Document 4: An Analysis of the Decisions in Our Research

Introduction:

The Nov. '97 of Galileo (an Israeli popular-science magazine) carried a critique of our research The authors of the critique, the psychologist Prof. Bar-Hillel, and the mathematicians Dr. Bar-Natan and Dr. McKay, raised the question of the integrity of our work: Was the original research conducted honestly and sincerely? Their analysis was as follows: After all, "any researcher in any investigation is obliged to make various kinds of decisions regarding his experimental methodology" - our original research also required decisions of this kind. If the decisions were made a priori, that is to say, without knowing in advance how they would effect the outcome, then one would expect the number of "beneficial" choices (those which improve the results) should just about equal the number of choices which were "not beneficial" (those which have a deleterious affect on the results). "Yet surprise, surprise, it turns out that in almost every case, if not in every case, their supposedly blind choices paid off." Their conclusion: The chances that the original research was conducted honestly and sincerely are "extremely slim."

Yet a closer examination of their analysis reveals that it was carried out in a way which was defective in the extreme. When one examines their analysis of our choices one reaches exactly the opposite conclusion!

Any analysis, done in order to reveal whether there was any bias in the decision-making process, must adhere to the following criteria:

1. Obviously the one conducting such an analysis must guard himself against bias. Therefore he must make himself a list, a priori, of the "choices" he is going to investigate, before he investigates whether the choices were "beneficial" or not.

Then, in order for his conclusions to be persuasive, he must present all of the "choices" which he investigated. Otherwise, there is no value whatsoever in his declaration that "such-and-such choices were 'beneficial' and such-and-such choices were 'not beneficial'."

2. The evaluation must be made only for those choices which were not dictated by the nature of the research. For example, when Galileo decided to investigate whether Jupiter has moons using a telescope instead of a microscope, this was a decision which was dictated by the needs of the investigation. Similarly, it would be somewhat ludicrous to level a charge of bias against a researcher investigating which company produces the most popular steak for not having chosen to conduct his survey among a village of vegetarians.

The authors of the critique failed to fulfill either one of these conditions. A full analysis of their "analysis" is presented below. Here let it suffice to say that on two previous occasions the authors of the critique presented lists of the "choices" they investigated. In the winter of '97 Prof. Bar-Hillel presented a list of 13 choices at a convention held by the Center for the Research of Rationality (of the Hebrew University). She claimed that every one of our 13 choices were "beneficial." But mathematician Prof. Robert J. Auman presented in response a totally different analysis: Of the 13 choices only one was "beneficial." The rest were either choices dictated by the needs of the research (for example, the "choice" to use correct dates rather than incorrect ones!), or choices which were "not beneficial."

In the first chapter, below, we will discuss this at length. Here we will mention just two examples of decisions cited in Prof. Bar-Hillel's paper. One example of a choice which was "not beneficial" was the choice not to use the form of the Hebrew date בא בתשרי(b'alef b'Tishrei). To our sorrow Yaakov Auerbach, is no longer with us to explain his reasons for making this choice, but one thing is clear: If we had used this form it would have improved the results both for the first list as well as for the second. Therefore it was "not beneficial" to have omitted it. How strange it is that this example was cited in the critique as an example of a "beneficial" choice!

A second example: The second list included Rabbis whose entries in Margalioth's Encyclopedia were from 1.5 to 3 columns. They claimed that we erred in measuring the column lengths, and in their opinion we should have omitted and added certain personalities. It turns out that had we operated according to their recommendations, the results would have improved! At the request of Prof. Auman we performed this calculation using the permutation test, as well, and the results improved from 4 in a million to one in 5 million!

On Apr. 22, '97 Dr. Brendan McKay published on the Internet a report of an investigation of our research. In this report he lists 20 choices which he investigated (principally regarding the method of measurement). In Chapter 2 we will deal with his assertions at length. Here, we will just note that even if we ignore the fact that some of the "choices" he mentions were completely imposed by the requirements of the research, it turns out that at least 12 choices out of 20 (by his own calculation) were "not beneficial" for us. One was particularly "not beneficial": If we had made a slightly differently decision the results for the first list would have improved by a factor of approximately 1,000, and the results for the second list would have improved by a factor of more than 1,000!

Summarizing the 20 decisions, it turns out that:

A. 3 of the decisions were dictated by the nature of the research: 3, 14, 19.

B. For 12 of the decisions the results would have improved had we chosen otherwise: 1, 2, 6, 7, 11, 12, 13, 15, 16, 17, 18, 20.

C. 2 of the decisions would not have effected the results at all: 4, 9.

D. In only 3 of the decisions would the results have turned out poorer had we made a different choice: 5, 8, 10.

Here we have an accounting which indicates clearly that our decisions were made objectively, a priori and without bias!

It seems rather strange that of all these investigations, which they themselves conducted, there is no mention in their critique! In Chapter 3 we will see that a number of additional choices, cited by the authors in their attack on Prof. Havlin's guidelines and their execution, also turn out to have been "not beneficial" for us.

I. BAR-HILLEL'S CHOICES

At a session of the Center for the Research of Rationality, which took place in the winter of '97, Prof. Bar-Hillel presented a list of 13 choices which

she asserts we made in the course of our original research. According to Prof. Bar-Hillel all 13 of the choices she presents were in our favor. In each case there were two possibilities to choose from, so the chances of making a favorable choice should have been 1/2. The chances of this occurring 13 times would be 1/2 to the power of 13, that is, a probability of one in eight thousand (approx.). Therefore, she concludes, that is the probability of our having made these decisions a priori.

As we noted in the Introduction, one indispensable criterion which must be met in an analysis of decisions, is that the researcher acknowledge all of the decisions investigated.

No one knows how many decisions Bar-Hillel really investigated: But at any rate, we have seen that in her defective report in Galileo she somehow forgot to mention those decisions which indicate that our research was indeed a priori. That being the case, how much more so is the burden of proof upon her to show that the list of choices she originally presented was compiled a priori. If it was not - then her entire report becomes totally irrelevant!

Decision 1: Concerning the first list: When an incorrect date is cited in Margalioth's Encyclopedia we chose to replace it with the correct date, rather than use the wrong one.

Response: The research hypothesis was that conceptually related concepts tend to converge in letter-skipping form in the Book of Genesis. The experiment was designed to test this hypothesis by investigating convergences between famous personalities and their dates of birth and death. Obviously there is no connection between a personality and the incorrect date of his birth or death!...

This is a classic example of decision which was dictated by the nature of the research (see criterion 2 in the Introduction). According to the research hypothesis we would naturally anticipate that correct dates would prove more successful.

Decision 2: The same as 1, but for the second list.

Response: The same as the Response to 1. But here the attack is even more surprising: After all, this decision was made for the sake of compiling the first list, and as Bar-Hillel notes in Galileo, all the decisions were adhered to without modification in the compilation of the second list as well. In other words, for the second list there was no room for decision at all!

Decision 3: Concerning the first list: When an incorrect date is cited in the Encyclopedia, we chose to replace it with the correct date, rather than omitting the date altogether.

Response: This is another classic example of a decision which was dictated by the nature of the research: After all, we had to have dates in order to

establish name-date word pairs, in order to test our research hypothesis!

Decision 4: The same as 3, but for the second list.

Response: The same as the Response to Decision 2.

Decision 5: Concerning the first list: We chose to use both the date of birth and the date of death. In her opinion we could have chosen to use only the date of death.

Response: We are of the opinion that dates of birth and dates of death are of equal relevance, and there is no justification for omitting either one. According to the research hypothesis, the more pairs available, the more successful the result should be. Therefore this is another example of a decision dictated by the nature of the research.

Decision 6: The same as 5, but for the second list.

Response: See the response to Decision 2.

Decision 7: Concerning the first list: We chose to indicate the 15th and the 16th of the month in two forms: tet"vav and yud"hey, tet"zayin and yud"vav. On the other hand, we could have chosen to use only the forms tet"vav and tet"zayin.

Response: Following the standard convention, after yud"gimmel, yud"dalet (13^{th} and 14^{th}) should come yud"hey, yud"vav, (15^{th} and 16^{th}) except that these latter two forms consist of letters from the Divine Name of G-d. Therefore their use is avoided. In their place it is customary to use the forms tet"vav for 15^{th} and tet"zayin for 16^{th}.

In the Torah itself, however, the Divine Name occurs many times. Therefore if we are looking for dates appearing in the Torah, there is no reason to alter the form yud"hey to tet"vav! Thus the research hypothesis dictates that we must investigate this form, which is why Prof. Rips suggested that we do so.

Decision 8: The same as 7, but for the second list.

Response: See the response to Decision 2.

Partial Summary: These eight decisions actually boil down to the four decisions regarding the first list: When it came to compiling the second list there was no longer any decision to be made. Furthermore, all four were dictated by the nature of the research. It is somewhat surprising that Prof. Bar-Hillel ignores this fact.

Let us continue with the remaining decisions:

Decision 9: A decision was made not to use the form of the date "be'alef be'Tishrei", in addition to the other three forms: "alef Tishrei", "be'alef

Tishrei", and "alef be'Tishrei".

Response: In Dec. '96, more than two weeks before the session at which Prof. Bar Hillel presented her report, she received our response to this issue in writing. (See Document 2: "Bar-Hillel and Bar-Natan Ask; Witztum and Rips Respond"). I quote from this document:

Indeed, we did not use the form בא בתשרי in our list of dates. We were made aware of this by your comments, as well as by those of one other person. To our sorrow, we are unable to ask the linguist Yaakov Orbach, of blessed memory, why he did not include this form in his recommendation. However, in order to remove any suspicion, we hereby declare that the list of dates was prepared by Yaakov Orbach, exactly as stated in the first preprint, before the experiment had been carried out on the first list. We used the exact same forms of the dates with regard to the second list.

If anyone still suspects that some hidden motive lay behind the omission of the form בא בתשרי, in order to improve our results, we invite him to consider -- as we did when we first heard this criticism from Dr. Dror Bar-Natan -- what would have happened had we included the form בתשרי. בא Recall that the only measures of success which were used regarding both the first and second lists, as was stated in the first preprint (the "White Preprint") , the second preprint (the "Blue Preprint"), as well as in the preprint which was originally sent to PNAS, were the over all probability figures: P1 and P2. The randomization test of Professor Diaconis was suggested at a later stage.

(i) The results which were calculated for the first sample were:

P_1 = 0.000000001334 and P_2 = 0.00000000145.

If we had used the form בא בתשרי as well, the results would have been:

P'_1 = 0.000000000349 and P'_2 = 0.00000000207.

In other words, the best result would have improved by a factor of 3.8.

(ii) The results which were calculated for the second sample were:

P_1 = 0.0000000331 and P_2 = 0.00000000201.

If we had used the form בא בתשרי the results would have been:

P'_1 = 0.00000000507 and P'_2 = 0.00000000171.

In other words, the best result would have improved by a factor of 1.18.
[end of quotation]

In other words, it is clear from here that the form in which the date was to be presented was established a priori, and that the decision not to include the form "be'alef be'Tishrei" was in fact not beneficial. It is very strange that even after her case had been refuted, Bar-Hillel continued to cite this example, not only at the conference, but months later in the Galileo article.

Decision 10: The same as 9, but for the second list.

Response: The same as the response to 9, with the additional point that for the second list there was no longer any decision to be made (just as with decisions 2, 4, 6 and 8).

Decision 11: Concerning the first list: We erred in estimating the length of the Encyclopedia entry for one of the personalities on the first list (it was a line short of the requisite 3 columns). Prof. Bar-Hillel counts this as a decision, vis-a-vis the possibility of having prepared an accurate list without such an error.

Decision 12-13: Similar charges arose concerning the second list except that, as Prof. Robert J. Aumann noted in a letter he wrote to Prof. Bar-Hillel (Jan. 17, '97), "In this case Maya examined two rather complicated alternatives, but did NOT (!) examine the alternative of simply using the correct second list."

Response to 11-13: On Dec. 25, '96, more than two weeks before the conference in which Prof. Bar-Hillel presented her case, she received our written response to these charges. (see Document 2: "Bar-Hillel and Bar-Natan Ask - Witztum and Rips Respond"). I quote from this response:

. We were asked why we included R. Aharon of Karlin, R. Yehudah Ayash, and R. Yehosef Ha-Nagid in the second list, despite the fact that their entries in Margalioth's encyclopedia are, in the opinion of the inquirers, less than a column and a half. Similarly, we were asked why R. Meir Eisenstat was not included in the second list, despite the fact that his entry is, in their opinion, exactly a column and a half. We were also asked why R. David Ganz was included in the first list rather than the second, despite the fact that in their opinion his entry is just short of three columns.

The answer is simple: The inquirers measured the size of the entry by counting lines. On the other hand, when Doron selected the 34 personalities, he did so (as best as he can recall) by a visual estimate. This is how he selected the 32 personalities of the second list, as well. As it turns out, it was an error in judgement to rely on a visual estimate, and the measure used by the inquirers is a better one.

If anyone suspects that this was done in order to improve the results through manipulation, let him examine the results of the following experiment, which we carried out as soon as we were made aware of this complaint:

(i) We recalculated the results of the first sample omitting R. David Ganz. Let us compare:

The results we originally received were:

$P_1 = 0.000000001334$ and $P_2 = 0.00000000145$.

If we recalculate, omitting R. David Ganz we receive:

$P'_1 = 0.00000001336$ and $P'_2 = 0.00000000276$.

In other words, the best result became worse by a factor of 2.07.

(ii) We recalculated the second sample, omitting R. Aharon Karlin, R. Yehudah Ayash and R. Yehosef Ha-Nagid, and adding in R. David Ganz and R. Meir Eisenstat (the names and appellations which were used to refer to him were delivered to us by Professor Havlin on the 22nd of December, '96. They are: רבי מאיר אזנשטט, מאיר איזנשטט, מהר"ם א"ש , "פנים מאירות", בעל "פנים מאירות", מאיר איזנשטאט, מהר"א איזנשטאט). For the sake of comparison, here again are the results of the second sample:

$P_1 = 0.0000000331$ and $P_2 = 0.00000000201$

If one recalculates, incorporating the changes mentioned above, one receives:

$P'_1 = 0.00000000422$ and $P'_2 = 0.00000000129$.

In other words, the best result improved by a factor of 1.56.

To summarize, one can clearly see that the results we would have received would have been of the same order of magnitude.
[End of quotation]

Clearly the error in the first list was indeed in our favor, but the errors in the second list were to our disadvantage to the same degree.

To sum up: Out of the 13 decisions which Prof. Bar-Hillel presented, she succeeded in identifying only a single choice which was in our favor!

The reader must be astonished: How could she have erred to such an extent?!

The answer is that her fundamental error lay in not distinguishing between arbitrary decisions, and decisions that are dictated by the nature of the research. Furthermore, when she evaluated the "benefit" of any particular decision she used a methodology which was only agreed to (between Prof. Diaconis and Prof. Aumann) after the relevant experiments had already been conducted. Therefore her calculations are inappropriate to an analysis of choices made in the course of the research. It is reasonable to suppose that had she evaluated the decisions correctly, we would never have heard the tale of the "13 Choices."

However, her work raises a more disturbing question: Since she evaluated the "benefit" according to this newer methodology, she should have been aware that using the corrected second list (Decisions 12-13), the results improve dramatically: Instead of a probability of 4 in a million, which we received in our original work, with the corrected list we receive a probability of one in five million! Why did this important fact slip her attention both at the conference of the Center for the Research of Rationality, as well as in the Galileo article? Why did she propose "two rather complicated alternatives," rather than the more obvious alternative of "simply using the correct second list"? (to quote from Prof. Aumann's letter mentioned above)

II. MCKAY'S DECISIONS

In a paper which appeared on Apr. 22 '97, Dr. McKay reports on 20 variations in the methodology of measurement. In his report he labels the experiments he conducted as V1 - V20. Below we will deal with each of the "choices" (the variations) he examined. The nature of the subject is such that much of the discussion concerning the methodology of measurement will be technical, and assumes that the reader is thoroughly familiar with the Statistical Science article. But before slipping into such a technical discussion, I would like to present one particularly outstanding and revealing example:

In our methodology every pair of expressions is given a numerical score between 0 and 1. If the convergence was successful, the value will be closer to

0. If the convergence was unsuccessful, the value will be closer to 1. In order to evaluate the "overall tendency of convergence" between pairs on the first list, we used two measures: P1 and P2. These figures were designed to measure whether or not there are "many" successful convergences. That is to say, are there "many" convergences whose value is small, approaching 0.

P1 is very simple: It is simply a count of how many convergences had values between 0 and 0.2. This is an easy way to calculate the probability. Imagine that everything is random. In that case the values of the convergences should be more or less evenly spread between 0 and 1. Therefore, out of 100 values we would expect about 20 of them to fall between 0 and 0.2. If the phenomenon is indeed random it would be very unlikely that we would find 40 values between 0 and 0.2 rather than 20. The probability of this happening can be calculated (using the binomial distribution).

For the first list there were 152 values of convergence. Of these, 63 fell between 0 and 0.2, rather than the approximately 30 which we would have expected to find in the random case. The probability of such an enormous deviation is extremely small: 0.000000001334.

Dr. McKay claims (rightly so) that there is nothing “holy” about the cutoff we selected, that is 0.2. We could have chosen a different cutoff, and he himself investigated a number of alternative cutoffs.

It turns out that according to McKay's own calculations, if the cutoff had been fixed at 0.33 the results would have improved by a factor of 1000, and if it had been established at 0.5 the results would have improved by a factor of 3300! (In fact, his calculations are imprecise, and for a cutoff of 0.33 the results would have actually improved by a factor of 2200, and for a cutoff of 0.5 by a factor of 7500 --- in other words, we would have reached a probability of one in 5000 billion!).

Thus in the experiment labeled V20 of his report, Dr. McKay has succeeded in demonstrating very nicely that we did indeed operate in a manner which was a priori and without bias. No one could have known what cutoff we originally established, or if we established a cutoff at all. A researcher wanting to improve his results would certainly not hesitate to choose a cutoff which would improve his results 7500-fold, and it would leave no "footprints"! It turns out that for the second list, as well, a change in the cutoff would bring about a dramatic improvement in the results. Dr. McKay himself made the calculations. (However, as I mentioned above, all the parameters established for the first experiment were preserved for the second experiment, therefore this is irrelevant to the evaluation of our decisions).

Let us now move on to the other investigations. According to the calculations I have made for purposes of comparison, it is clear that Dr. McKay's figures are not accurate. However, I do not have at my disposal the tremendous computer and other resources that he has, therefore I have not tried to replicate all of his results. In a few instances (which I will

indicate), I will rely upon my own calculations, but in general I will have to rely on the data appearing in his report. Here, too, I will relate to the decisions which were made for the first list, unless otherwise specified.

Variation V1: In the original experiment we established a cutoff for the size of the skip length, for which ELS's of a given expression would be sought. In each case the cutoff would be such that the expected number of ELS's would be 10. Dr. McKay investigated what would happen if instead of 10, it were established at 15, 20, etc. He reports that the he did not complete the experiment for a cutoff of 20. A pity. Otherwise he would have learned that the results in fact improve. And if he would have gone on and investigated a cutoff of 25 he would have discovered that the results continue to improve!

Variation V2: This time Dr. McKay investigated what would happen if we had defined differently the function "delta." It turns out that if the "distance" between one ELS and another had been defined thus: 1/distance squared (and not as we defined it in our article), the results would have improved by a factor of 30!

Variation V3: In our experiment we only examined expressions of 8 letters or less. The reason for this was simple: The value of the convergences was established by comparing convergences between expressions as ELSs with convergences of expressions in PLSs (perturbed letter sequences). Words of 9 letters or longer are much less likely to appear at random in the Book of Genesis. Therefore for expressions of this length we would expect there to be "no competition" - the expressions in non-equidistant letters would not show up for the race. For this reason our choice of cutoff was not at all arbitrary, and was completely justified by the research (see criterion 2 in the Introduction). McKay, on the other hand, arbitrarily created an artificial cutoff point at 7 letters, and even at 6. Thus he received fewer pairs on which to run the experiment, and naturally the results were poorer.

McKay writes that the data for expressions of 9 letters was "not yet prepared." Again a pity! This would have been the most interesting investigation. He would have discovered that for the first list there are no expressions of this length appearing as ELS's in Genesis! He might then have noticed that using our cutoff, all the expressions on the list which appear as ELS's, were able to be used in the experiment!

Variation V4: Here McKay examines another technical point: When we calculated the row length in a table of letters we rounded 1/2 up. McKay investigated what would have happened had we rounded it down. The result: No change! (By the way, he reports that for the second list rounding down would have led to slightly poor results. He erred in his calculations; the results in fact improve!)

Variation V5: Now he explores what would have happened if instead of using a summation of all convergences (TOTAL), we had used only the most successful convergence (BEST). He found that the results deteriorate.

Variation V6: In our calculation of TOTAL, the same row length was often included several times (due to rounding). He examines the alternative of using each row length only once. It turns out that according to his own calculations the results improve when this method is used.

Variation V7: He found that according to the program used in the original research, the minimal row length allowed was 2 letters, whereas in the article no such limit was imposed. He did not realize that this was simply a "bug" in the program, so he listed it as a "choice." Therefore he tried using a minimal row length of 1, and discovered that it made no difference in the results. He then tried using a minimum of 10 and discovered that the results improved!

Variation V8: He tried using a different definition of the "domain of minimality," and received a slightly poorer result. (I would eventually like to try verifying his figures).

Variation V9: Here he investigates a truly marginal issue, which he himself admits makes no practical difference, and the results are essentially the same.

Variation V10: McKay tried several alternative ways of assigning a "weight" to a convergence besides the method we used (he did not finish examining them all). If we can accept his figures, it turns out that for the first list the results deteriorate slightly (although for the second list they improve).

Variation V11: In this experiment he altered the PLSs are chosen. The results improved (for the second list as well).

Variation V12: He investigated what would happen if size of the PLS were altered. He found that when the set was enlarged the results improved (for the second list as well). Of course, if one reduces the set of PLSs, the success will be reduced.

Variation V13: In our experiment the function "sigma" was defined so that it was the summation of 10 tables. McKay experimented with various numbers of tables, and discovered that when "sigma" summed up only 5 tables instead of 10 the results improved dramatically!

Variation V14: In the original experiment we did not take into consideration the values of convergences which would have few "competitors" (PLSs)- we drew the line at a minimum of 10. There was a very good reason for this: Imagine that there exists a spectacularly successful convergence of ELS's in the Book of Genesis - a convergence in which the ELS's are not only close together, but "rare" as well, (i.e.that the probability of these ELS's themselves appearing by chance is extremely small). Because of this small probability it will be impossible to find PLSs for competition. They simply will not materialize. Only the ELS's will participate in the race, and the value of the convergence will wind up being

1/1 (since the number of competitors was reduced to one). According to our methodology of evaluating the results, a value of 1/1 represents total failure! (Remember, the closer a result is to 0, the more successful it is, and the closer to 1, the worse).

Even if one other competitor did appear, the outcome would still be an tremendous distortion of the results. 1/2 is a result which does not indicate success! In order to eliminate such distortions we placed a limit of no less than 10 competitors.

Dr. McKay simply did not understand this point, which is why he proposed setting the limit at 2 or 5 competitors. This is an example of a decision arrived at through a lack of understanding of the research. (By the way, it turns out that the effect on the results was small). It is interesting to note that when he experimented with limits of 15 or 20 competitors there was no change in the results.

Variation V15: He proposed setting a minimum skip distance of something more than 2. He discovered that for a minimum of 3, for example, the results improved!

Variation V16: McKay proposed using a different methodology for measuring the distance between PLSs. Here, too, the results he received were better than the original ones.

Variation V17: McKay investigated what would have happened in the permutation test, had we omitted the names of personalities for which there are no dates. (In our original experiment we did not omit them). This point is only relevant to the second list, because in our original research the permutation test was only run on the second list (also, as mentioned above in response to Bar-Hillel, the permutation test was only proposed after the second list of names and dates had already been compiled).
Nevertheless, once again McKay discovered that the best result for the second list improved!

Variation V18: McKay investigated what would happen if we were to include the form of the date "be'alef beTishrei". Once more it turns out that the results improve (for the second list as well)! (Cf. the Response above to Decision 9 of Bar-Hillel; regarding his proposal to use the forms he labels F5 and F6 see Document 2, sec. 4Biii).

Variation V19: He examines what would have happened had we not included the format yud"hey and yud"vav to represent the numbers 15 and 16. See above, the Response to Decision 7 of Bar-Hillel, where it is explained that this was a decision deriving from the research hypothesis.

Variation V20: This investigation was discussed above at the beginning of the section dealing with McKay's work.

A Summary of the 20 Investigations

Of the 20 investigations carried out by McKay, it turns out that:

A. 3 of the decisions were dictated by the nature of the research: 3, 14, 19.

B. In 12 cases the results would have improved had we made different decisions: 1, 2, 6, 7, 11, 12, 13, 15, 16, 17, 18, 20.

C. In 2 cases the results did not change at all: 4, 9.

D. In 3 cases the results would have been poorer had we made different decisions: 5, 8, 10.

The above summary cleary indicates that our decisions were made objectively, a priori, and without bias!

III. OTHER CHOICES NOTED BY OUR CRITICS

In this chapter we will deal with several other decisions which our critics have questioned, and which are mentioned in the documents posted at the internet site.

1. BNMK write that in Prof. Havlin's report: "Havlin acknowledges making many mistakes in preparing the list and says that if he were to do it again, he would have done it differently." A closer look at the Report (Document 1) reveals quite a different picture.

In the Report no such expression is to be found. What we do find is that in the section where Prof. Havlin explains his reasons for not including in the second list certain appellations which appear in the Bar Ilan Responsa database, he indicates a number of appellations which were left out inadvertently, or for which he could no longer recall the reason they were omitted.

We tallied these omissions and found that in all only 10 of the omitted names should have been on the list. That is to say, only ten of the omitted names were between 5 and 8 letters long. Of the ten, three do not appear as ELSs in Genesis at all.

We decided to investigate what would have happened if the remaining seven names had been included in the original list. Here are the results (recall that in the original experiment the statistics P1 and P2 served as the measures of probability. This is how they were presented in the "Blue Preprint" - the permutation experiment was suggested at a later date):

Originally, the best result was: P2 = 0.00000000201
If we add to the list the seven names which were omitted, we receive: P'2 =0.000000000101

In other words, the results improve by a factor of 20!

This should make it perfectly clear that Prof. Havlin did not omit these names in order to improve the result. Nevertheless, BNMK may have intended that it would be more proper to evaluate the statistical significance (using the permutation test) for Prof. Havlin's emended list. In response to this challenge we performed the permutation test with the addition of the seven names. In an experiment in which we ran 100,000,000 permutations, and P4 came in eighteenth place, that means that the probability is less than 1/5,500,000!

2. In his letter, Prof. Cohen asserted that an appellation associated with a certain personality should be used even if it is also associated with a different personality, and even if the other personality is more famous, rather than following the guideline established by Prof. Havlin. It must be reiterated that Prof. Havlin established this guideline before preparing the first list. As it turns out, the only time this guideline had to be invoked in compiling the first list was with regard to the acronym "Rivash" for Rabbi Israel Baal Shem Tov. Havlin decided to omit this acronym because it is the standard appellation for one of the "Rishonim"- Rabbi Isaac Bar-Sheshet. It so happens that if Prof. Havlin had acted according to Prof. Cohen's recommendation and used the appellation "Rivash" for R. Yisrael Baal Shem Tov, the results would have improved:

The results for the first list were: P1 = 0.000000001334 and P2 = 0.00000000145
With the addition of "Rivash" the results were: P'1 = 0.000000000412 and P'2 = 0.00000000117

That is, the best result improved by a factor of 3.24!

3. Prof. Cohen claims that we should have used all the variants of the appellations, and not just the most common and accepted ones as Prof. Havlin laid down.

When we read Prof. Cohen's criticism, it occurred to us to investigate what would have happened to the results of the first list if Prof. Havlin had not established this rule: That is to say, if he had included all related variants of the appellations, as well. As it turns out, the results would have improved!

The results for the first list were: P1 = 0.000000001334 and P2 = 0.00000000145
With the addition of related variants the results are:
P'1 = 0.000000000262 and P'2 = 0.00000000109

In other words, the best result would have improved by a factor of more than 5!

4. Several other examples of this kind can be found in the document "A Refutation Refuted."

Summary: Here as well, the facts plainly show that the decisions were made in a manner which was a priori and objective.