctity of a Torah scroll. Special mechanisms were established by the Sages to ensure its accurate transmission through the generations (see, for example, Megilah 18b; YD #274). (From the wording of the Rambam, it appears that this is true even if the wanton letter does not affect the meaning of the word. This is also the ruling of the Tikunei ha'Zohar (#25), Ramban end of Introduction to the Torah, Magen Avraham and Vilna Gaon OC 143:4, Sha'agat Aryeh (#36), Chatam Sofer (OC #52), in contrast to Minchat Chinuch's ruling (#613) that a missing or additional letter does not invalidate a Torah scroll unless it affects either a word's pronunciation or its literal or exegetical meaning.) Originally, the Torah was so well preserved that every letter was counted (Kiddushin 30a), which is why the early scribes were given the title "Soferim" ("Counters/Scribes"). Thousands of traditions were handed down specifying orthographic details. One of the more well-known is that the letter 'Vav' of the word 'Gachon' Parasha Vayikra (11:42) is the middle letter of the Torah (Kiddushin, ibid. -- refer to Rabbi Kornfeld's "Torah from the Internet" p. 122 for an in-depth discussion of this and similar traditions.)

Indeed, the text of today's Torah scrolls the world over are uniform, with very few exceptions. As we will demonstrate, the Mesorah (transmitted tradition) of our text was well tended to; its margin of error appears to be less than .00004, and to involve only insignificant letters at that. However, upon investigation it is evident that there existed many variants among older Torah scrolls. This prompts us to ask a number of questions: (a) First, one must ask how it came to be that there existed such diverse texts. Did they derive from individual copyists' errors, or were there differing Mesorot? (b) Second, one must ask how we came to accept at present one text as "correct" from among the many that once existed. (c) Third, can we have any degree of certainty that the present day unified text is the accurate text of the Torah as transmitted to and transcribed by Moshe? In this essay, we will attempt to address these questions.

II

Originally, it was easy to attend to the Mesorah of the Torah text. A Torah scroll written in Moshe's own hand was kept in or near the Holy Ark in the Holy of Holies (Bava Batra 14a). This text, which apparently was accessible to the Kohanim (Rashi Bava Batra 14b s.v. Sefer; see also Tosefot, Bava Batra 14a s.v. Shelo), undoubtedly served as the proof text for all other texts. The scroll which each Jewish king was required to write and bear at all times was likewise copied from this scroll (Rambam, Hil. Sefer Torah 7:2, based on Yerushalmi Sanhedrin 2:6). The kingly scrolls, in turn, served as proof texts after their owner's death.

The destruction of the first Beit ha'Mikdash most likely brought with it the destruction of these proof texts. Ezra the Scribe, who led the people back to Eretz Yisrael and began to rebuild the Beit ha'Mikdash, set to reestablishing a proof text. At this point, a defining event occurred. According to the Talmud Yerushalmi (Ta'anit 4:2), three ancient scrolls were found in the Temple confines which had slightly variant texts. (Although the Yerushalmi does not specify when this occurred, other sources relate that it happened in the days of Ezra and according to some versions, it was Ezra himself who found the scrolls -- see Torah Sheleimah, Shemot 24:25.) The Yerushalmi then relates that the correct version of the Torah was determined by virtue of a majority of 2 against 1.

Throughout the period of the Second Beit ha'Mikdash, a scroll referred to as 'Sefer Ezra' or 'Sefer Ha'azarah' (Moed Katan 18b) served as the standard for all others. Sefer Ha'azarah was either the very scroll that was written by Ezra the Scribe or one that was copied from it (Rashi, ibid.). Professional Soferim were employed at the Beit ha'Mikdash to correct private scrolls based on this scroll (Ketuvot 106a; Shekalim 10b). These highly accurate scrolls and their copies remained the standard until well after the destruction of the second Beit ha'Mikdash. The Talmud in Kiddushim (30a) establishes that the accurate counting of the letters of the Torah was preserved at least until Tanaitic times (2nd century CE).

III

A century or so later, in the times of the Amora'im, Rav Yosef commented that this accuracy was already somewhat diluted. Such a lack of accuracy can only have been made apparent by the existence of divergent texts. The Gemara makes it clear that even this dilution of accuracy was only with regard to Malei and Chaser. (Malei and Chaser refer to unpronounced letters, such as 'Vav' and 'Yud,' which lend added accent to vowels. Their presence or absence does not affect the meaning of a word). Nor does the Gemara state in how many instances doubts arose regarding orthography. It is possible that these uncertainties were limited to a very few instances. In fact, nowhere in the Talmud or Midrashic sources is there recorded a dispute over the orthography of a specific Malei or Chaser, either before or after the time of Rav Yosef. (It should be pointed out that according to some, Rav Yosef was merely stating that *he* could not determine the exact number of letters in the Torah, since he himself was blind and could not count them by heart and he was not willing to rely on another person's count -- see Rav Reuvain Margulies in "HaMikra V'HaMesorah," #4).

Due to the dispersal of the Jewish people and the lack of a central supervising authority, variations in scrolls continued. Authorities in Israel and Bavel, independently, undertook to produce one highly accurate text. These authorities, called the Masorites, thrived and produced such works between the 8th and 10th centuries. Their methodology, which was based on the system described by the Yerushalmi Ta'anit (above, section II), may be called the "eclectic process," or majority rule. Simply stated, this process involves surveying a great variety of Torah scrolls whereby each letter of the text is compared and contrasted. The correct orthography is determined based on the majority of texts, and hence errors are weeded out. For example, if in a survey of 200 Sifrei Torah, 198 were found to have in one particular place a spelling of "honour" and 2 were found to have the spelling as 'honor', it may be assumed that the former is the correct orthography, while the latter were introduced by careless scribes. (Of course, the eclectic process can only be employed using older texts of good standing to some degree. This is evident from the fact that only the three scrolls found in the Temple confines were considered for the process, in the time of Ezra. After all, certainly hundreds of scrolls were in existence at the time.)

The crowning jewel of the master texts produced in this manner was the one produced in Teveryah by Aharon ben Moshe ben Asher (known simply as "Ben Asher") of the late 10th century. The Rambam extols his text as being extremely accurate and it was adopted by the Rambam and many others as the standard (Rambam, Hil. Sefer Torah, beginning of 8:4). In the Rambam's time, this Torah was known to be in Alexandria, Egypt. (Traditionally, the "Keter Aram Tzova," or Aleppo Codex, presently in Yerushalayim, is purported to be the Ben Asher manuscript. Unfortunately, only the Nevi'im and Ketuvim sections of this manuscript remain intact, as virtually the entire Torah section of the manuscript was lost to fire a few decades ago.)

Today, the Teimani (Yemenite) Torah scrolls are very likely exact copies of this text. It is well known that the Yemenite Jews adhered firmly to the Rambam's rulings in every matter of Halachah. The limited size and dispersion of their community throughout the generations made it much easier for them to preserve their Mesorah. Indeed, there is no variance among Teimani scrolls today.

Despite the Rambam's efforts to ensure the perpetuation of one standardized text, divergent scrolls began to propagate once again. A contemporary of the Ramban, the RaMaH (Rav Meir Halevi Abulafia -- early 13th century), undertook to reestablish a text of exceptional accuracy. The RaMaH again used the eclectic process, surveying hundreds of old and reputable scrolls. (RaMaH did not have the Ben Asher manuscript at his disposal.) The resultant text was published in his work "Mesores Seyag la'Torah." Given the great effort that RaMaH invested in this project and his standing as a leading Halachic authority, his work became the definitive standard until today, certainly with regard to orthography (see Ohr Torah, Minchat Shai and Keset ha'Sofer).

We have thus answered the first two of our questions: (a) Since a standard, approved Mesorah for the Torah text existed throughout much of our history, in all probability the variant texts of early Torahs may be attributed to sloppy copyists, who did not carefully

compare their work with the Masoretic proof-text of the times, or were not able to do so. (b) The manner in which the mistaken texts were weeded out from the correct ones was the eclectic process of the Yerushalmi in Ta'anit, which has been employed regularly since the time of Chazal in order to ensure proper transmission of the Torah.

IV

(c) However, we have not yet addressed our third question: Can it be scientifically demonstrated that our text is indeed the correct one (i.e., that the eclectic process worked)? Halachically, we are secure in our reliance on the eclectic process (Teshuvot Ginat Veradim 1:2:6). This does not mean, though, that our Mesorah is 100% in agreement with the original text that was handed to us by Moshe. It only means that we are doing our best and are following the dictates of Halachah in determining how to write our Torahs. In fact, many authorities write that our texts may very well not match up with the true Mosaic text (authorities in OC 143:4, Sha'agat Aryeh, Chatam Sofer and Minchat Chinuch cited at the beginning of section I, see Hagaon Rav Moshe Sternbuch in "Mitzvat ha'Yom," pp. 32-43, who discusses the Halachic aspects of this statement in detail.). But does that mean that our texts may be "wildly inaccurate", or that "one or two" discrepancies may exist? Or, returning to our first question, can it be proven that enough attention was given to preserving the Mesorah and that copyists' errors were usually nipped in the bud before assuming the part of "Mesorah?" Or did too long a time pass between Masoretic overhauls, and many errors became independent Mesorahs over the years? (This theoretical question has been brought to the forefront in recent years by the great Torah Codes debate.) An exercise regarding this very question has been conducted by Dr. Mordechai Breuer of Yerushalayim, with fascinating results.

In his work, "The Aleppo Codex and the Accepted Text of the Torah", Dr. Breuer describes his years of meticulous research and discusses his conclusions in attempting to demonstrate the scientific usefulness of the eclectic process. In fact, Dr. Breuer's purpose was to demonstrate that a single Mesorah already existed in the years prior to the RaMaH, even though the RaMaH did not have such a Mesorah at his disposal. (The existence of such a single Mesorah is flatly rejected by many academicians.) Dr. Breuer began by selecting four texts of ancient origin to compare and contrast in his study. Each of these texts predates the RaMaH. The texts were all of the type written by the Tiberian Masorites (as opposed to the Babylonian Masorites) yet clearly differed from each other in certain significant formatting areas, indicating that they were not copied from an immediate common source. In addition, he included the text of the Mikra'ot Gedolot of Yaakov ben Chaim, printed in Venice, 1525. (It should be noted that the orthography of these 5 texts differed widely from one another, in one case by more than 200 letters from the others.) Using the eclectic process, he suggested that if a broad majority of 4 out of 5 texts (and not just 3 of the 5) agreed with each other, it could be assumed that the fifth, inconsistent text was a copyist's error. His results were startling. There are 304,805 letters in the Torah. All five texts were in **total** agreement in all but about 220 letters. Of these, all but **20** were resolved by a majority of at least 4 texts against 1! Of the 20 remaining conflicts, Dr. Breuer was able to clarify all but **6** by applying another Masorite method, that of carefully studying thousands of early Masoretic notes (a broader topic similar in style to the eclectic process). These final 6 he

was not able to clarify because three of the Torahs presented one spelling, while the remaining **two** presented another. It was apparent that nearly all of the inconsistencies between the Torahs were caused by copyists errors, and not by Masoretic uncertainties.

Next, the resultant 'eclectic' text was compared with the RaMaH's text (i.e., our present text). It was found that the RaMaH differed in but **6** places from the eclectic. That is, the margin of uncertainty of our Torah scrolls is probably not more than 12 (out of 304,805!) letters -- the 6 indeterminate ones, plus the six in which the RaMaH's text differed from Dr. Breuer's eclectic! When he compared the results of his experiment with the Teimani text (which, as we mentioned, is probably identical to that of Ben Asher), the results were even more startling. The texts were in perfect agreement! Their margin of uncertainty may be no more than 6 letters! Equally amazing is that **all** the above mentioned differences involve Vavs and Yuds, which do not affect the meaning of the word at all. (As for the remaining six uncertainties in Dr. Breuer's eclectic survey, in three of the instances the RaMaH and Teimani texts agreed with the 3-against-2 majority text. In the other three cases, the RaMaH and Teimani texts were themselves split over the same variant spellings as were the pre-RaMaH texts. In total, that means that the Teimani text differs from the RaMaH's text in but **9** letters -- see endnotes for details.)

In conclusion, the transmission of our Torah text has been well tended to and well preserved. The methods of Chazal have proudly withstood the tests of time. Such demonstrations of the strength of our Mesorah are indeed a Kiddush Hashem.

The author welcomes your comments on the above article:

ENDNOTES: Torah variants of Dr. Breuer's results, as compared to our (=RaMaH's) Torahs, in order of appearance (E=eclectic; T=Teimani): (1) Bereishit 4:13 "Mineso" (E&T w/o Vav); (2) Bereishit 7:11 "Ma'ayanos (E&T w/o Vav); (3) Bereishit 9:29 "Vayehi" (E&T Vayiheyu); (4) Bereishit 46:13 "v'Shimron" (E with Vav); (5) Shemot 14:22 "Chomah" (E w/o Vav); (6) Shemot 25:31 "Te'aseh" (E&T w/o Yud); (7) Shemot 28:26 "ha'Efod" (E&T w/o Vav); (8) Bamidbar 1:17 "b'Shemot" (T w/o Vav); (9) Bamidbar 10:10 "Chodsheichem" (T with Yud); (10) Bamidbar 22:5 "Be'or" (T w/o Vav); (11) Bamidbar 33:52 "Bamotam" (E w/o Vav); (12) Devarim 23:2 "Daka" (E&T with Alef instead of Heh. Lubavitch Chassidic texts are in agreement with T in this matter).

APPENDIX B

An English translation of
APPROBATIONS OF LEADING RABBIS AND SAGES

Our acquaintance with Mr. Doron Witztum goes many years back. He studies Torah, and in recent years has engaged in researching the hints found in Torah at equidistant intervals. The results of one of these studies, which he performed with his friend Eliyahu Rips, a math professor, was published in a prominent scientific statistical journal. Lately the above researchers were accused of doctoring and dishonestly portraying the results of their study with Rav Prof. Shlomo Zalman Havlin, their advisor regarding Torah bibliographies. Their opponents claim that the list of names and titles of Torah leaders which Prof. Havlin prepared for the purposes of the study were forged so as to insure highly successful results. According to their claim, the fraud is obvious since the rules according to which Prof. Havlin worked (and which were publicized by him) are unconvincing, particularly since in many of the cases, the composition of the lists contravened the rules themselves. We are not expert in the science of statistics, but the claims mentioned above are unrelated to this wisdom. The question is if an act of fraud was perpetrated or not, and regarding this, our view is definite. We hereby incontrovertibly affirm:

A. Rav Witztum and Prof. Rips are well known to us as decent and upright men of truth, to whom few in Israel can compare regarding their integrity and aversion to falsehood. It is abominable that such individuals are being accused of fraud and deception.

B. We investigated into Rav Havlin's character and discovered that he too is known as a reputable, decent and upright person whose trustworthiness is not under question, particularly in his area of competence and especially in so transparent a matter.

C. We checked the rules according to which Prof. Havlin formulated his list of names and titles of Torah leaders, and we found that it was commensurate with both professional standards and common sense.

D. The list is in keeping with the principles. We found that all the opponents' individual claims concerning deviations from the principles are false, and are a testimony to their glaring ignorance and unfamiliarity with the subject.

In the light of the above, we hereby affirm that the work of Rav Doron Witztum, Prof. Eliyahu Rips and Rav Prof. Shlomo Zalman Havlin does not contain an iota of fraud or deception, and the claims of their opponents are a reprehensible libel. Whoever assists them, will face judgment one day.

Hereby confirmed on Elul, 5758
Boruch Shmuel Hakohen Deutsch Shlomo Fisher

I wish to add my testimony to the signatories above, who are distinguished Torah leaders, and whose words may be relied upon. Although I am not acquainted with the issue, it is clear to me that all those who cast aspersion on the Torah's hints, which are deduced through letters of equidistant intervals, are doing a great injustice. Those observant Jews who join them are partners in their impure intentions, and will face judgment one day. May the Almighty give all of them a new change of heart.
Who signs for the honor of heaven,

Shmuel, son of the Gaon Rav Shlomo Zalman Auerbach

2 Marcheshvan 5759

It is known that a way exists to discover hints and matters from the Torah by reading letters at equidistant intervals. This method is found in the commentary of Rabeinu Bechai on the Torah and the works of Rav Moshe Cordovero. More recently, the tzadik, Rav Weissmandel, revealed wondrous things with this method. To my surprise, I have heard that opponents to this method have arisen claiming that various deceptions were performed by those who are involved in this method today. It is astonishing to me that they were not intimidated to state their claims after the Gaon Rav Shlomo Zalman Auerbach gave his clear agreement to the method of equidistant intervals, and after Rav Auerbach's son, the Gaon Rav Shmuel, censured them. I am uncertain of these opponent's intentions. Are they troubled that at seminars to make people religious, that these astonishing matters are sometimes taught, and their effect on the listeners is profound? Is this what is bothering them? Whatever the reason of those who oppose this method, it is certainly not a dispute for the sake of heaven, and we must strengthen those who are engaged in this method, for it is a totally honest endeavor. May they be encouraged and see blessed results to their deeds, and may they continue to increase the honor of our holy Torah and its influence on all who seek the Almighty's closeness. May the Almighty help those who engage for the sake of heaven in this honest discipline of studying equidistant intervals,

Shlomo Wolbe

When more can I add in writing to the view already mentioned by Torah leaders that there is a procedure and truth to the method of equidistant intervals. Moreover, the findings of this method are used only as a way to reinforce and increase wonderment at the divinity of the Torah's wisdom, and not as a proof that the Torah is from heaven. It is well known that our faith is built on a reliable tradition, which our ancestors saw with their own eyes and heard with their own ears, and publicizing this method has merely reinforced this impression.

Mattisyahu Chaim Salomon

Translated by:

Miriam Samsonowitz, Studio Har Chotzvim

Translations between Hebrew and English and 20 other languages

POB 45143, Jerusalem, Israel 91450

Tel: +972-2-587-0790 Fax/Modem: +972-2-587-0950

With the help of G-d, Erev Shabbos Kodesh, Parshas Chukas – Balak, 5759

Two “Tzanteri D’dahava” (Aramaic; literally, “Golden vessels”, i.e., people of impeccable character), may they live long, came before me and set before me the allusions in the Torah by way of ELS in the Torah. Although I am not adept in the discipline of statistics, what has clearly emerged from their words, was that there was no trickery or dishonesty on the part of these who search for the Word of Hashem and whose will is only to uplift Torah and to exalt her.

The sources in the words of our antecedent sages that there is in fact this way to find allusions in the Torah is known. Thank G-d, we in our days have been privileged to the means to find many more [such allusions] for his Namesake.

There surely are those that will falsify and misconstrue and about them the verse says, “the righteous shall go in [the ways of Hashem] and...” but this is no reason not to use the aforementioned allusions to open the blind eyes of many of our blinded Brethren of B’nei Yisroel. Maybe through these things they will return to drink from the Living Waters of our holy torah.

Let the strength and power of those who bring the hearts of K’lal Yisroel back to their father in heaven be strengthened and we should all merit the Light of Moshiach speedily in our days.

With blessings of success,

Shmuel Kamenetsky

Translated into English by Rabbi Sholom Kamenetsky