דורון ויצטום

סיגנל של מערכת צפני *ELS*: דעיכה מול הגברה

המאמר "Equidistant Letter Sequences in the Book of Genesis" מאת דורון ויצטום, אליהו ריפס ויואב רוזנברג (להלן: ור"ר),[1] אשר פורסם בכתב העת *Statistical Science* בשנת ה'תשנ"ד, הציג תוצאות מחקר מדעי של הרמז בדילוג השווה בספר בראשית. מחקר זה עסק בקבוצה של מופעי ביטויים בדילוגים שווים, שהם מינימליים בקטעים גדולים בטקסט הנחקר. מופעים כאלה נכנה כאן "צפני *ELS*". במוקד המחקר היה ניסוי שבו נמדדה הקורלציה בין צפני *ELS* של השמות והכינויים של קבוצת אישים (גדולי תורה) לבין צפני *ELS* של תאריכי הלידה והפטירה שלהם. המדידה הראתה "סיגנל" ברור. המובהקות הסטטיסטית של הסיגנל נמדדה באמצעות מבחן פרמוטציות, דהיינו, רנדומיזציה של הנתונים. התוצאה הייתה מובהקת ביותר, ברמה של **0.00002**. תיאור מפורט בעברית של העבודה פורסם בבד"ד.[2]

מסקנת הניסוי שבספר בראשית מוצפנים פרטים ביוגרפיים של גדולי התורה הציתה פולמוס לוהט. במאמר ביקורת,[3] שהתפרסם חמש שנים אחרי כן ב־*Statistical Science*, טענו מבקרי המחקר (להלן מבב"ק – על פי ראשי התיבות של שמות המחברים) כי התוצאה הנ"ל הושגה במעשה תרמית, באמצעות "תפירה" של נתונים. "תפירה" של נתונים מאפשרת לקבל סיגנל גבוה באופן מלאכותי בטקסט, אף על פי שהופעת צפני *ELS* בטקסט

1 D. Witztum, E. Rips, and Y. Rosenberg, "Equidistant Letter Sequences in the Book of Genesis", *Statist. Sci.* 9 (3) (1994), pp. 429-438.
נמצא גם באתר "צופן בראשית" בקישור:
http://www.torahcode.co.il/pdf_files/pub/wrr.pdf

2 ד' ויצטום, "על הרמז בדילוג השווה: מדידה חדשה של מדגם גדולי חכמי התורה", בד"ד 7 (קיץ תשנ"ח). נמצא גם באתר הנ"ל בקישור:
http://www.torahcode.co.il/pdf_files/pub/bdd.pdf

3. B. D. McKay, D. Bar-Natan, M. Bar-Hillel, and G. Kalai, "Solving the Bible Code Puzzle", *Statist. Sci.* 14 (2) (1999), pp. 150-173.

היא אקראית לחלוטין. הדבר נעשה באמצעות התאמה של הנתונים למערכת הצפנים (למשל, על ידי בדיקת כמות גדולה של נתונים, בחירת "המצליחים" והסתרת "הנכשלים"). כדי להוכיח כי אפשר היה לרמות, הם "תפרו" באופן מוצהר רשימת שמות וכינויים עבור אותה קבוצת אישים שבניסוי המקורי, וכך יצרו באופן מלאכותי סיגנל בעל מובהקות חזקה, שאינה פחותה מזו של החוקרים – אך בספר "מלחמה ושלום".[4] בקצרה, הם טוענים: עשינו "אותו דבר" בספר "מלחמה ושלום".

הדיון בשאלה אם מבקרי המחקר אכן הצליחו לעשות "אותו דבר" בספר "מלחמה ושלום" או לא נמצא בתחום הלשוני והביבליוגרפי-תורני, ויוזכר כאן בנספח בלבד. במאמר הנוכחי ננסה לבדוק באמצעים מתמטיים-סטטיסטיים אם ניתן להבחין בין **הצלחה אמיתית** של מערכת צפני *ELS* לבין **הצלחה מלאכותית** הנוצרת כתוצאה מ"תפירה" של נתונים.

סיגנל הנוצר באופן מלאכותי באמצעות "תפירה" של מערכת צפני *ELS* בקטע מוגדר מטקסט, צפוי לדעוך כאשר המדידות לזיהויו נעשות בקטע ארוך יותר, המכיל את הקטע המקורי. הדבר נבדק עבור רשימת הנתונים שנמדדה בעבודת המבקרים[5] בספר "מלחמה ושלום". תחום המדידה המקורי היה קטע בן 78,064 אותיות מתחילת הספר הנ"ל. תחום המדידה החדש הינו קטע מתחילת "מלחמה ושלום", הכפול באורכו מתחום המדידה המקורי. מתברר כי בתחום החדש הסיגנל אכן דועך באופן בולט.

לעומת זאת, בדיקה דומה בספר התורה עבור רשימת הנתונים שנמדדה בעבודה המקורית של החוקרים[6] בספר בראשית, העלתה תוצאה הפוכה. תחום המדידה המקורי היה קטע בן 78,064 אותיות מתחילת ספר התורה (כלומר, כל ספר בראשית). תחום המדידה החדש הינו קטע מתחילת ספר התורה, הכפול באורכו מתחום המדידה המקורי. מתברר כי בתחום המדידה החדש הסיגנל **מתחזק באופן בולט ומובהק** במקום לדעוך.

מבוא

א. רשימת הנתונים בניסוי שלפנינו

1. רשימת הנתונים שנמדדה בעבודתם של דורון ויצטום, אליהו ריפס ויואב רוזנברג (להלן ור"ר),[7] בנויה מזוגות ביטויים *(w, w′)*. בכל זוג, *w* הוא שם אישיות (או כינויה) מקבוצת אישים (גדולי תורה) ו־*w′* – תאריך הלידה או הפטירה שלה.

4 ל"נ טולסטוי, מלחמה ושלום, לאה גולדברג (מתרגמת), מרחביה: ספריית פועלים, 1953.
5 לעיל הערה 3.
6 לעיל הערה 1.
7 לעיל הערה 1.

2. השמות והכינויים של כל אישיות הם משני סוגים: שמות וכינויים המיוחדים לאותה אישיות, והכינוי הסטנדרטי "רבי פלוני" (כאשר "פלוני" הוא שמו העברי הפרטי), המשותף לכמה אישים.

לדוגמה: לאישיות *#1* ברשימה יש כינוי "אישי" הראב"י, וכינוי סטנדרטי **רבי אברהם**. הכינוי רבי אברהם ניתן לכל חכם ששמו הפרטי אברהם, ולכן הוא משותף לארבעת האישים הראשונים בקבוצת האישים.[8]

3. לכן, ניתן להציג את רשימת הנתונים השלמה, ***LIST 2***, כאיחוד של שתי תת־רשימות:

- תת־הרשימה *L2* הבנויה מאותם זוגות ביטויים *(w, w')*, שבהם *w* הוא מן השמות והכינויים המיוחדים.
- תת־הרשימה *L'2* הבנויה מאותם זוגות ביטויים *(w, w')*, שבהם *w* הוא הכינוי רבי פלוני.

$$LIST\,2 = \text{הרשימה הכוללת} = L2 + L'2$$

אבחנה זו נעשתה לפני ביצוע מבחן הפרמוטציות על ידי ור"ר (לעיל הערה 1, עמ' 436 בסופו – 437). מבחן הפרמוטציות, שפרטיו סוכמו מראש בין ישראל אומן לפרסי דיאקוניס, נועד למדוד את מובהקותם של ארבעה סטטיסטים: שני סטטיסטים לגבי הרשימה כולה *LIST 2* ואותם שני סטטיסטים לגבי *L2*. (על סטטיסטים אלה ייכתב בהמשך, בסעיף ב3.)

4. המובהקות החזקה ביותר, 4 למיליון, התקבלה עבור אחד משני הסטטיסטים עבור תת־הרשימה *L2*, וכך נקבעה הצלחת הניסוי.

5. טענתם המרכזית של מבקרי המחקר ברנדן מק־קי, דרור בר־נתן, מיה בר־הלל וגיל קלעי (להלן מבב"ק) במאמרם [לעיל הערה 3] היא כי תת־הרשימה *L2* ניתנת "לתפירה" על ידי בחירה או השמטה מכוונת של שמות וכינויים (בניגוד לכינויים הסטנדרטיים רבי פלוני שהם קבועים). לכן, כאשר הכינו באופן מוצהר רשימה "תפורה" של נתונים כדי שתצליח בספר "מלחמה ושלום", רק תת־הרשימה (ללא הכינויים הסטנדרטיים), אותה נכנה כאן *BM2*, תרמה להצלחה.

6. לגופה של טענת מבב"ק כבר ענינו באופן פרטני במקום אחר (ראו בנספח, סעיף א). כאן אנו מציעים ניסוי אשר בודק, אף מבלי להיכנס לגוף הטענה, אם *L2* מתנהגת כרשימת נתונים "תפורה".

8 כינויים סטנדרטיים אלה נמצאים בעמודה מיוחדת בטבלת השמות והכינויים שהכין המומחה החיצוני, פרופ' שלמה זלמן הבלין (ראו במאמר בבד"ד, לעיל הערה 2, עמ' 73–74).

ב. הגדרות

1. יהא T טקסט, נגדיר קטע בן n אותיות D, בטקסט T כך:

$$D \equiv ([t_{d_1}, t_{d_n}], T)$$

כאשר t_{d_1} הוא המספר הסידורי ב־T של d_1, האות הראשונה בקטע D, ו־t_{d_n} הוא המספר הסידורי ב־T של d_n, האות האחרונה בקטע D.

2. יהא T טקסט, D קטע ממנו, $LIST$ רשימה של זוגות ביטויים, ו־$ELS(LIST, D)$ קבוצת צפני ELS המייצגים את הביטויים ב־$LIST$ והמוגדרים ב־D. בניסוי המקורי של ור"ר (לעיל הערה 1, עמ' 434-435; לעיל הערה 2, עמ' 60-61), הוגדרה "מידת הקִרבה המכוילת" $c(w, w')$ עבור כל זוג ביטויים (w, w'). "מידת הקרבה המכוילת" בין הביטויים, $c(w, w')$, היא מספר בין 0 ל־1: ערכו קרוב ל־0 כאשר צפני ELS של הביטויים (w, w') נפגשים בצורה מכונסת במיוחד; ערכו קרוב ל־1 כאשר צפני ELS של הביטויים (w, w') רחוקים ומפוזרים במיוחד.

3. כדי לסכם את "הנטייה הכוללת לקִרבה" עבור כלל הזוגות ב־$LIST$, הוגדרו בניסוי המקורי של ור"ר (לעיל הערה 1, עמ' 436) שתי "מידות לנטייה הכוללת לקִרבה" (המשמשות כסטטיסטים) באמצעות שני אופרטורים $\boldsymbol{P_1}$ ו־$\boldsymbol{P_2}$, המיושמים לגבי הטקסט, הקטע והרשימה הנתונים. המידות המתקבלות כאן יהיו:

$$P_i \equiv \boldsymbol{P_i}(LIST, D), \quad i = 1, 2.$$

הגדרה מתמטית מלאה של שני הסטטיסטים נמצאת בנספח, סעיף ב. כאן רק נעיר כי כל אחת משתי המידות P_i משמשת מעין "גלאי" להצפנה המשוערת.

- P_1: לפי מידה זו מונים את מספר התוצאות $c(w, w')$ ב"אזור ההצלחה", אשר הוגדר אפריורי כמרווח בין 0 ל־0.2, ומחשבים מה הסיכוי לקבל באקראי את הערך המתקבל.
- P_2: מידה זו נבנתה כך שהיא רגישה לגודלם של **כל** המספרים $c(w, w')$.

לכל סטטיסטי, P_1 או P_2, יש יתרונות וחסרונות.

(א) למשל, ל־P_1 יש יתרון, שאין השפעה של "גודל הכישלון" של הזוגות "הנכשלים". אנחנו מחפשים הצפנה – זוגות "מוצפנים". לא מעניין אותנו כיצד בדיוק מתנהגים אותם זוגות שאינם מוצפנים. עניין זה עצמו הוא חיסרון של P_2, הרגישה לתוצאות באזור הכישלון (סמוך ל־1).

(ב) לעומת זאת ל־P_2 יש יתרון, שהיא רגישה לחדות ההצלחה. עניין זה עצמו חסר ל־P_1, שעבורה כל תוצאה ב"אזור ההצלחה" – ערכה שווה.

יש להדגיש, כי שני הסטטיסטים הוגדרו אפריורי, עוד לפני הניסוי הראשון של רשימת גדולי התורה, וכי מאז שימשו **בקביעות** לסיכום הנטייה הכוללת לקִרבה בכל המדגמים הרבים שמדדנו ופרסמנו במרוצת השנים.[9] מתברר שההחלטה להשתמש בשני סוגי "גלאים" הוכיחה את עצמה.[10] פעמים רבות P_2 הצביע על מובהקות חזקה (כמו במקרה של *L2*), ופעמים רבות P_1 הוא שהצביע על מובהקות חזקה יותר (למשל, עבור "מדגם העמים"[11] או עבור "מדגם אישי ספר בראשית"[12]). בעבודה הנוכחית, בה אנו עוסקים בתת־הרשימה *L2*, ברירת המחדל היא לדבוק באותם שני סטטיסטים.

4. נגדיר **סיגנל** S_i באופן הבא:

$$S_i(LIST, D) \equiv 1 / P_i(LIST, D), \quad i = 1, 2.$$

בנספח (סעיף ב) מבואר כי כל אחד מהסטטיסטים P_i מקבל ערך **הקטֵן** ככל שהמפגשים בין צפני ה־*ELS* **יותר "מוצלחים"**. לכן, הסיגנל יקבל ערך **הגדל** ככל שהמפגשים בין צפני ה־*ELS* **יותר "מוצלחים"**. (הגדרת הסיגנלים S_i אינה מחדשת דבר מהותי לעומת ההגדרה המקורית של P_i, אבל עשויה לסייע לתפיסה אינטואיטיבית של הנדון כאן, לקוראים מתחומי מדע מגוונים.)

את מובהקות הסיגנל, כלומר הסיכוי "לקבל סיגנל כה גבוה במקרה", ניתן לקבוע באמצעות מבחן רנדומיזציה (ראו בנספח, סעיף ג).

5. יהא D' קטע ב־T, ארוך מ־D והמכיל אותו. הסיגנלים המתקבלים עבור *LIST* ב־D' יסומנו $S_i(LIST, D')$. אם הסיגנל של *LIST* ב־D התקבל באופן מלאכותי, על ידי מניפולציה של נתוני *LIST* – אזי במדידה חדשה בקטע D', אנו מצפים לדעיכת הסיגנל:

9 למשל: D. Witztum, E. Rips, and Y. Rosenberg, "Equidistant Letter Sequences in the Book of Genesis: II. The Relation to the Text" (preprint, 1995); ד' ויצטום, א' ריפס וי' רוזנברג, "צפן חבוי בדילוג שווה בספר בראשית: מובהקות סטטיסטית של התופעה", פרה-פרינט, אביב ה'תשנ"ו; ד' ויצטום, צופן בראשית, ירושלים ה'תשס"ד. ראו בעמ' 217 (טבלה 1-20) ריכוז עבודות, מלבד אלה הפזורות בפרקי הספר השונים.

10 אף על פי ששילמנו "מחיר" סטטיסטי על כך שהיו לרשותנו שתי אפשרויות.

11 ראו במאמר הראשון בהערה 9 לעיל, וגם: ד' ויצטום, "מדגם העמים", ה'תשס"ה. נמצא באתר הנ"ל בקישור: http://www.torahcode.co.il/amim_heb.htm.

12 ד' ויצטום, "מפגשים בין שמות אישי ספר בראשית לתאריכי הלידה שלהם", ה'תשנ"ט. נמצא באתר הנ"ל בקישור: http://www.torahcode.co.il/leida_heb.htm.
המאמר, שעבר שיפוט קפדני, פורסם בה'תשס"ו, במסגרת הכנס הבינלאומי ה־18 על זיהוי תבניות (Pattern Recognition). נמצא באתר הנ"ל בקישור:
http://www.torahcode.co.il/pdf_files/pub/leida1.pdf

$$S_i(LIST, D') < S_i(LIST, D)$$

הסבר: בהגדלת הקטע מ־D ל־D' מקבלים במקום $ELS(LIST, D)$ עליה בוצעה המניפולציה, קבוצה גדולה יותר $ELS(LIST, D')$, שבה צפני ELS חדשים המופיעים באקראי ושהנתונים לא הותאמו אליהם. הם יוצרים "רעש" אקראי המקטין את הסיגנל.

6. נגדיר את **מדד ההגברה** Q_i:

$$Q_i \equiv S_i(LIST, D') / S_i(LIST, D)$$

$Q_i < 1$ מצביע על דעיכה, היחלשות S_i, ואילו $Q_i > 1$ מצביע על הגברת S_i.

ייתכן כי עבור i שונים תתקבל מידת הגברה שונה. אף ייתכן ש־Q_1 יצביע על הגברה, בעוד Q_2 מצביע על דעיכה, או להיפך. אנו מעוניינים **במדד ההגברה המרבית** Q:

$$Q \equiv \max\{Q_1, Q_2\}$$

ג. דוגמה:

רשימת הנתונים שנמדדה בעבודת מבב"ק בספר "מלחמה ושלום" הוכנה באופן מוצהר על ידי מניפולציה של נתוני תת-רשימה, אותה כינינו לעיל (בסעיף א5) $BM2$. תחום המדידה המקורי, D, היה הקטע $[1, 78{,}064]$, מתחילת הספר הנ"ל. תחום המדידה החדש, D', הינו הקטע $[1, 156{,}128]$ מתחילת "מלחמה ושלום".

לפי מדידה:

$$S_1(BM2, D) = 4.9293\text{E+4}, \quad S_2(BM2, D) = 3.4723\text{E+4}.$$

לעומת זאת, בקטע המוגדל מקבלים:

$$S_1(BM2, D') = 7.1855, \quad S_2(BM2, D') = 3.4247\text{E+1}.$$

מתברר כי בתחום המוגדל הסיגנל אכן דועך באופן בולט:

$$Q_1 = 1.4577\text{E}{-4}, \quad Q_2 = 9.8629\text{E}{-4},$$

ולכן:

$$Q = \max\{Q_1, Q_2\} = 9.8629\text{E}{-4}.$$

ד. מה יקרה אם קיימת הצפנה אמיתית של $LIST$ ב־T?

ישנן שתי אפשרויות:

1. אם תחום ההצפנה של *LIST* ב־*T* מוכל בקטע *D*, אזי הרחבת הקטע מוסיפה רק "רעש" אקראי, ולכן הסיגנל ייחלש וידעך.
2. אם תחום ההצפנה של *LIST* ב־*T* גדול מן הקטע *D*, אזי הרחבת הקטע לא תחליש את הסיגנל, וייתכן שאף תגביר אותו.

ה. לפי ניסיון שהצטבר[13] מאז פורסמה עבודת ור"ר, נראה הדבר כי התופעה שלה טענו ור"ר אינה מוגבלת לספר בראשית, אלא מצויה בתורה כולה. אנו רואים את כל התורה כחטיבה אחת לעניין התופעה שלפנינו,[14] כאשר נושאים שונים מוצפנים באזורים שונים בספר. בראייה זו, ספר בראשית שבו נמדדה רשימה *L2*, אינו אלא הקטע [1, 78,064] מספר התורה. נבדוק את *L2* בקטע גדול מספר התורה, המכיל את ספר בראשית. לפי טענת מבב"ק כי *L2* אינה אלא רשימת נתונים "תפורה", אנו מצפים אך ורק לדעיכת הסיגנל המקורי.

חלק א: מערכת המדידה ותוצאות המדידות

בחלק זה שלוש מדידות עיקריות.
במדידה 1, נמדד אם הסיגנל שנמדד בספר בראשית עבור *L2* מתגבר או דועך כאשר אנו עוברים למדוד בקטע גדול מן התורה המכיל את ספר בראשית. נבנתה מערכת מדידה כדי לקבוע את מובהקות השינוי, ומוצגות תוצאות הרצת מערכת המדידה.
במדידה 2, נערכו מדידות ביקורת.
במדידה 3, יש חזרה על המדידות שנעשו במדידה 1 עבור *L2* בשינוי אחד: כמה מן התאריכים ברשימת הנתונים שונו בהתאם לטענות מבב"ק (כל שאר התאריכים והפרטים, בפרט השמות והכינויים – נשארים כפי שהם, ללא שינוי).

מדידה 1: בדיקת *L2* בקטע גדול מספר התורה

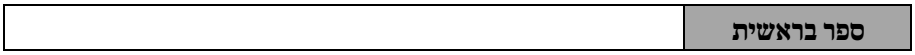
הקטע המקורי, *D*, היה ספר בראשית (*G*). אנו צריכים להרחיב את *D* בקטע "גדול דיו" כדי לצפות לדעיכת הסיגנל. ככלל אצבע בחרנו ב־*D′* הכפול באורכו מ־*D*.
נתוני הקטעים:

13 הנה כמה מן העבודות שנעשו בספר התורה (גם מחוץ לספר בראשית): באתר הנ"ל, במדור "פרסומים מדעיים", http://www.torahcode.co.il/pub_index_heb.htm, מאמרים 5, 8, 9. וכן במאמרים http://www.torahcode.co.il/kriah1_heb.htm, http://www. torahcode.co.il/weis2_heb.htm

14 זאת ניתן לראות בצפנים המחברים בין החומשים (למשל, בשני הקישורים האחרונים).

- הקטע המקורי D, שהוא ספר בראשית (G), הוא "הקטע $[1, 78{,}064]$ מן התורה" (1 הוא המספר הסידורי של האות הראשונה בתורה, ו-78,064 הוא המספר הסידורי של האות האחרונה בספר בראשית). ראו ציור 1.
- הקטע D' הוא הקטע $[1, 156{,}128]$ מן התורה: $D' = (TORAH, [1, 156{,}128])$. ראו ציור 2.

התורה כולה



D

ציור 1. ספר בראשית כקטע מן התורה

התורה כולה

$D' = D + D''$

D D''

ציור 2. הקטע המורחב D' כקטע מן התורה

א. ערכנו שתי מדידות:

- בספר בראשית (G).
- בקטע D'.

1. התוצאות בספר בראשית (G):

$$S_1(L2, G) = 1.1923\text{E+6}, \qquad S_2(L2, G) = 1.0776\text{E+8}.$$

2. התוצאות בקטע D':

$$S_1(L2, D') = 2.7504\text{E+9}, \qquad S_2(L2, D') = 2.2676\text{E+7}.$$

מכאן:

$$Q_1 = 2.307\text{E+3}, \qquad Q_2 = 2.104\text{E}{-1}, \qquad Q = \max\{Q_1, Q_2\} = 2.307\text{E+3}.$$

ב. צריך להעריך מהי המובהקות של התוצאה הנ"ל, דהיינו: "מה הסיכוי שיתקבל במקרה Q כה גדול?".

1. לשם כך בנינו אוסף גדול של N טקסטים, T_j, "דומים" לקטע D' מספר התורה, המוגדר בסעיף א. הקטע D' בנוי משני חלקים (ציור 3):

- מחלקו הראשון, D = ([1, 78,064], *TORAH*), הוא ספר בראשית בדיוק,
- ומתוספת – הקטע D'' מן התורה: D'' = ([78,065, 156,128], *TORAH*).

$$D' = D + D''$$

ספר בראשית	תוספת
D	D''

ציור 3. הקטע D'

כל טקסט T_j אף הוא בנוי משני חלקים (ציור 4):

- מחלקו הראשון, D = ([1, 78,064], *TORAH*), הוא ספר בראשית בדיוק,
- ומתוספת – הקטע D''_j – שהוא ערבוב אקראי של הקטע D'' מן התורה (ערבוב אקראי של מילים בתוך פסוקים. הדבר נעשה בדיוק כפי שנעשה בניסוי המקורי של וו"ר לגבי טקסט U שם).

$$T_j = D + D''_j$$

ספר בראשית	תוספת מעורבבת
D	D''_j

ציור 4. טקסט T_j

2. כאשר מודדים את *L2* בטקסט T_j, התוצאה הסופית "נהנית" מכל היתרונות (והחסרונות) שהיו במדידה המקורית של *L2* בספר בראשית. במילים אחרות: כל רץ T_j נהנה בדיוק מאותה "רוח גבית".

בודקים עבור N טקסטים T_j בכמה טקסטים n מתקבל ערך Q גדול או שווה ל-**2,307**.

הסיכוי לקבל כזו תוצאה במקרה הוא $p = n/N$.

ג. מובהקות התוצאה:
מתוך **50,000** טקסטים T_j היו רק **48** טקסטים שבהם נתקבל Q כה גדול. לפיכך:

$$p < 0.001.$$

מדידה 2: בדיקות ביקורת

א. היפוך סדר הקטעים

יצרנו טקסט נוסף, D^*, על ידי הפיכת הסדר של הקטעים D ו־D''. דהיינו (ציור 5):

- החלק הראשון (D^* ,[1, 78,064]) הוא D'', דהיינו הקטע [78,065, 156,128] מן התורה.
- החלק השני (D^* ,[78,065, 156,128]), הוא ספר בראשית, דהיינו הקטע [1, 78,064] מן התורה.

$$D^* = D'' + D$$

תוספת	ספר בראשית
D''	D

ציור 5. טקסט D^*

התוצאות בקטע D^*:

$$S_1(L2, D^*) = 1.0655\text{E}{+3}, \quad S_2(L2, D^*) = 4.00\text{E}{+4}.$$

מכאן:

$$Q_1 = 8.937\text{E}{-4}, \quad Q_2 = 3.7106\text{E}{-4}, \quad Q = \max\{Q_1, Q_2\} = 8.937\text{E}{-4}.$$

ב. מדידה נוספת

בניסוי המקורי של ור"ד נמדדה המובהקות של "המידות הכוללות לקִרבה" P_i באמצעות מבחן פרמוטציות. בכל פרמוטציה מוצמדים באקראי התאריכים של אישיות i מרשימת הנתונים לאישיות j מרשימה זו, וכך לכל $i, j = 1,2,3,\ldots,32$. לכן כל הביטויים הנמצאים ברשימת הפרמוטציה הם שמות, כינויים ותאריכים הנמצאים ברשימת הנתונים המקורית.

התוצאה הסופית נקבעה למעשה על ידי הדירוג של P_4 במבחן הפרמוטציות, שהיה 4 מתוך מיליון. כלומר, היו שלוש פרמוטציות אשר "ניצחו" את הרשימה המקורית: הסיגנל S_4 עבורן היה חזק יותר מן הסיגנל המקורי. מעניין לבדוק מה יקרה לסיגנלים אלה כאשר נרחיב את תחום המדידה מספר בראשית לקטע D' (הנ"ל). להלן נשתמש בסימון של המאמר הנוכחי בו האינדקסים 3 ו-4 הוחלפו ל-1 ו-2 בהתאמה (ראו בנספח, סעיף ב1).

1. הפרמוטציה שהייתה ראשונה מתוך מיליון פרמוטציות (PER1).
מתברר שהסיגנל נחלש מאוד:

$$Q = \max\{Q_1, Q_2\} = 9.3927\text{E}{-7}.$$

פרטי PER1 והחישובים נמצאים בנספח, סעיף ה. כאן רק נציין כי תוצאה זו מפתיעה. לא זו בלבד שהביטויים ב־PER1 הם השמות, הכינויים והתאריכים הנמצאים ברשימת הביטויים המקורית, אלא אף שני אישים "שהצליחו" בעבודה המקורית במפגשים עם תאריכיהם (מס' 1 ומס' 31), נמצאים עם תאריכיהם, ללא כל שינוי, גם ב־PER1.

2. הפרמוטציה שהייתה שנייה מתוך מיליון פרמוטציות (PER2).
גם כאן מתברר שהסיגנל נחלש מאוד:

$$Q = \max\{Q_1, Q_2\} = 1.7945\text{E}{-5}.$$

פרטי PER2 והחישובים נמצאים בנספח, סעיף ה. מה שנכתב לעיל לגבי PER1 רלוונטי גם כאן. במקרה זה מדובר על ארבעה אישים "שהצליחו" בעבודה המקורית במפגשים עם תאריכיהם (מס' 1, מס' 14, מס' 22 ומס' 23), הנמצאים עם תאריכיהם, ללא כל שינוי, גם ב־PER2.

3. הפרמוטציה שהייתה שלישית מתוך מיליון פרמוטציות (PER3).
מתברר שהסיגנל נחלש:

$$Q = \max\{Q_1, Q_2\} = 1.95\text{E}{-3}.$$

פרטי PER3 והחישובים נמצאים בנספח, סעיף ה. מה שנכתב לעיל לגבי PER1 ו-PER2 רלוונטי גם כאן. במקרה זה מדובר על שלושה אישים "שהצליחו" בעבודה המקורית במפגשים עם תאריכיהם (מס' 14, מס' 23 ומס' 30), הנמצאים עם תאריכיהם, ללא כל שינוי, גם ב־PER3.

ג. מדידה בספר מלחמה ושלום

נחזור לדוגמה שבסעיף ג של המבוא. כאמור שם, רשימת הנתונים שנמדדה בעבודת מבב"ק בספר "מלחמה ושלום", הוכנה באופן מוצהר על ידי מניפולציה של נתוני תת־רשימה, אותה נכנה *BM2*. תחום המדידה המקורי, D, היה הקטע [1, 78,064] מתחילת הספר הנ"ל. תחום המדידה החדש, D', הינו הקטע [1, 156,128] מתחילת "מלחמה ושלום".
לפי מדידה מתברר, כי בתחום המוגדל הסיגנל אכן דועך באופן בולט:

$$Q_1 = 1.4577\text{E}{-4}, \quad Q_2 = 9.8629\text{E}{-4},$$

ולכן:

$$Q = \max\{Q_1, Q_2\} = 9.8629\text{E}{-4}.$$

לפי בקשת אחד הקוראים של טיוטת המאמר, מדדנו מה הסיכוי לקבל מדד הגברה חלש כמו Q_2 או חזק ממנו (כלומר, דעיכה כזו של הסיגנל S_2 או קטנה ממנה).

השתמשנו באוסף של **1000** טקסטים. מחציתו הראשונה של כל טקסט מן האוסף היא הקטע [1, 78,064] מתחילת הספר הנ"ל, ומחציתו השנייה היא הקטע [78,065, 156,128] ממנו, המעורבב באופן אקראי. מתברר כי ב־**396** טקסטים מתוך **1000** נתקבל $Q'_2 \geq Q_2$.

מדידה 3: בדיקת *L2M* בקטע גדול מספר התורה

במאמרם העלו מבב"ק טענות אחדות בנוגע לתאריכים ולצורות התאריך שננקטו ברשימות הנתונים של ור"ר. מטענות אלה רלוונטית רק טענה אחת ויחידה לגבי תת־הרשימה *L2*: כי עבור ארבעה אישים יש להוסיף או לשנות תאריך (הפרטים בנספח, סעיף ו). נגדיר את הרשימה *L2M* כתת־הרשימה *L2* עם השינויים הנ"ל בתאריכים (כל שאר התאריכים והפרטים, בפרט השמות והכינויים – נשארים כפי שהם, ללא שינוי).

נחזור על **מדידה 1** כאשר אנו מודדים את *L2M* במקום *L2*.

א. ערכנו שתי מדידות:

- בספר בראשית (G).
- בקטע D'.

1. התוצאות בספר בראשית (G):

$$S_1(L2M, G) = 1.1837\text{E}{+7}, \qquad S_2(L2M, G) = 1.2739\text{E}{+9}.$$

(שימו לב, כי הסיגנל עבור *L2M* בספר בראשית חזק פי **10** מאשר הסיגנל עבור *L2*!)

2. התוצאות בקטע D':

$$S_1(L2M, D') = 2.5023E+12, \qquad S_2(L2M, D') = 2.9412E+9.$$

מכאן:

$$Q_1 = 2.114E+5, \qquad Q_2 = 2.309, \qquad Q = \max\{Q_1, Q_2\} = 2.114E+5.$$

ב. מובהקות התוצאה:

מדידה זו נעשתה בדיוק כמו במדידה 1 סעיף ב. מתוך **50,000** טקסטים T_j היו רק **13** טקסטים שבהם נתקבל Q כה גדול. לפיכך:

$$p = 0.00026.$$

חלק א: מסקנות

א. תוצאת הניסוי בתורה מורה, כי הסיגנל מתחזק בקטע המורחב D'. עובדה זו מפריכה את טענת מבב"ק כי הסיגנל שנתקבל בקטע D המקורי היה מלאכותי ונוצר ממניפולציה בנתוני *L2*.

ב. תוצאה א של מדידה 2 מראה כי התגברות הסיגנל אינה תולדה של סכום ההצפנות בקטעים D בפני עצמו ו־D'' בפני עצמו. סדר החיבור הנכון בין הקטעים הוא קריטי להתגברות הסיגנל.

ג. נראה שתחום ההצפנה של *L2* קרוב לתחום המורחב, D', יותר מאשר לתחום המקורי, D (ספר בראשית). לכן, יש עניין רב לערוך ניסויים נוספים בנושא *L2* בתחום המורחב. [יש להדגיש, כי אין אנו טוענים כי בתחום D' מוצפנת *L2* באופן אופטימאלי. דבר זה כלל לא נבדק.]

חלק ב: ניסויים נוספים ב"תחום המורחב"

בחלק א הסקנו, כי תחום ההצפנה של רשימת הנתונים *L2* קרוב יותר לתחום המורחב, שהוא הקטע [1, 156,128] מן התורה, מאשר לתחום המקורי, הקטע [1, 78,064] מספר התורה. לכן, יש עניין רב לערוך ניסויים נוספים בנושא רשימת נתונים זו בתחום המורחב.

בחרנו לבדוק אם קיימת התגברות הסיגנל גם עבור שני וריאנטים של *L2*, וזאת בעקבות רעיון של מבב"ק. נמצא כי גם כאן הסיגנל מתגבר ובמובהקות חזקה.

א. הטענה של חוסר סימטריה

1. נזכיר כי רשימות הנתונים ששימשו לניסויי ור"ר הן רשימות של זוגות ביטויים. בכל זוג, אחד מבני הזוג הוא שם או כינוי של אישיות מסוימת, ובן זוגו הוא תאריך הלידה או הפטירה של אישיות זו.
כל תאריך הוצג בשלוש תצורות. למשל היום הראשון בחודש תשרי הוצג כך:
A – "א' תשרי",
B – "בא' תשרי",
C – "א' בתשרי".
הדבר נקבע באופן אפריורי לפני הכנת רשימת הנתונים הראשונה (ראו על כך בנספח, סעיף ז).

2. לטענת מבב"ק חסרה כאן תצורה רביעית: *D* – "בא' בתשרי".
הרציונל העומד מאחורי טענת "חסר" זו מצריך סימטריה: D מתייחס ל־C כמו B ל־A.
נסביר: B נוצר מ־A באמצעות תוספת "ב" השימוש – ב"א' תשרי".
לכן יוצרים את D מ־C באמצעות תוספת "ב" השימוש – ב"א' בתשרי".
טענה זו הועלתה ממש מתחילת הוויכוח. העובדה שור"ר לא השתמשו בתצורה D הוצגה בהבלטה רבה בכל פרסומי מבב"ק כראיה למניפולציה (ראו על כך בנספח, סעיף ז. מבואר שם כי "ראיה זו למניפולציה" הופרכה כבר על ידינו, וכי אדרבה, היה "כדאי" לור"ר להשתמש בתצורה D).

3. אם שלישיית התצורות ABC נחשבת ללא־סימטרית בגלל שתצורה C חסרה "בן זוג", אפשר לתקן זאת בשתי דרכים:
(א) הוספת תצורה D – כפי שהציעו מבב"ק – וכך מקבלים את רביעיית התצורות ABCD.
(ב) הורדת תצורה C – וכך מקבלים את צמד התצורות AB.
נסמן ב־*L2(ABCD)* את רשימה *L2* כאשר התאריכים בה ניתנים ברביעיית התצורות ABCD.
נסמן ב־*L2(AB)* את רשימה *L2* כאשר התאריכים בה ניתנים בצמד התצורות AB.
אנו נעמיד את שתי האפשרויות במבחן. להלן נמדוד את הסיגנלים של *L2(ABCD)* ושל *L2(AB)* בספר בראשית ובקטע המורחב.

ב. המדידות

בחלק זה ארבע מדידות.
מדידה 1: נמדוד אם הסיגנל שנמדד בספר בראשית עבור *L2(ABCD)* מתגבר או דועך כאשר אנו עוברים למדוד בתחום המורחב, ומה מובהקות התופעה.
מדידה 2: מדידה דומה עבור *L2(AB)*.
מדידה 3: מדידה דומה עבור *L2M(ABCD)* (ראו הגדרת *L2M* במדידה 3 של חלק א).
מדידה 4: מדידה דומה עבור *L2M(AB)*.

מדידה 1: בדיקת *L2(ABCD)* בתחום המורחב

א. נתוני הקטעים הם כמו במדידה 1 של חלק א. ערכנו שתי מדידות:

- בספר בראשית (*G*).
- בקטע *D′*.

1. התוצאות בספר בראשית:

$$S_1(L2(ABCD), G) = 4.5979\text{E+6}, \quad S_2(L2(ABCD), G) = 1.7889\text{E+8}.$$

2. התוצאות בקטע *D′*:

$$S_1(L2(ABCD), D') = 3.9210\text{E+11}, \quad S_2(L2(ABCD), D') = 2.0161\text{E+8}.$$

לכן:

$$Q_1 = 8.5277\text{E+4}, \quad Q_2 = 1.1270, \quad Q = \max\{Q_1, Q_2\} = 8.5277\text{E+4}.$$

ב. מובהקות התוצאה:
הערכת מובהקות התוצאה נעשית כמו בחלק א, באמצעות N טקסטים, T_j, מעורבבים כנ"ל. מתוך 50,000 טקסטים T_j היו רק 4 טקסטים בהם נתקבל Q כה גדול. לפיכך המובהקות היא:

$$p = 0.00008.$$

מדידה 2: בדיקת *L2(AB)* בתחום המורחב.

א. נתוני הטקסטים כנ"ל. ערכנו שתי מדידות:

- **בספר בראשית (*G*).**
- **בקטע *D′*.**

1. התוצאות בספר בראשית:

$$S_1(L2(AB), G) = 3.0097E+4, \quad S_2(L2(AB), G) = 8.4746E+6.$$

2. התוצאות בקטע *D′*:

$$S_1(L2(AB), D') = 2.9279E+10, \quad S_2(L2(AB), D') = 4.5249E+7.$$

לכן:

$$Q_1 = 9.7283E+5, \quad Q_2 = 5.3394, \quad Q = \max\{Q_1, Q_2\} = 9.7283E+5.$$

ב. מובהקות התוצאה:

בהשוואה לטקסטים כנ"ל, נמצא כי מתוך 50,000 טקסטים T_j היו רק 19 טקסטים בהם נתקבל Q כה גדול. לפיכך המובהקות היא:

$$p = 0.00038.$$

מדידה 3: בדיקת *L2M(ABCD)* בתחום המורחב.

א. נתוני הטקסטים כנ"ל. ערכנו שתי מדידות:

- **בספר בראשית (*G*).**
- **בקטע *D′*.**

1. התוצאות בספר בראשית:

$$S_1(L2(ABCD),G) = 1.1832E+8, \quad S_2(L2(ABCD),G) = 8.2651E+9.$$

2. התוצאות בקטע *D′*:

$$S_1(L2(ABCD), D') = 4.2589E+14, \quad S_2(L2(ABCD), D') = 1.3228E+10.$$

לכן:

$$Q_1 = 3.5994E+6,\ \ Q_2 = 1.6004,\ \ Q = \max\{Q_1, Q_2\} = 3.5994E+6.$$

ב. מובהקות התוצאה:

בהשוואה לטקסטים כנ"ל, נמצא כי מתוך 50,000 טקסטים T_j היו רק 3 טקסטים שבהם נתקבל Q כה גדול. לפיכך:

$$p = 0.00006.$$

מדידה 4: בדיקת *L2M(AB)* בתחום המורחב.

א. נתוני הטקסטים כנ"ל. ערכנו שתי מדידות:

- **בספר בראשית (G).**
- **בקטע D'.**

1. התוצאות בספר בראשית:

$$S_1(L2(AB),G) = 9.0106E+4,\qquad S_2(L2(AB), G) = 4.2194E+7.$$

2. התוצאות בקטע D':

$$S_1(L2(AB), D') = 1.8211E+13,\qquad S_2(L2(AB), D') = 3.4723E+9.$$

לכן:

$$Q_1 = 2.0211E+8,\ \ Q_2 = 8.2294E+1,\ \ Q = \max\{Q_1, Q_2\} = 2.0211E+8.$$

ב. מובהקות התוצאה:

בהשוואה לטקסטים כנ"ל, נמצא כי מתוך 50,000 טקסטים T_j היו רק 5 טקסטים בהם נתקבל Q כה גדול. לפיכך:

$$p = 0.00010.$$

סיכום כל המדידות

בטבלה 1 מרוכזות התוצאות עבור רשימה *L2* (מחלק א) והווריאנטים שלה (מחלק ב):

טבלה 1

מספר הטקסטים המתחרים n שלהם ערך $Q'_j \geq Q$ בתחרות ובה **50,000** מתחרים

L2	*L2(ABCD)*	*L2(AB)*	רשימה
48	**4**	**19**	n

בטבלה **2** מרוכזות התוצאות עבור רשימה *L2M* (מחלק א) והווריאנטים שלה (מחלק ב):

טבלה 2

מספר הטקסטים המתחרים n שלהם ערך $Q'_j \geq Q$ בתחרות ובה **50,000** מתחרים

L2M	*L2M(ABCD)*	*L2M(AB)*	רשימה
13	**3**	**5**	n

חלק ב: מסקנות

א. בכל ארבע המדידות מדד ההגברה המרבית הצביע על התחזקות ברורה וחדה של הסיגנל ובמובהקות חזקה.

ב. העובדה כי רביעיית תצורות התאריך ABCD, ואף צמד התצורות AB, מצליחים יותר מאשר שלישיית התצורות ABC של ור"ר – עשויה להצביע על האופן שבו מקודדים תאריכים באמצעות צפני *ELS* בספר התורה. כמובן, נדרש מחקר רב נוסף לברר נקודה זו.

חלק ג: דיון

טיוטת מאמר זה נשלחה על ידי המערכת לקורא אנונימי מטעמה, כנהוג. התפתח דיון ממושך שבו העלה הקורא טענות ורעיונות שתרמו להבהרת הנושאים הנידונים במאמר. הדברים משוקעים בחלקם בקטעים מן המבוא ובנספח (בעיקר בסעיף ג). כאן מרוכזים כמה מן הדברים שנדונו, נכתבו ובוצעו בעקבות הדיאלוג עם הקורא.

א. תוצאות מבחן הטקסטים עבור Q_1 ו־Q_2

במאמר עסקנו במדד ההגברה המרבית Q: $Q \equiv \max\{Q_1, Q_2\}$.

בטבלאות 3 ו־4 להלן, המקבילות לטבלאות 1 ו־2 דלעיל, מובאות גם תוצאות מבחן הטקסטים עבור Q_1 ועבור Q_2.

1. הנה התוצאות עבור *L2*:

טבלה 3

מספר הטקסטים המתחרים n_i שלהם ערך $Q'_{ij} \geq Q_i$ בתחרות ובה 50,000 מתחרים, ומספר הטקסטים המתחרים n שלהם ערך $Q'_j \geq Q$ בתחרות ובה 50,000 מתחרים

	L2	*L2(ABCD)*	*L2(AB)*	רשימה
Q_1	45	4	19	n_1
Q_2	641	169	266	n_2
$Q \equiv \max\{Q_1, Q_2\}$	48	4	19	n

2. הנה התוצאות עבור *L2M*:

טבלה 4

מספר הטקסטים המתחרים n_i שלהם ערך $Q'_{ij} \geq Q_i$ בתחרות ובה 50,000 מתחרים, ומספר הטקסטים המתחרים n שלהם ערך $Q'_j \geq Q$ בתחרות ובה 50,000 מתחרים.

	L2M	*L2M(ABCD)*	*L2M(AB)*	רשימה
Q_1	11	3	5	n_1
Q_2	222	122	67	n_2
$Q \equiv \max\{Q_1, Q_2\}$	13	3	5	n

3. הערות

(א) יש הבדל ניכר בין התוצאות עבור Q_1 ועבור Q_2. הדבר נובע בראש ובראשונה מכך שהסיגנל S_2 נחלש מעט במעבר לתחום המורחב, בעוד הסיגנל S_1 מתחזק מאוד.

(ב) במבוא (סעיף ב3) מבואר כי כל אחד משני הסיגנלים משמש כגלאי לתופעה שונה (אומנם קיימת תלות מסוימת). הדבר מבואר עוד וביתר עומק בנספח (סעיף ד).

(ג) הסיגנל S_2 מושפע מאוד מהצטברות תוצאות "מידת הקִרבה המכוילת" בזנב ההתפלגות בקרבת הערך 1. הנה השוואת התוצאות עבור *L2* בזנב ההתפלגות (עבור ספר בראשית היו בסך הכול 126 תוצאות ועבור התחום המורחב – 148):

טבלה 5

מרווח	(0.92, 1]	(0.8, 1]	(0.6, 1]
בראשית	2	8	20
ת' מורחב	5	16	35

העלייה התלולה במספר התוצאות בזנב ההתפלגות עבור התחום המורחב השפיעה לרעה על הסיגנל S_2. זה בדיוק החיסרון הגדול של S_2 המוזכר במבוא: ההשפעה של "גודל הכישלון" של הזוגות "הנכשלים". הדבר האפיל על העובדה כי "באזור ההצלחה" – דהיינו, קרוב ל-0, קיימת עלייה משמעותית במספר התוצאות:

טבלה 6

מרווח	(0, 0.2]	(0, 0.08]
בראשית	49	23
ת' מורחב	63	27

(ד) בניסוי הוספנו לספר בראשית קטע גדול באורך ספר בראשית עצמו. יש לשים לב כי אם אין הצפנה בתחום המורחב, נוספה כאן כמות גדולה מאוד של "רעש" אקראי. יש לכך השלכות הרסניות על הסיגנל. ראו דוגמאות במדידות הביקורת שבמאמר. לכן, אם הסיגנל התחזק רק במעט או אפילו נחלש פי חמישה – זה אירוע לחלוטין לא טריוויאלי, שההסתברות שלו קטנה בהרבה מן הנדרש בניסוי סטטיסטי מקובל. זו אכן התוצאה בטבלאות 3-4 עבור Q_2.

ב. הגברת/דעיכת סיגנל המנורמל לפי מבחן פרמוטציות

בנספח (סעיף ג) הסברנו כי לדעתנו מבחן המובהקות הנכון והמתאים יותר למחקר שלפנינו הוא "מבחן הטקסטים". אולם, לבקשת הקורא מטעם המערכת נחשב את מקדמי ההגברה גם ביחס לסיגנלים שהם מנורמלים לפי מבחן הפרמוטציות – כפי שיבואר לקמן. מבחן הפרמוטציות נערך בכל מקרה ומקרה בדיוק כפי שנעשה במאמר של ור"ר (על מבחן זה ראו בקצרה לעיל חלק א, מדידה 2, סעיף ב).

1. נגדיר את r_i להיות הדירוג של הסטטיסטי P_i במבחן הפרמוטציות, מחולק במספר הפרמוטציות. המספר r_i הוא המובהקות הסטטיסטית של P_i במבחן הפרמוטציות.

2. במקביל להגדרות שבמבוא (סעיף ב):

(א) נגדיר **סיגנל מנורמל במבחן פרמוטציות** כך:

$$Sr_i \equiv 1 / r_i, \quad i = 1, 2.$$

(ב) נגדיר את **מדד ההגברה** Qr_i:

$$Qr_i \equiv Sr_i(LIST, D') / Sr_i(LIST, D)$$

$Qr_i < 1$ מצביע על היחלשות Sr_i ואילו $Qr_i > 1$ מצביע על הגברת Sr_i.

(ג) נגדיר את **מדד ההגברה המרבית** Qr:

$$Qr \equiv \max\{Qr_1, Qr_2\}$$

3. נציג בקצרה את תוצאות מדידת הגברת/דעיכת הסיגנל המנורמל במבחן הפרמוטציות נוסף על תוצאות המדידות שבגוף המאמר. בכל טבלה אנו מציגים

- את ערכי Sr_i בתחום המורחב D'.
- את ערכי Qr_i ו־ Qr.

(א) בטבלה 7 מסוכמות התוצאות עבור רשימה *L2* והווריאנטים שלה:

טבלה 7

L2	*L2(ABCD)*	*L2(AB)*	
3.95E+5	**5.26E+6**	**3.33E+6**	$Sr_1(D')$
2.35E+2	**8.77E+3**	**1.67E+4**	Qr_1
2.20E+5	**5.08E+5**	**2.94E+5**	$Sr_2(D')$
1.51E−1	**5.94**	**1.62**	Qr_2
2.35E+2	**8.77E+3**	**1.67E+4**	Qr

(ב) בטבלה 8 מסוכמות התוצאות עבור רשימה *L2M* והווריאנטים שלה:

טבלה 8

L2M	*L2M(ABCD)*	*L2M(AB)*	
1.92E+7	1E+9	1E+9	$Sr_1(D')$
5.04E+4	1.86E+6	1.95E+7	Qr_1
4.76E+6	5.00E+7	4.35E+7	$Sr_2(D')$
1.20E+1	1.75E+2	1.15E+3	Qr_2
5.04E+4	1.86E+6	1.95E+7	Qr

(ג) בטבלה 9 מסוכמת מדידה דומה עבור רשימת *BM2* של מבב"ק:

טבלה 9

BM2	
1.73E+2	$Sr_1(D')$
1.38E−3	Qr_1
4.08E+3	$Sr_2(D')$
2.86E−3	Qr_2
2.86E−3	Qr

ג. בדיקות אחרות באמצעות "קיזוזים"

אפשר לחשוב על בדיקות שונות שבהן מקזזים את התרומה של צפני *ELS* בקטע *D* מן התוצאות בקטע המורחב *D′*. אפשר לעשות זאת בדרכים רבות:

א. לבדוק את המפגשים רק בקטע התוספת *D″* (הדבר דומה למי שירצה לבדוק את החיוניות של היד באמצעות חיתוכה מן הגוף. ראו עוד בנספח סעיף ח).
ב. מחיקת כל צפני *ELS* בקטע *D*.
ג. מחיקת תרומת צפני *ELS* המוכלים בקטע *D*.
ד. מחיקת תרומת מפגשים של זוגות צפני *ELS* המוכלים בקטע *D*.

האופציות ב-ד כרוכות בשינויים בפונקציה *c(w, w′)*. ישנם סוגים שונים של ניתוח כירורגי שניתן לבצע במעי הפונקציה להשגת אופציה זו או אחרת. וודאי ישנן גם אפשרויות

נוספות. אפילו אם נניח לשם הדיון, שהניתוח המבוקש אינו פוגע בהצפנה המשוערת (הנחה שאינה נכונה, ראו בנספח סעיף ח), הרי לפנינו מספר לא קטן, ואף לא ידוע היטב, של בחירות אפשריות.

1. הפתרון שבמאמר: לא לגעת בשום דבר – אין אופציות, אין ניתוחים כירורגיים, אין שום שינוי בפונקציה $c(w, w')$ המקורית. אפילו הפרמטרים של הפונקציה נשארו בדיוק כמו בניסוי המקורי של ור"ר. השגנו את הקיזוז המבוקש באמצעות השוואה לטקסטים דומים כמבואר במאמר. לכן, כל היתרונות (והחסרונות) והתרומות של צפני *ELS* בקטע D (ספר בראשית), מקוזזים באופן אוטומטי. כל המתחרים נהנים מאותה "רוח גבית" בדיוק.

2. נתבקשנו על ידי הקורא האנונימי לערוך מדידה ("לשם אינדיקציה בלבד") בקטע מן התחום המורחב D', המורכב מן הקטע שנוסף D'', ומקטע בסוף ספר בראשית שבו נמצאים צפני *ELS* העוברים ב"תפר" שבין ספר בראשית D לבין D''. הקורא הנ"ל העריך כי אם מורידים מתחילת התחום המורחב D' קטע שאורכו כארבע חמישיות מספר בראשית, אזי מתקבל הקטע המבוקש. דהיינו, הקטע המבוקש הוא כחמישית האחרונה בספר בראשית בתוספת הקטע D''.
(א) כמבואר לעיל, אין זו הגישה המחקרית שלנו. אנו מדדנו את ההסתברות המותנית לקבלת סיגנל כה גבוה בתחום המורחב, בהינתן הסיגנל המקורי בספר בראשית. לעומת זאת, כאן מוצע לערוך בדיקה **שונה לגמרי**: לבדוק אם ישנה הצפנה נפרדת בקטע התוספת יחד עם תחום "התפר".
(ב) בנספח סעיף ח1 הוספנו כי מלבד ההבדל המתודולוגי הנזכר ב(א), קיימת כאן בעייתיות נוספת הנובעת מהיות "מידת הקִרבה המכוילת" פונקציה **מסכמת**.

3. מבדיקה שערכנו עולה, כי אם רוצים שהקטע המבוקש יכלול את כל צפני *ELS* החוצים את "קו התפר", צריך לשייר קטע גדול יותר מספר בראשית, בערך שליש מאורכו. ואם משיירים כרבע מספר בראשית – כוללים את רוב הצפנים הנ"ל.

4. לפיכך יצרנו שלושה קטעים חלקיים שונים:
קטע א הוא D' ללא שני השלישים הראשונים של ספר בראשית.
קטע ב הוא D' ללא שלושת הרבעים הראשונים של ספר בראשית.
קטע ג הוא D' ללא ארבע החמישיות הראשונות של ספר בראשית.

5. התוצאות עבור *L2* ניתנות בטבלה 10:

טבלה 10

	L2(AB)		*L2(ABCD)*		*L2*	
	S_2	S_1	S_2	S_1	S_2	S_1
קטע א	9.43E+2	3.70E+3	3.82E+2	1.97E+3	2.92E+2	6.71E+2
קטע ב	1.01E+3	4.63E+3	2.22E+2	2.31E+3	3.07E+2	2.98E+3
קטע ג	6.41E+2	4.57E+2	1.09E+2	1.36E+2	1.77E+2	1.36E+2

6. התוצאות עבור *L2M* ניתנות בטבלה 11:

טבלה 11

	L2M(AB)		*L2M(ABCD)*		*L2M*	
	S_2	S_1	S_2	S_1	S_2	S_1
קטע א	8.85E+3	1.57E+5	1.10E+4	8.06E+5	2.69E+3	4.39E+4
קטע ב	1.24E+4	2.05E+5	6.25E+3	9.90E+5	4.44E+3	2.56E+5
קטע ג	3.69E+3	2.99E+3	1.43E+3	6.76E+3	1.33E+3	1.70E+3

7. הקורא הנ"ל ביקש להדגיש את הערכתו "שמובהקות של הצפנה בתפר ובהמשך היא בסביבות 0.001".

ד. בדיקות באמצעות הגדרות אחרות לסיגנל

הקורא האנונימי הציע להגדיר סיגנל חדש: $S' \equiv 1/\min\{P_{03}, P_2\}$,
כאשר הסטטיסטי P_{03} מוגדר עבור התוצאות בקטע [0, 0.3] כפי ש- P_1 (ראו הגדרת P_1 בנספח סעיף ב2) מוגדר עבור התוצאות בקטע [0, 0.2].

1. בחירתו של הקורא הנ"ל בסטטיסטי P_{03} נעשתה באופן אפוסטריורי לאחר שהיו לפניו ערכי $c(w, w')$ עבור זוגות הביטויים ברשימות *L2* ו-*L2M* בתחום המורחב. לדבריו, בחירה זו נבעה מהערכתו כי עבור תוצאות עבודת ור"ר ההצטברות בקטע [0, 0.3] היא ההצטברות החזקה ביותר. אומנם, בדיקה מדוקדקת יותר מעלה, כי גם עבור *L2* וגם עבור *L2M* ההצטברות המרבית היא בקטע [0, 0.25]. וכן הדבר עבור *L2(ABCD)* ועבור *L2M(ABCD)*.

2. לכן הגדרנו את הסטטיסטי P_{025} באופן דומה לנ"ל לגבי הקטע $[0,\ 0.25)$, והגדרנו את הסיגנל החדש להיות: $S' \equiv 1/\min\{P_{025},\ P_2\}$.

3. הגדרנו מדד הגברה חדש: $Q' \equiv S'(LIST,\ D') \ /\ S'(LIST,\ D)$.

4. נבדקו **10,000** טקסטים מעורבבים כמתואר במאמר. התוצאות בטבלה **12**:

טבלה 12

מספר הטקסטים המתחרים n שלהם ערך $Q''_j \geq Q'$ בתחרות ובה **10,000** מתחרים

רשימה	*L2M(AB)*	*L2M(ABCD)*	*L2M*
n	**4**	**2**	**20**

5. הערות

(א) הקורא מטעם המערכת הבחין נכוחה כי בתוצאות עבודת ור"ר קיימת הצטברות חזקה בקטע $[0,\ 0.3)$.

(1) אכן, הערך הגולמי (לא מנורמל) של הסטטיסטי P_{03} עבור הרשימה השנייה של ור"ר חזק פי **1,700** מזה של P_1, והערך הגולמי של P_{03} עבור תת־הרשימה $L2$ חזק פי **37,000** מזה של P_1 (לגבי הסטטיסטי P_{025} יש להכפיל פקטורים אלה בערך פי **2**).

(2) גם בתוצאות עבודת ור"ר הראשונה (הרשימה הראשונה של גדולי ישראל) קיימת הצטברות חזקה בקטע $[0,\ 0.3)$. ערכו הגולמי של P_{03} חזק פי **1,400** מזה של P_1.

(3) עובדות פשוטות אלה גלויות לכל עין הסוקרת את התפלגות התוצאות. ור"ר יכלו בקלות רבה לאמץ את הסטטיסטי P_{03} תחת הסטטיסטי P_1 ולהציג הצלחה גדולה בכמה סדרי גודל.

(4) לכן לפנינו הוכחה כי ור"ר עבדו ביושר לפי קנה מידה אפריורי, ולא חיפשו דרכים לאופטימיזציה של תוצאת המחקר, בניגוד גמור לטענות מבב"ק.

(ב) לגבי עצם ההצעה להשתמש עתה בסיגנל אחר, עמדתנו ברורה, שאין לשנות באופן אפוסטריורי את הגדרת הסיגנל. נוסף על כך, התפתח דיון מעניין בנוגע לשאלה מהי המדידה הנכונה (או לא נכונה) של התופעה. הדיון, שהביא להעמקה בנושא זה (מעבר למה שנכתב לעיל במבוא סעיף ב3), נמצא בסעיף מיוחד בנספח (סעיף ד).

נספח

א. בעניין הפולמוס

כאמור במבוא, מבב"ק טוענים במאמרם[15] כי הפריכו את עבודת ור"ר.[16] הם טוענים כי הכול מעשה תרמית ("תפירה" של נתונים), וכי המומחים החיצוניים שלמה זלמן הבלין והמנוח יעקב אורבך ז"ל שיתפו פעולה עם החוקרים במעשה התרמית.

מבב"ק טוענים: הוכחנו שאפשר לרמות, על ידי "תפירה" מוצהרת של רשימת שמות וכינויים עבור אותה קבוצת אישים שבעבודת ור"ר, "כמו" *L2*, שמצליחה לא פחות ממנה, אבל בספר "מלחמה ושלום".

בקצרה, הם טוענים: עשינו "אותו דבר" בספר "מלחמה ושלום".

1. הניסוי במאמר לפנינו מוכיח כי השתנות הסיגנל שיצרו בספר "מלחמה ושלום" עם הרחבת קטע המדידה המקורי (**דעיכה** בשלושה סדרי גודל), **שונה באופן קיצוני** מזו של הסיגנל בעבודת ור"ר עם הרחבה דומה בספר בראשית (**הגברה** לפחות בשלושה סדרי גודל). הגברת הסיגנל בעבודת ור"ר אינה מתיישבת עם "תפירת" נתונים.

2. לדעתנו, גם ללא השוואת השתנות הסיגנלים ניתן לברר כי מבב"ק כלל לא הצליחו לעשות "אותו דבר". אנו מפנים את הקורא למאמרים המנתחים את עבודתם הן מן הבחינה הלוגית והעובדתית,[17] והן מן הבחינה הביבליוגרפית והלשונית.[18]

15 לעיל הערה 3.

16 לעיל הערה 1.

17 ד' ויצטום, "על מדע ועל פרודיה: הפרכה גמורה של הטענה המרכזית של מבב"ק", ה'תשס"א. נמצא באתר הנ"ל, במדור "ההתנגדות למחקר" בקישור http://www.torahcode.co.il/paro_ heb.htm

18 ד' ויצטום, "סקירה של המאמצים להפריך את מחקר הצפנים בתורה", ה'תשנ"ט (עדכון אחרון: ה'תשס"ט), פרק ג (והמאמרים והמסמכים המקושרים אליו). נמצא באתר ובמדור הנ"ל בקישור http://www.torahcode.co.il/rev1_heb.htm. במיוחד ראו במקור הבא: ד' ויצטום, "הפרכה מופרכת, או: איך נכשלה רשימת הרבנים ב'מלחמה ושלום'", ה'תשנ"ח. חלק א בקישור http://www.torahcode.co.il/pdf_files/oppose/maindoch.pdf, וחלק ב בקישור http://www.torahcode.co.il/pdf_files/oppose/maindoc2.pdf וכן במסמכים הנלווים.

ב. המידות הכוללות לקִרבה

1. בניסוי המקורי של וו"ר (לעיל הערה 1, עמ' 436) הוגדרו ארבע "מידות כוללות לקִרבה" P_1, P_2, P_3 ו- P_4. שם (עמ' 431) מבואר כי:

- מידות P_1 ו-P_2 הוגדרו כשני סטטיסטים שונים לגבי *LIST 2*.
- כאשר מידות P_1 ו-P_2 מיושמות לגבי *L2*, הן מסומנות כ-P_3 ו-P_4.

במילים אחרות: ישנם שני אופראטורים, $\boldsymbol{P_1}$ ו-$\boldsymbol{P_2}$.
היישום שלהם על רשימת הנתונים השלמה נותן את P_1 ו-P_2:

$$P_1 \equiv \boldsymbol{P_1}(LIST\ 2),\ \ P_2 \equiv \boldsymbol{P_2}(LIST\ 2)$$

והיישום שלהם על תת-הרשימה *L2* נותן את P_3 ו-P_4:

$$P_3 \equiv \boldsymbol{P_1}(L2), \qquad P_4 \equiv \boldsymbol{P_2}(L2)$$

בניסוי הנוכחי קיימת רק רשימת נתונים אחת *L2*, לכן איננו זקוקים עוד לאינדקסים 3 ו-4, ונסמן כאן:

$$P_1 \equiv \boldsymbol{P_1}(L2), \qquad P_2 \equiv \boldsymbol{P_2}(L2)$$

וכן לגבי *BM2*.

2. להקל על הקורא נביא כאן את הגדרת $P1$ ו-$P2$ מור"ר.

(א) הגדרת מידת "הנטייה הכוללת לקִרבה" P_1.
לפי מידה זו מונים את מספר התוצאות ב"אזור ההצלחה", אשר הוגדר (שרירותית) כמרווח בין 0 ל-0.2, ומחשבים מה הסיכוי לקבל באקראי את הערך המתקבל. המדגם העומד לבדיקה הוא קבוצה של זוגות ביטויים. "מידת הקִרבה המכוילת" של כל זוג ביטויים (w, w') ניתנת לחישוב על ידי $c(w, w')$. כך מקבלים N מספרים, שכל אחד מהם הוא בין 0 ל-1. נניח שמספר הזוגות (w, w') עבורם $c(w, w') \le 1/5$ הוא k. נגדיר

$$P_1 \equiv \sum_{j=k}^{N} \binom{N}{j} \left(\frac{1}{5}\right)^j \left(\frac{4}{5}\right)^{N-j}$$

כדי להבין הגדרה זאת, נשים לב לכך, שאם המספרים $c(w, w')$ הם משתנים אקראיים בלתי-תלויים המתפלגים בצורה אחידה (אוניפורמית) בין 0 ל-1, אזי P_1 היא ההסתברות, כי

לפחות k מ־N המספרים קטנים או שווים ל־0.2. אומנם, איננו עושים כאן כל שימוש בהנחה של אחידות ואי־תלות. לכן, אף על פי ש־P_1 מכויל כהסתברות, הוא משמש רק כמדד סידורי. נשים לב, כי המידה P_1 מתעלמת מכל ערכי $c(w,w')$ הגדולים מ־0.2, ומעניקה אותו המשקל לכל ערכי $c(w,w')$ הקטנים מ־0.2. כלומר, אנו מתמקדים במפגשים המצליחים ללא הבחנה באיכותם, ואיננו מתעניינים לדעת באיזו מידה נכשלו אלה שלא הצליחו.

(ב) הגדרת מידת "הנטייה הכוללת לקרבה" P_2.

המידה P_2 נבנתה כך שהיא רגישה לגודלם של **כל** המספרים $c(w,w')$. המשמעות של P_2 היא ש**אם** המספרים $c(w,w')$ הם משתנים אקראיים בלתי־תלויים המתפלגים בצורה אחידה בין 0 ל־1, אזי P_2 היא ההסתברות, שמכפלת ערכי $c(w,w')$ תהיה קטנה כפי שהיא, או קטנה מזה.

אנו מחשבים את המכפלה $\prod c(w,w')$ ואחר כך אנו מגדירים

$$P_2 \equiv F^N\left(\prod c(w,w')\right)$$

כאשר
$$F^N(X) \equiv X\left(1-\ln X+\frac{(-\ln X)^2}{2!}+\cdots+\frac{(-\ln X)^{N-1}}{(N-1)!}\right)$$

כדי להבין הגדרה זו, נשים לב כי אם $x_1, x_2, \ldots, x_N$ הם משתנים אקראיים בלתי־תלויים המתפלגים בצורה אחידה בין 0 ל־1, אזי ההתפלגות של מכפלתם $X \equiv x_1 x_2 \cdots x_N$ ניתנת על ידי

$$\Pr(X \le X_0) = F^N(X_0)$$

[הדבר נובע מתוצאה (3.5) של פלר[19] מכיוון ש־$-\ln x_i$ מתפלגים באופן אקספוננציאלי,

וגם
$$-\ln X = \sum_i (-\ln x_i)\text{].}$$

אומנם, איננו עושים כאן כל שימוש בהנחה של אחידות ואי־תלות. לכן, אף על פי ש־P_2 מכוילת כהסתברות, היא משמשת רק כמדד סידורי.

19 הספר: W. Feller, *An Introduction to Probability Theory and Its Applications* 2 Wiley, New York 1966

ג. מדידת מובהקות הסיגנל

1. כאשר ור"ר ביצעו בספר בראשית מדידות של רשימות הנתונים, הם מדדו את האפקט הכולל באמצעות מידות כוללות P_i. כדי למדוד את מובהקות התוצאות היה צריך לחזור על אותן מדידות בטקסטים רבים "דומים". דבר זה לא היה בהישג ידם של ור"ר לפני 25 שנים (מבחינת כמות החישובים).[20] הוצע להם לערוך מבחן פרמוטציות, אשר פרטיו סוכמו בין פרסי דיאקוניס וישראל אומן, ואושרו לפני עריכת הניסוי על ידי ארבעה סטטיסטיקאים נודעים נוספים. מבחן זה בוצע במאמר שהוזכר לעיל, הערה 1. מבחן הפרמוטציות מעריך את מובהקות הסטטיסטים P_i שנתקבלו בספר בראשית עבור רשימת הנתונים המקורית (מפגשים של שמות אישים עם תאריכי הלידה ו/או הפטירה שלהם), באמצעות השוואה לערכי הסטטיסטים P'_i המתקבלים בספר בראשית עבור הרבה רשימות נתונים "מעורבבים" (מפגשים של אותם שמות אישים עם תאריכי הלידה ו/או הפטירה של אישים אחרים מן הרשימה המקורית).

2. כחמש וחצי שנים לאחר ביצוע מבחן הפרמוטציות המתואר במאמר שבהערה 1, סבר אליהו ריפס שהתקדמותם המהירה של אמצעי החישוב מאפשרת לשוב לרעיון הפשוט ביותר של מדידת המובהקות: באמצעות רנדומיזציה של טקסטים, דהיינו, באמצעות השוואה לטקסטים רבים "דומים".

במכתב לדוד קשדן[21] הציע אליהו ריפס את העקרונות הבאים לגבי מדידת מובהקות של רשימה בספר בראשית:

- להשתמש ב־1000 (או יותר) טקסטים T_j, כל T_j נוצר על ידי ערבוב אקראי של מילים בתוך פסוקים (הדבר נעשה בדיוק כפי שנעשה במאמר של ור"ר לגבי טקסט ביקורת U).
- להעריך את מובהקות הסטטיסטים P_i בספר בראשית, באמצעות השוואתם לערכי P'_i המתקבלים בטקסטים T_j.

מבחן מסוג זה יכונה להלן "**מבחן טקסטים**".

3. השוואה בין מבחן הפרמוטציות למבחן הטקסטים – שיקול כללי

כל אחד משני המבחנים מודד דבר שונה:

20 להרצת רשימה בגודלה של *LIST2* בספר בראשית נדרשו כשלושה שבועות.

21 בדואר אלקטרוני מיום 15 במאי שעה 16:40:13, שנת 1997 (למניינם).

- מבחן הפרמוטציות מודד עד כמה "מוצלחים" המפגשים של שמות האישים עם תאריכי הלידה ו/או הפטירה שלהם **יותר** מן המפגשים של אותם שמות אישים עם תאריכים אחרים (שהם תאריכי לידה/פטירה של אישים אחרים מאותה רשימת נתונים).
- מבחן הטקסטים מודד מה הסיכוי לקבל במקרה P_i כה קטנים. דהיינו: מה הסיכוי שהצטברו במקרה כל כך הרבה מפגשים "מוצלחים".

במילים אחרות: מבחן הטקסטים מודד את מובהקות הכמות המצטברת של מפגשים "מוצלחים", ואילו מבחן הפרמוטציות מודד את המובהקות של "ספציפיות הצופן".

לכן אנו מצפים אפריורי להבדל בתוצאות שני המבחנים עבור אותה רשימת נתונים, שהרי הם מודדים דברים שונים.

4. דוגמה:

רשימת הנתונים *BM2* שנמדדה בעבודת מבב"ק בספר "מלחמה ושלום" הוכנה על ידי מניפולציה מוצהרת. המניפולציה נעשתה כדי להשיג מובהקות חזקה במבחן הפרמוטציות. ואכן, הושגה מובהקות חזקה מאוד עבור הסטטיסטי הנמדד:

$$p = 7\text{E}{-7}.$$

לעומת זאת, כאשר מודדים את המובהקות של אותו סטטיסטי במבחן טקסטים, מתברר כי ב־12 מתוך 10,000 טקסטים "דומים" (שנוצרו מערבוב אקראי של הטקסט המקורי) הושגה תוצאה חזקה יותר. לכן, המובהקות חלשה בהרבה:

$$p = 1.2\text{E}{-3}.$$

אכן, קיים הבדל גדול מאוד בין התוצאות.

5. המבחן המתאים למצבו העכשווי של מחקר הצפנים בתורה

המחקר המדעי של הצפנים בתורה, עד כמה שבא לידי ביטוי במדידת רשימות גדולות של נתונים, עסק עד כה בהוכחת קיומו של צופן מסוים. הוא לא עסק באפיון מדויק או בפענוח של הצפנים. הדבר הוצהר במפורש, בהקדמת ור"ר (בעמוד הראשון של המאמר. בעניין זה ראו גם לעיל הערה 2, עמ' 52-57). לכך מתאים דווקא מבחן הטקסטים, ולא מבחן פרמוטציות המודד את המובהקות של "ספציפיות הצופן" (ור"ר השתמשו במבחן פרמוטציות רק באין ברירה, כמבואר לעיל).

6. השוואה בין מבחן הפרמוטציות למבחן הטקסטים – שיקול ספציפי

בדרך כלל משתמשים במבחן פרמוטציות במקרה הבא:

ישנה קבוצה א שלגביה השערת המחקר טוענת טענה מסוימת, וקבוצת ביקורת ב שלגביה אין השערת המחקר טוענת דבר. לפי השערת האפס אין תופעה גם בא וגם בב. כלומר יש ויכוח בין השערת האפס להשערת המחקר ביחס לקבוצה א ויש הסכמה ביחס לקבוצה ב.

(א) בוחרים סטטיסטי לאפיון התופעה.

(ב) מגרילים באקראי פריטים (בכמות של פריטי א) מקבוצה ג, שהיא האיחוד של א וב, ומודדים את הסטטיסטי לגביהם.

(ג) חוזרים על (ב) פעמים רבות, ורואים את המיקום של קבוצה א בהתפלגות ערכי הסטטיסטי.

לעומת זאת, מבחן הפרמוטציות שהוצע לור"ר[22] היה שונה.

המדגם המקורי הוא קבוצה א. במבחן הפרמוטציות, אנו מצמידים לכל אדם תאריך של חברו לרשימה. קבוצת כל הזוגות של שמות אישים שלהם מוצמדים תאריכים של אחרים אמורה להיות קבוצת הביקורת ב.

אבל כאן, השערת המחקר אינה מסכימה שלגבי ב ודאי אין תופעה. וכי אפשר לטעון (למשל) כי י"ד בניסן, יום הולדתו של הרמב"ם, אינו קשור לראב"ד? וכי הראב"ד לא עשה (או חידש) דברים מהותיים בערב פסח בכל ימי חייו? וכי יום כ"ו בכסלו (נר שני של חנוכה), יום מותו של הראב"ד, לא היה משמעותי בחיי הרמב"ם? ומי בכלל יכול לטעון, כי תאריך מסוים אינו קשור לאישיות מסוימת?!

לכן, ייתכן מאוד שהמדגם המקורי מתחרה עם מדגמים משובשים המכילים זוגות מקבוצה ב האמורים להצליח אם קיימת התופעה, והדבר צפוי *להחליש* את המובהקות הנמדדת.

7. במבחן הטקסטים לא קיימת בעיה זו.

22 לעיל הערה 1.

ד. בנוגע להגדרות אפוסטריוריות אחרות של הסיגנל

מאחר שעלתה לדיון השאלה מהי הדרך הנכונה למדוד את "התופעה", יש להבהיר את המושגים, ובראש ובראשונה את המונח "התופעה". אסביר כאן מהו נושא המחקר שעליו דן המאמר. נגיע אליו בכמה שלבים בסוד הצמצום המדעי.

1. התופעה הרחבה: תופעת הרמז בתורה (ראו מיון דרכי הרמז ב"פרדס רמונים" להרמ"ק זללה"ה, שער ל – שער הצרוף).
2. אנו מצטמצמים לתופעה מצומצמת – אחת מדרכי הרמז – דילוג שווה של אותיות.
3. תופעה זו מורכבת מאוד ומכילה ריבוי עניינים. למשל:
   (א) תופעה של מילים בדילוגים שווים מיוחדים.
   (ב) תופעה של צבירי מילים בדילוגים שווים.
   (ג) תופעה של ביטויים ארוכים.
   (ד) תופעה של מפגשים של ביטויים בדילוגים מינימליים עם הטקסט.
   (ה) תופעה של מפגשים בין ביטויים בדילוגים מינימליים.
   (ו) קריאה פשוטה על פני טבלאות.
   (ז) יש עוד תופעות, אך דיינו בזה.
4. עוד לפני עבודת ור"ר הראשונה צמצמנו את תופעה 3(ה) כדלקמן:
   "מידת הקִרבה המכוילת" (היא הפונקציה $c(w,w')$) היא הקובעת את איכות המפגש בין שני ביטויים בדילוג שווה על פני טבלאות דו־ממדיות. [זה כולל פרטים רבים שהם עצמם צמצום של המושג "מפגש".]
5. בשלב הבא צמצמנו את המושג "מפגש" למפגשים "מוצלחים", מפגשים כאלה שעבורם ערך "מידת הקִרבה המכוילת" הוא בקטע $[0,\ 0.2)$. התופעה המצומצמת היא אוסף המפגשים "המוצלחים" לפי הגדרה זו. אנו מתעניינים ועוקבים דווקא אחרי מפגשים מקטגוריה זו [הסטטיסטי P_1 נועד לבדוק בניסוי מבוקר מה השכיחות של "מפגשים מוצלחים"].
6. אחר כך הוצע לבדוק במקביל גם את שכיחות המפגשים "המוצלחים מאוד", שעבורם ערך "מידת הקִרבה המכוילת" קרוב יותר ל־0, אך לא באמצעות קביעת קטע מסוים מראש אלא באופן אחר [הסטטיסטי P_2 נועד לטפל בזאת].
7. בעבודת ור"ר הראשונה נמצאים צמצומים 5 ו־6.
   לפיכך הסטטיסטי P_1 מודד תופעה מצומצמת אך מוגדרת היטב: שכיחות המפגשים מסוג מסוים.
   השימוש בסטטיסטי P_1 נעשה עוד לפני עבודת ור"ר הראשונה, וגם בכל העבודות הרבות שעשיתי אחר כך. בעבודות רבות (ראו על כך במבוא למאמר הנוכחי) P_1 הראה כי התופעה המצומצמת המתוארת בסעיף 5 אכן קיימת במובהקות עצומה.

(א) כבר מאז עבודת ור"ר הראשונה (הרשימה הראשונה של גדולי ישראל) התברר כי אם היו מגדירים מפגשים "מוצלחים" כאלה שעבורם ערך "מידת הקִרבה המכוילת" הוא בקטע [0, 0.3), היו מקבלים ערך חזק פי 1,400 מזה שנתקבל עבור הגדרה 5 דלעיל. למרות זאת ור"ר לא אימצו הגדרה חדשה זו למפגשים "מוצלחים", ולא רק משום שהיו נאמנים להגדרה האפריורית ולא חיפשו "רווחים קלים". וכן עשו בעבודת הרשימה השנייה בה יכלו "להרוויח בקלות" פקטור של 1,700 לטובתם.

(ב) הסיבה היא שהסדר הלוגי בעבודה מדעית (ובכלל) הוא: **קודם** מגדירים תופעה **ואחר כך** מודדים אותה. ההגדרה בסעיף 5 לעיל של "מפגשים מוצלחים" באה לאחר בחינה של מפגשים קודמים ומדידתם באמצעות "מידת הקִרבה המכוילת". רצינו לאפיין אותם מפגשים שהם אכן מפגשים נראים היטב לעין (בהצגתם בטבלאות דו־ממדיות) – ולא רק ברמה של מספרים ערטילאיים בעלמא.

(ג) המאמר מתאר מה קורה לכל אחת משתי התופעות המוזכרות בסעיף 7, באמצעות שני הסטטיסטים, כאשר עוברים לתחום המורחב. הדבר בא לענות על טענות המבקרים כי נתוני *L2* "נתפרו" בהתאם לסטטיסטים אלה.

(ד) והנה, בחישוב אפוסטריורי עבור קבוצת זוגות חלקית של *L2M* נמצא כי כמות הזוגות במרווח [0.2, 0.3) עבור התחום המורחב קטנה מזו עבור ספר בראשית. מה משמעות הדבר? – אין בזה יותר מן האמירה כי ההצטברות במרווח הנ"ל של מפגשים בקטגוריה **אחרת** מזו המבוקשת אפריורי, דעכה במעבר לתחום המורחב.

(ה) לעומת זאת, מקובל לחלוטין כי לקח או תצפית הנלמדים באופן אפוסטריורי מתוצאות מחקר מסוים, יכולים לשמש במחקר עתידי, לגביו הם יהוו בבחינת הגדרות אפריוריות.

ה. פרטי המדידה הנוספת (מדידה 2ב בחלק א)

1. פרמוטציה PER1

זו הייתה פרמוטציה מס׳ 808,836 במרוץ המקורי:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	17	8	13	25	19	30	29	23	14	5	18	11	10	4	32

17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
15	9	2	26	28	12	3	16	24	6	21	22	20	7	31	27

(א) התוצאות בספר בראשית (G):

$$S_1(PER1, G) = 1.2680\text{E+7}, \qquad S_2(PER1, G) = 1.8485\text{E+9}.$$

(ב) התוצאות בקטע D':

$$S_1(PER1, D') = 1.1910\text{E+1}, \qquad S_2(PER1, D') = 1.9802\text{E+2}.$$

מכאן:

$$Q_1 = 9.3927\text{E}{-7}, \qquad Q_2 = 1.0713\text{E}{-7}, \qquad Q = \max\{Q_1, Q_2\} = 9.3927\text{E}{-7}.$$

2. פרמוטציה PER2

זו הייתה פרמוטציה מס' 788,884 במרוץ המקורי:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	20	32	11	17	19	8	29	7	9	15	18	26	14	6	28

17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
10	16	2	3	27	22	23	21	4	31	25	30	13	24	12	5

(א) התוצאות בספר בראשית (G):

$$S_1(PER2, G) = 3.380\text{E+7}, \qquad S_2(PER2, G) = 6.8970\text{E+8}.$$

(ב) התוצאות בקטע D':

$$S_1(PER2, D') = 3.6614\text{E+2}, \qquad S_2(PER2, D') = 1.2376\text{E+4}.$$

מכאן:

$$Q_1 = 1.0833\text{E}{-5}, \qquad Q_2 = 1.7945\text{E}{-5}, \qquad Q = \max\{Q_1, Q_2\} = 1.7945\text{E}{-5}.$$

3. פרמוטציה PER3

זו הייתה פרמוטציה מס׳ 777,442 במרוץ המקורי:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
24	20	17	12	28	19	31	18	25	8	7	27	9	14	32	10

17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
22	15	2	4	3	6	23	1	29	11	21	5	13	30	16	26

(א) התוצאות בספר בראשית (*G*):

$$S_1(PER3, G) = 1.1392\text{E+8}, \qquad S_2(PER3, G) = 1.47\text{E+8}.$$

(ב) התוצאות בקטע *D′*:

$$S_1(PER3, D') = 3.2854\text{E+3}, \qquad S_2(PER3, D') = 2.86\text{E+5}.$$

מכאן:

$$Q_1 = 2.8838\text{E}{-5}, \qquad Q_2 = 1.95\text{E}{-3}, \qquad Q = \max\{Q_1, Q_2\} = 1.95\text{E}{-3}.$$

ו. פרטים בנוגע ל-*L2M*

במאמרם העלו מבב״ק טענות אחדות בנוגע לתאריכים ולצורות התאריך שננקטו ברשימות הנתונים של ור״ר. הם טענו כי לור״ר היה חופש פעולה בתחום זה, וכי ור״ר נקטו תמיד בבחירה שנתנה אופטימיזציה של התוצאות.

במאמרנו ״על בחירת התאריכים למדגמים של ור״ר״,[23] דחינו את טענות מבב״ק. לענייננו, הראינו כי מטענותיהם רלוונטית אך ורק טענה אחת ויחידה לגבי תת הרשימה *L2*: הטענה כי עבור 4 אישים (מתוך 32 האישים ברשימת הנתונים) יש להוסיף או לשנות תאריך:

1. תיקון תאריך עבור אישיות מס׳ 20 (י׳ איר במקום ד׳ איר).
2. תיקון תאריך עבור אישיות מס׳ 21 (א׳ איר במקום ל׳ ניסן).

23 ד׳ ויצטום, ״על בחירת התאריכים למדגמים של ור״ר״, ה׳תשס״א. ראו באתר הנ״ל בקישור: http://www.torahcode.co.il/date_heb.htm

3. הוספה של תאריך לידה עבור אישיות מס׳ 24 (ט״ו סיון, י/ה סיון).
4. הוספה של תאריך לידה עבור אישיות מס׳ 30 (ט״ו תמוז, י/ה תמוז).

נגדיר את הרשימה *L2M* כתת הרשימה *L2* עם השינויים הנ״ל בתאריכים (כל שאר הפרטים, בפרט השמות והכינויים – נשארים כפי שהם, ללא שינוי). אגב, בניגוד לטענת מבב״ק, מתברר מן התוצאות כי השינויים בתאריכים אינם גורמים לירידה בסיגנל.

ז. פרטים בעניין תצורות התאריך

1. רשימות הנתונים ששימשו לניסויי ור״ר הן רשימות של זוגות ביטויים. בכל זוג, אחד מבני הזוג הוא שם או כינוי של אישיות מסוימת, ובן זוגו הוא תאריך הלידה או הפטירה של אישיות זו.

כל תאריך הוצג בשלוש תצורות. למשל היום הראשון בחודש תשרי הוצג כך:

A - ״א׳ תשרי״,

B - ״בא׳ תשרי״,

C - ״א׳ בתשרי״.

הדבר נקבע באופן אפריורי לפני הכנת רשימת הנתונים הראשונה, על ידי הבלשן יעקב אורבך ז״ל ששימש יועץ לור״ר בעניינים הלשוניים. לפי הנחיה זו הוכנה רשימת הנתונים הראשונה ובוצע עליה ניסוי. פירוט הרשימה הראשונה ותוצאות הניסוי פורסמו בפרה-פרינט. בעקבות פרסום זה נדרשו ור״ר להכין רשימת נתונים נוספת. רשימת נתונים נוספת זו, הידועה כ־*LIST 2*, הוכנה אף היא באותו אופן, והיא הרשימה אשר לגביה נערך מבחן הפרמוטציות שתוצאותיו פורסמו במאמר שהוזכר לעיל בהערה 1.

2. טענת מבב״ק, שהועלתה ממש בתחילת הוויכוח, והייתה השאלה הראשונה הנוגעת לגוף הנתונים,[24] היא: ״מדוע ברישום התאריכים השתמשתם רק ב־3 צורות מ־4 (א׳ תשרי, א׳ בתשרי, בא׳ תשרי, אך לא בא׳ בתשרי)?״ הם הוסיפו כי באנציקלופדיה מרגליות,[25] שהייתה נקודת המוצא לרשימת האישים, נעשה שימוש בארבע הצורות.

טענה זו חזרה בהבלטה בכל פרסומיהם. למעשה, הם טענו כי נבחרו רק שלוש הצורות ABC ולא D, כי כך השיגו ור״ר תוצאה טובה יותר.

24 ״מסמך 2״ בקישור http://www.torahcode.co.il/pdf_files/oppose/docum2h.pdf. ראו שם את השאלה הרביעית.

25 מ׳ מרגליות (עורך), אנציקלופדיה לתולדות גדולי ישראל, תל אביב 1961.

3. ובכן, איננו יודעים את שיקוליו של יעקב אורבך ז"ל, בפרט איננו יודעים אם בדק או התחשב בכתוב באנציקלופדיה מרגליות. אבל אנו יודעים שלושה דברים בבירור:
(א) כי זו בחירה ראויה. למשל, כאשר האנציקלופדיה העברית[26] השתמשה בצורות סטנדרטיות בערכי האישים שבמדגם, היא השתמשה בדיוק בשלוש הצורות ABC, ולא בצורה D.
(ב) כי אי אפשר להכחיש שזו בחירה אפריורית לגבי הרשימה השנייה (ראו סעיף 1 לעיל).
(ג) כי לפי אמות המידה הסטטיסטיות ששימשו למדידת המובהקות בזמן יצירת הרשימה הראשונה – שימוש בתצורה D היה דווקא משפר את המובהקות! (ראו המאמר המצוין בהערה 23).

4. השאלה לגבי צורות התאריך מעניינת כשלעצמה (ראו מסקנה ב לחלק ב).

ח. שתי הערות לגבי הנדון במאמר חלק ג סעיף ג

1. יש לקחת בחשבון כי "מידת הקִרבה המכוילת" **מסכמת** את המפגשים השונים של צפני *ELS* עבור זוג מילים. ולכן ייתכן בהחלט שהצלחה עבור זוג מילים מסוים בתחום המורחב נובעת דווקא **מסיכום** ההצלחה בקטע המקורי עם הצלחה חדשה בתוספת ואולי גם "בתפר" – בבחינת **זה וזה גורם**. זה אכן מאפיין עשרות זוגות המצליחים בתחום המורחב. הפעלת המתווה המוצע מחסלת הצלחות כאלה.

2. אדם שאינו עוסק ממש בפרטי המחקר הנדון עשוי לחשוב כי "ניתוח כירורגי" כזה או אחר במעי "מידת הקִרבה המכוילת" אפשרי בהחלט לצורך קיזוז השפעת המפגשים המקוריים בספר בראשית על התוצאה בתחום המורחב. אבל אין הנחה זו נכונה. אפילו ניתוח "עדין" במיוחד, כמו באופציה ד – "מחיקת תרומת מפגשים של זוגות צפני *ELS* המוכלים בקטע *D*" – עשויה לפגוע בהצפנה המשוערת.
לדוגמה: נניח שמפגש של זוג צפני *ELS* המוכלים בקטע *D* נמצא ברביע האחרון של *D*, ואף שהמפגש ביניהם בטבלאות דו־ממדיות מוצלח מאוד מבחינה גיאומטרית, הוא קיבל דירוג שולי משום שצפנים אלה מינימליים רק על חלק קטן בסוף קטע *D* (זה מרכיב חשוב מאוד ב"מידת הקִרבה"). והנה מתברר כי כאשר מרחיבים את הקטע, בקטע התוספת (או

26 האנציקלופדיה העברית, ירושלים ה'תשמ"א.

בתפר) אין הופעות בעלות דילוג קצר יותר מהם. בזה הם הפכו להיות בעלי תחום מינימליות גדול, דבר שהפך את המפגש השולי קודם לכן, שהוא אכן מוכל בקטע D, למפגש מכריע. הניתוח התמים המוצע באופציה זו דיו לחסל הצפנה מסוג זה.

הכרת תודה

החישובים נעשו באמצעות תוכנה של יעקב רוזנברג. רוברט הראליק עורר אותי לחשב את המובהקות באמצעות השוואה לטקסטים "דומים". יִשְׂרון יצחק לוי ומיכל לוי עזרו בהכנת הקטע שנוסף ל"מלחמה ושלום". שלום סרברניק העלה, כדרכו, רעיונות מועילים. אני מודה לקורא המאמר מטעם המערכת שהשקיע זמן, מחשבה ועבודה בביקורת טיוטת המאמר ותרם בכך הרבה להבהרת כמה וכמה עניינים חשובים. תודה למבב"ק על שהסבו את תשומת ליבנו לעניין תצורות התאריך.