

מיסמך 2 : בר הלל ובר נתן שואלים – ויצטום וריפס משיבים

מיסמך זה כולל שלושה חלקים :

1. רשימה של שאלות שהוצגו על ידי ד"ר דרור בר-נתן ופרופ' מיה בר-הלל לדורות ויצטום ופרופ'

אליהו ריפס.

2. תשובותיהם של דורות ויצטום ופרופ' אליהו ריפס.

3. אפילוג.

ירושלים, 27.11.96

על פי בקשתו של ישראל אומן, אנו מעלים על הכתב כמה שאלות אשר התעוררו אצלנו תוך קריאת המאמר והדו"חות על גילוי קוד הדילוגים בתורה, ותוך הנסיונות לשחזר את העבודה המדווחת. נשמח לקבל את התשובות בכתב.

1. היינו רוצים לדעת פרטים על מהלך המחקר מבחינה כרונולוגית : למשל, מי גילה לראשונה את

הקרבה הגדולה של זוגות השמות-תאריכים, ומתי? מתי גובשה לראשונה הרשימה הראשונה? ומתי

השניה? מי דרש את הרשימה השניה? מתי הוגשה העבודה לראשונה לכתב עת מדעי, ואיזה? איך עלה

על דעתכם לבדוק את התאריכים המדווחים על ידי מרגליות? האם בדקתם את כל התאריכים, או רק

כאלה שהיו, משום מה, חשודים? או רק כאלה שידעתם מראש, איכשהו, על קיומו של תאריך אחר?

וכו'.

2. היינו שמחים לקבל מכם את הקלדת ספר "שמות" ויתר החומש, כדי לערוך עליהם בדיקות משלנו,

תוך הבטחה שאנחנו ואתם משתמשים באותו טקסט בדיוק. נודה לכם אם תואילו לתת לנו העתק

מהקלדתכם. כמו כן היינו רוצים לקבל את המובהקות של הרשימה הראשונה על פי חישוביכם,

למטרות השוואה עם חישובינו שלנו, כדי להבטיח שאין שגיאה בתוכנה.

3. אנציקלופדיית מרגליות שהשתמשתם בה יצאה לאור ב-1961. האם יש מקור מאוחר יותר לשמות של

גדולי ישראל? האם יש באיזה מקום data base ממוחשב, אולי? או מכל מקום, מקור בו ניתן, למשל,

לקבל רשימת תאריכי לידה ומוות של גדולי ישראל—לא רק אלה שהופיעו ברשימות שלכם, אלא באופן

כללי? כיצד אנשים כמונו יכולים לקבל רשימה עדכנית ומשופרת על פני זו של מרגלית?

3. מדוע ברישום התאריכים השתמשתם רק ב-3 צורות מ-4 (א תשרי, א בתשרי, בא תשרי, אך לא בא בתשרי)? ובכלל מדוע לא השתמשתם בכל צורות הרישום (כגון: מרחשון, אייר, כסליו, ר"ח אדר, ראש השנה, יא לתשרי, וכו')?

4. עשינו ניסיון ליצור בעצמנו רשימת גדולי ישראל מן האנציקלופדיה, על פי הקריטריון המדווח. היו כמה פערים שאנו רוצים לשאול עליהם:

--לרבי אהרון מקרלין יש טור ו-11 שורות, ולא 17 כנדרש. מדוע נכלל?

--הגאון מוילנא נפטר בג' דחול המועד סוכות. האין זה יח בתשרי (אצלכם: יט בתשרי)?

--לרבי דוד גאנו יש רק 2 טורים ו-31 שורות, ולפיכך הוא היה צריך להכלל ברשימה השניה. מדוע נכנס לראשונה?

--לרבי יהודה עייאש יש רק טור+15 שורות, למה נכלל?

--לרבי יהוסף הנגיד יש רק טור+16 שורות, למה נכלל? מה עוד שתאריכיו נפסלו.

-- את תאריך הלידה של ישראל בעל שם טוב(בעש"ט) לא מצאנו במרגליות, התוכלו להפנותנו אליו?

-- מדוע לא נכלל רבי מאיר אייזנשטט (מהר"ם א"ש), שלו בדיוק 1.5 טורים, היינו טור+17 שורות, ברשימה השניה?

אלו הן שאלותינו לפי שעה. תודה.

מיה בר הלל דרור בר נתן

ט"ו טבת התשנ"ז

ב"ה

לכב' פרופסור ישראל אומן שלוי רב:

הרי תשובה למכתב מיום 27 בנובמ. 96 (לפי מניינם), החתום ע"י פרופסור מיה בר-הלל וד"ר דרור בר-נתן. (מיספור הסעיפים לפי מיספור השאלות).

1. נשאלנו של מהלך המחקר מבחינה כרונולוגית. הרי מהלך העניינים:

א. באביב של שנת התשמ"ה גילינו שתי תופעות בנוגע למד"שים (מלים בדילוג שווה):

(i) נטייה למיפגשים בין מד"שים, שהם מינימליים על פני קטעים גדולים בספר בראשית, ואשר קיים קשר מושגי ביניהם, על פני טבלאות דו-ממדיות הנקבעות ע"י המד"שים עצמם (כפי שמתואר במאמר).

(i.i) נטייה למיפגשים בין מד"שים כנ"ל לבין ביטויים בטקסט על פני טבלאות כנ"ל (כפי שמתואר במאמר השני).

ב. בפגישה שנערכה בחודש סיון התשמ"ה, הציע אליהו לעבור מהתרשמות אינטואיטיבית למדידה כמותית. לצורך זה הוגדר פונקציונאל המסומן במאמרנו ב- $c(w, w')$. תוכנית למדידה כמותית כזאת הוכנה במהלך הקיץ ע"י יואב רוזנברג, ונוסתה ע"י מדידת מיפגשים ממאגר המימצאים הקיימים. ג. בסוף התשמ"ה הציע אליהו, שנבדוק את עצמנו ע"י מדידת רשימה א-פריורית של זוגות ביטויים. הוא הציע זוגות מסוג אדם - תאריך (לידה או פטירה) משום שהקשר המושגי מוגדר הייטב. הוא הציע שהאנשים יהיו גדולי חכמי ישראל משום שאם באמת קיימת תופעה כזאת, הרי סביר שגדולי חכמי התורה יזכו ל"סיקור" נרחב יותר. עד לאותו זמן לא נבדק נושא זה לגבי תופעות (I) ו-(i.i) וממילא לא היתה ידיעה אם קיימת נטייה למפגשים אדם-תאריך. לעומת זאת היתה דוגמא לקשר כזה, בעבודה קודמת של דורון שהיתה הרחבה של רעיונו של הרב וייסמנדל ז"ל לגבי מד"שים בדילוגים מיוחדים. (בדוגמא זו רואים את הצרוף הרצל-הולדו-בי"איר).

ד. אליהו הציג את תוכניתו לבדוק רשימה של גדולי חכמי התורה, לגבי מיפגש שמוותיהם עם תאריכי הלידה והפטירה, לפני שני אנשים (לפחות): פרופ' זווין מן המחלקה לפיסיקה, ופרופ' שנקובסקי מן המחלקה למתמטיקה. הוא אף הסביר להם את הפונקציונאל. תגובתו של פרופ' שנקובסקי היתה (בהומור מסוים): "אינני יודע אם לאחל לך הצלחה..."

ה. דורון היה צריך לטפל בהכנת הרשימה ומדידתה. כיוון שלא היה לו ניסיון קודם בתחום הביבליוגרפי והלשוני הנידרש, פנה לעזרתו של הבלשן יעקב אורבך ז"ל (נפטר בשנה שעברה). הוא הפנה אותו ל"אנציקלופדיה לתולדות גדולי ישראל" של ד"ר מרדכי מרגליות. דורון בחר את ה"גדולים ביותר" לפי הקריטריון הבא: ברשימה נכללו אלה, שערכם בן שלושה טורים ומעלה, ואשר תאריך הלידה ו/או פטירה נכלל בערכם. היו אלה 34 האישים שברשימה הראשונה.

מן המפורסמות, שקיימות שגיאות באנציקלופדיות, ולכן נתן דורון את רשימת האישים לבדוק ובודקת על מנת שישוו את התאריכים המופיעים באנציקלופדיה לכמה מקורות נוספים. עליהם היה לציין אם יש מידע נוסף לגבי התאריכים ומהו. לאחר שבועות מספר, כשקיבל דורון את שני הדו"חות (של הבדוק

ושל הבודקת), התברר לו כי הם מצאו :

(i) שישנם שני תאריכי פטירה השנויים במחלוקת מפורסמת : של הרי"ף ושל הבעש"ט.

(i i) שישנו תאריך אחד שגוי : של רבי דוד גאנו.

(i i i) שישנו תאריך נוסף : תאריך לידתו של הבעש"ט.

אחד הבודקים, הפנה את תשומת לבו של דורון לכך, שמן המקורות שבדק התרשם, שישנם שמות או כינויים נוספים או אחרים לכמה מן האישים.

ו. עם כל המידע הזה, חזר דורון ליעקב אורבך, וביקש יעוצו בנוגע לתאריכים ובנוגע לשמות וכינויים.

יעקב אורבך הציע עזרתו כבלשן : בכתיבה מדוייקת של התאריכים ושל השמות והכינויים.

הכללים שהינחו אותו מופיעים במאמר. באשר לשאלה מהם השמות או הכינויים של האישים הנ"ל, יעץ לפנות לפרופ' שלמה זלמן הבלין, הידוע כמומחה בתחום הביבליוגרפי הזה. דורון פנה לפרופ' הבלין, נתן לו את רשימת 34 האישים, וביקש שיכין רשימה של השמות והכינויים שלדעתו מאפיינים את האישים הללו.

ז. רשימת השמות והכינויים הוכנה ע"י פרופ' הבלין בחורף התשמ"ו. ואז, נערך המבחן לרשימה הראשונה. בעקבות הצלחת המבחן, החלטנו לכתוב מאמר. המאמר נדפס כפרה-פרינט ביום ג' תשרי התשמ"ז ("החוברת הלבנה").

ח. המאמר נשלח לפרופ' קשדן. הוא העביר אותו לפרופ' דיאקוניס. במכתב תשובה לפרופ' קשדן מיום 30 בדצמ. 86 (לפי מניינם), שהועבר מאוחר יותר אלינו, הציע פרופ' דיאקוניס להכין רשימה חדשה של אישים מפורסמים, ולבדוק אותה בדיוק באותה התוכנה.

ט. דורון הכין רשימת אישים חדשה לפי אותה אנציקלופדיה, הפעם היו בה האישים שערכם

באנציקלופדיה 3 - 1.5 טורים. נמצאו על ידו 32 אישים כאלה. כיוון שהנסיון הראה שברשימה

הראשונה היו תאריכים מסופקים ואפילו מוטעים, חשש דורון שלגבי אישים פחות מפורסמים, אישי הרשימה החדשה, צפויות טעויות רבות יותר. לכן החליט לבדוק את כל התאריכים, בהשתמשו בחומר המצוי בבית הספרים הלאומי. ואכן, נמצאו כמה טעויות וספקות, עליהם דווח בפרה-פרינט השני. את רשימת האישים העביר לפרופ' הבלין על מנת שיכין רשימה שניה של שמות וכינויים.

י. כאשר היו כל הנתונים בידינו, נערך המבחן לרשימה השנייה. ואז נערך מבחן נוסף שבו לכל אישיות

ברשימה השנייה הוצמד תאריך של האישיות הבאה ברשימה. גם מבחן זה נערך לפי הצעתו של פרופ'

דיאקוניס. המדידות הסתיימו באייר התשמ"ז, ונשלחו ב-19 במאי 87 (למניינם) לפרופ' דיאקוניס.

בעקבות מבחנים אלה נכתב המאמר בגירסא חדשה כפרה-פרינט השני ("חוברת כחולה") בחורף התשמ"ח.

יא. רצינו לשלוח את המאמר לפרסום ב-PNAS, וכיוון שהגשת מאמר לפרסום שם צריכה להעשות ע"י חבר האקדמיה הלאומית האמריקנית למדעים, פנינו לפרופ' ישראל אומן. עד מהרה התברר לנו שהעתון מגביל את אורך המאמרים שהוא מפרסם. כדי לעמוד בדרישות העתון, נאלצנו לקצץ בחלק מן המידע הכלול במאמר (כולל טבלאות השמות והכינויים). הכנו גירסא מקוצרת של המאמר. פרה-פרינט זה מצורף למכתבנו.

יב. לגבי פרה-פרינט זה התנהל משא ומתן בנוגע לפירסומו כמאמר. במשא ומתן זה הציע פרופ' דיאקוניס לראשונה במכתב מיום 3 לאוג. 88 (למניינם) מבחן המבוסס על כמות גדולה של תמורות אקראיות. בסופו של דבר, נקבעו פרטי המבחן, מספר התמורות וסף המובהקות, ע"י פרופ' דיאקוניס במכתב מיום 5 בספטמ. 90 (למניינם). מנגנון יצירת התמורות ע"י האלגוריתם P של קנות, הוצע על ידו עוד קודם לכן. כל הפרטים שהציע הם אלה המתוארים במאמר שפורסם. הוא הציע לערוך מבחן זה לרשימה השניה.

יג. לפי הצעתו של פרופ' אומן נכתב המאמר מחדש, לפני שבוצע המבחן, למעט תוצאות הניסוי שטרם בוצע. במתכונתו זו נשלח המאמר למבקרים נוספים. אחרי שהם אישרו את המבחן, וקבעו (כל אחד מהם) את סף המובהקות הנדרש, בוצע הניסוי החורף התשנ"ב. תוצאותיו הוכנסו למאמר. גם המאמר הזה יועד ל-PNAS, אך פורסם בסופו של דבר, כפי שהוא, ב-STATISTICAL SCIENCE.

2. א. את הניסויים ערכנו על ספר בראשית, לכן איננו מבינים את הבקשה לקבל את הקלדת שאר החומשים "כדי לערוך עליהם בדיקות משלנו, תוך הבטחה שאנחנו ואתם משתמשים באותו טקסט בדיוק". מכל מקום נוכל להעביר לכם את הקלדת ספר בראשית השייכת לנו. קובץ של כל התורה, שקיבלנו חנים לפני כמה שנים ממשמרת סת"ם בבני ברק, התחייבנו שלא להעבירו לאחרים. אם אתם מעוניינים דווקא בקובץ זה, תוכלו לפנות ישירות אליהם.

ב. נשאלנו מה המובהקות של המידגם הראשון לפי חשובינו.

התוצאות שחושבו לרשימה הראשונה הן:

$$P2=0.00000000145 \quad P1=0.000000001334$$

במבחן רנדומיזציה (הפרמוטציות) שנערך בדיוק כפי שנעשה לגבי הרשימה השניה, היה הדירוג של P2

22 מתוך 1,000,000. זו היתה התוצאה הטובה יותר.

3. איננו מומחים לביבליוגרפיה. בטוחים אנו שאיש מקצוע בתחום זה יתן לכם יעוץ הרבה יותר טוב.

4. א. נשאלנו מדוע ברישום התאריכים השתמשנו בצורות "א תשרי" "בא תשרי" ו"א בתשרי" ולא בצורה "בא בתשרי".

אכן, ברישום התאריכים לא השתמשנו בצורה "בא בתשרי". שמנו לב לכך בעקבות ההערות שלכם ושל אדם נוסף. לצערנו, איננו יכולים לשאול כעת את הבלשן יעקב אורבך ז"ל מדוע לא כלל בהצעתו גם את הצורה הזאת. אך כדי להסיר כל ספק, אנו מצהירים כאן, כי קביעת רישום התאריכים נעשתה ע"י יעקב אורבך, בדיוק כפי שהם כתובים בפרה-פרינט הראשון, עוד לפני ביצוע המבחן לרשימה הראשונה. אנו השתמשנו בתאריכים בדיוק באותה הצורה גם לגבי הרשימה השניה.

אם מישהו חושד, שהיתה כוונה נסתרת בהשמטת הצורה "בא בתשרי", על מנת לשפר את התוצאות, הרי הוא מוזמן לערוך בדיקה, כפי שעשינו לאחר ששמענו טענה (בעל פה) בעניין זה מד"ר דרור בר-נתן. בדקנו, מה היה קורה אם היינו משתמשים גם בצורה "בא בתשרי".

כזכור, אמת המידה היחידה שלנו להצלחה, גם ברשימה הראשונה וגם בשניה, כפי שפורסם בפרה-פרינט הראשון ("חוברת לבנה") וגם בפרה-פרינט השני ("חוברת כחולה"), וגם בפרה-פרינט ששלחנו בתחילה ל-PNAS, היתה הסתברויות הכוללות שחישבנו: P_1 ו- P_2 .

מבחן הרנדומיזציה של פרופ' דיאקוניס הוצע בשלב מאוחר יותר.

(i) התוצאות שחושבו למידגם הראשון הן: $P_1=0.000000001334$ $P_2=0.00000000145$

אם היינו משתמשים בצורה "בא' בתשרי" היינו מקבלים

$P_1'=0.000000000349$ $P_2'=0.000000000207$

כלומר, התוצאה הטובה ביותר השתפרה פי 3.8.

(i i) התוצאות שחושבו למידגם השני הן: $P_1=0.00000000331$ $P_2=0.00000000201$

אם היינו משתמשים בצורה "בא' בתשרי" היינו מקבלים

$P_1'=0.000000000507$ $P_2'=0.000000000171$

כלומר, התוצאה הטובה ביותר השתפרה פי 1.18.

כלומר, מוכח מכאן, שצורות התאריך נקבעו באופן א-פריורי.

ב. נשאלנו מדוע לא השתמשנו בעוד צורות רישום לתאריכים (כגון: מרחשון, אייר, כסליו, ר"ח אדר,

ראש השנה, יא לתשרי).

ננסה לענות במקום מי שקבע את צורת התאריכים ולצערנו אינו יכול להשיב.

(i) "מרחשון" - איננו יודעים מדוע לא השתמש בו, בנוסף ל"חשון". על כל פנים בדקנו עתה ומצאנו,

שאם היינו משתמשים גם ב"מרחשון" התוצאות כלל לא היו משתנות.

(i i) "אייר", "כסליו" - לפי הכללים הלשוניים שנקבעו (ראה לעיל 1 ו'), הכתיב הוא "אייר", "כסליו" וכך עשינו.

(i i i) "יא לתשרי" - אין זו הצורה המקובלת. למשל, הן אנציקלופדיה מרגליות והן האנציקלופדיה

העברית השתמשו בתצורות שאנו השתמשנו, ולא בתצורה זו. אין ספק שאלו התצורות הנפוצות

בכלל. ערכנו בדיקה, באמצעות מאגר השו"ת הממוחשב של אוניברסיטת בר-אילן. והרי תוצאות

הבדיקה לגבי השמוש בתצורות היסודיות הבאות, במאגר פוסקי זמננו:

נסמן את התצורה "א תשרי" + "בא תשרי" ב-1,

את התצורה "א בתשרי" + "בא בתשרי" ב-2,

את התצורה "א לתשרי" + "בא לתשרי" ב-3.

הטבלה הבאה מסכמת את שכיחויות השימוש ב-1, ב-2 וב-3.

| החודש | 1 | 2 | 3 |
|-------|-----|-----|---|
| תשרי | 178 | 51 | 2 |
| חשון | 364 | 130 | 1 |
| כסלו | 409 | 90 | 0 |
| טבת | 375 | 108 | 0 |
| שבט | 434 | 190 | 4 |
| אדר | 582 | 159 | 6 |
| ניסן | 303 | 126 | 0 |
| אייר | 359 | 82 | 0 |
| סיון | 319 | 86 | 0 |
| תמוז | 419 | 181 | 2 |
| אב | 68 | 263 | 0 |
| אלול | 286 | 86 | 0 |

(i v) לגבי "יר"ח אדר", "ראש השנה" - איננו מבינים כלל מה הפסול בכך שהשתמשנו בצורה

סטנדרטית ומקובלת, שבה נמצאים רוב התאריכים באנציקלופדיה, ולא בצורות שונות וחריגות. סביר

להניח, שאם היינו משתמשים בכל מיני צורות, היו טוענים מדוע לא השתמשנו בתצורה סטנדרטית...

5. א. נשאלנו מדוע מופיע אצלנו התאריך י"ט בתשרי כתאריך פטירתו של הגאון מוילנא, בעוד שידוע

שהוא נפטר בג' דחול המועד סוכות.

התשובה לכך פשוטה: בחוץ לארץ יש יום טוב שני של גלויות. לכן ג' דחול המועד בוילנא הוא י"ט תשרי.

ב. נשאלנו היכן נמצא תאריך הלידה של הבעש"ט. תאריך זה נמצא למשל, בלקסיקון "החסידות", של יצחק אלפסי, בהוצאת ספריית מעריב.

ג. נשאלנו מדוע נכללו ר' אהרן מקרלין, ר' יהודה עיאש, ור' יהוסף הנגיד ברשימה השניה למרות שערכם באנציקלופדיה מרגליות, לפי דעת השואלים, הוא פחות מטור וחצי. כמו כן נשאלנו מדוע לא נכלל ר' מאיר אייזנשטט ברשימה השניה למרות שערכו לדעתם הוא טור וחצי בדיוק. נשאלנו גם מדוע נכלל ר' דוד גאנו ברשימה הראשונה ולא בשניה, למרות שלדעתם ערכו הוא מעט פחות מ-3 טורים. התשובה לכך פשוטה: השואלים מדדו את גודל הערך לפי שורות. דורון, שבחר את 34 האישים לרשימה הראשונה, עשה זאת (עד כמה שהוא מסוגל לשחזר) בטביעת עין. וכן עשה אחרי כן בבחירת 32 האישים שברשימה השניה. מתברר, שהיתה זאת טעות לסמוך על טביעת העין, ושהמודד של השואלים טוב יותר. אבל אם מישהו חושד, שהיה כאן נסיון ליצור הצלחה יש מאין באמצעות מניפולציה, יעיין בבקשה בתוצאות הבדיקה הבאה, שערכנו עם היוודע לנו טענה זאת.

(i) חישבנו את המידגם הראשון גם ללא ר' דוד גאנו. הבה נשווה:

התוצאות שחושבו למידגם הראשון הן: $P_1=0.00000001334$ $P_2=0.00000000145$

אם מחשבים אותו ללא ר' דוד גאנו מקבלים: $P'_1=0.00000001336$ $P'_2=0.00000000276$

כלומר, התוצאה הטובה ביותר נעשתה גרועה פי 2.07.

(ii) חישבנו את המידגם השני בהשמטת ר' אהרן מקרלין, ר' יהודה עיאש ור' יהוסף הנגיד, ובהוספת

ר' דוד גאנו ור' מאיר אייזנשטט (השמות והכינויים השייכים לו נמסרו לנו בי"ב טבת התשנ"ז ע"י

פרופ' הבלין, והם: רבי מאיר, איזנשטט, איזנשטאט, מאיר איזנשטט, מאיר איזנשטאט, מהר"ם

א"ש, "פנים מאירות", בעל "פנים מאירות"). לשם השוואה:

התוצאות שחושבו למידגם השני הן: $P_1=0.0000000331$ $P_2=0.00000000201$

אם מחשבים אותו עם השנויים הנ"ל מקבלים: $P'_1=0.00000000422$ $P'_2=0.00000000129$

כלומר, התוצאה הטובה ביותר השתפרה פי 1.56.

בסך-הכל, רואים בבירור, שמתקבלות תוצאות מאותו סדר גודל.

בכבוד רב

דורון ויצטום
אליהו ריפס

אפילוג:

בחורף התשנ"ז התקיים כנס של המרכז לחקר לרציונליות (של האוניברסיטה העברית) בזרקה שבירדן. באותו כנס העלתה פרופ' בר הלל השגות על עבודת המחקר שלנו. המתמטיקאי פרופ' ישראל אומן השתתף באותו כנס. לפני שיצא לזרקה, הוא ביקש מדורון ויצטום להריץ מחדש את ניסוי הפרמוטציות על הרשימה השנייה, הפעם לפי התיקון שהוצע ע"י בר הלל בסעיף האחרון דלעיל, בזמן הקצר שעמד לרשותו, הספיק דורון להריץ 10,000,000 פרמוטציות. לא רק שהתוצאה לא הורעה, אלא שהסטטיסטיקה P4 השתפרה, ובמקום ערך של 4 ל- 1,000,000 קיבלה ערך של 1 ל--10,000,000! (מאוחר יותר, הספיק דורון להריץ 100,000,000 פרמוטציות, והתוצאה היתה 19 ל- 100,000,000).