

רציונליות

מ-א ועד ת



ספר I

מפה ושטח

אליעזר יודקובסקי

רציונליות: מ-א ועד ת | אליעזר יודקובסקי

רציונליות

מ-א ועד ת

מאנגלית: אבנר בן זמרה

עריכה: דביר שדה

ספר I

מפה ושטח

אליעזר יודקובסקי

Eliezer Yudkowsky
Rationality: From AI to Zombies

אליעזר יודקובסקי
רציונליות: מ-א' ועד ת'
(או, רציונליות: מאינטליגנציה מלאכותית ועד תחיית המתים)

תרגום: אבנר בן זמרה
עריכה והבאה לדפוס: דביר שדה

זהו מיזם ללא מטרות רווח, האנשים הנ"ל עבדו על הספר בהתנדבות. הם אינם אחראים לעיצוב של הספר, להדפסתו, לאופן הפצתו אל לכל שימוש שיעשה בו, והם מבקשים לעשות בספר זה שימוש חוקי בלבד.

יצא לאור באנגלית בשנת 2015 ע"י
Machine Intelligence Research Institute (MIRI)
תחת רישיון CC BY-NC-SA 3.0
הספר בעברית זמין בחינם בכתובת: <https://rationality.co.il>
הספר באנגלית זמין בחינם בכתובת: <https://www.readthesequences.com>

אין למכור ספר זה
אין לעשות בספר זה כל שימוש מסחרי

הדפסה ראשונה של כרך זה
נדפס בישראל 2026

תוכן עניינים

9	בפתח המהדורה העברית
12	הקדמה
16	הטיות: מבוא

שגוי באופן צפוי

34	1. למה אני מתכוון כשאני אומר 'רציונליות'?
40	2. להרגיש רציונלי
43	3. למה לי לדעת את האמת? ו...
47	4. ...תזכיר לי מה זו הטיה?
51	5. זמינות
54	6. פרטים מכבידים
59	7. כשל התכנון
64	8. אשליית השקיפות: למה אף אחד לא מבין אותך
68	9. ציפייה למרחקי היסק קצרים
71	10. העדשה שרואה את פגמיה שלה

אמונות מזויפות

76	11. גם אמונות צריכות לשלם שכירות (בציפיות לעתיד)
80	12. משל על מדע ופוליטיקה
85	13. אמונה באמונה
90	14. ג'ודו בייסיאני

- 92 .15 להעמיד פני חכם
- 97 .16 תביעת הדת להיות בלתי ניתנת להפרכה
- 102 .17 הצהרות וקריאות עידוד
- 105 .18 אמונה כסממן קבוצתי
- 107 .19 נורת מחיאות הכפיים

לשים לב לבלבול

- 112 .20 למקד את אי-הוודאות
- 116 .21 מהי ראייה?
- 120 .22 ראייה מדעית, ראייה משפטית, ראייה רציונלית
- 124 .23 כמה ראיות נדרשות?
- 128 .24 יהירותו של איינשטיין
- 131 .25 תערו של אוקאם
- 136 .26 כוח העל שלך כרציונליסט
- 139 .27 היעדר ראיות הוא ראייה להיעדר
- 142 .28 חוק שימור הראיות הצפויות
- 145 .29 החוכמה בדיעבד מפחיתה מערכו של המדע

תשובות מסתוריות

- 150 .30 למקד את אי-הוודאות
- 153 .31 לנחש את מילת הקוד של המורה
- 157 .32 מדע כסממן קבוצתי
- 160 .33 סיבתיות מדומה
- 165 .34 תמרורי עצור סמנטיים

169	35. תשובות מסתוריות לשאלות מסתוריות
173	36. חוסר התוחלת של מערכות מתהוות
176	37. אל תגיד "סיבוכיות"
180	38. הטיה חיובית: להביט אל תוך האפלה
184	39. אי-ודאות שיטתית
188	40. נעוריי הפרועים והפוחזים
192	41. חוסר היכולת ללמוד מההסטוריה
194	42. להנגיש את ההיסטוריה
198	43. להסביר/לסגוד/להתעלם?
200	44. "מדע" כעוצר סקרנות
204	45. חלק ממך באמת
209	אינטרלוד: האמת הפשוטה

בפתח המהדורה העברית

בהתרגשות גדולה אנו מגישים לקורא העברי את הכרך הראשון של רציונליות: מ-א' ועד ת' – חיבור מכוון שהפך ל"ספר הלימוד" הבלתי־רשמי של שוחרי רציונליות ברחבי העולם.

המחבר, אליעזר יודקובסקי, מוכר כיום כאחד ההוגים המשפיעים בתחום הבטיחות של בינה מלאכותית. אולם, מפעל־חיים משמעותי לא פחות של יודקובסקי הוא פיתוחה והנגשתה של הרציונליות כמשנה סדורה. לאורך השנים כתב יודקובסקי רבות על אודות החשיבה האנושית, ובין היתר הקים את קהילת LessWrong – במה עולמית לחקר הרציונליות שהפכה למוקד של חשיבה מקורית המשלבת בין תורת ההחלטות, פסיכולוגיה קוגניטיבית, תורת ההסתברות ופילוסופיה של המדע. דא עקא, כתיבתו הפורה של יודקובסקי והשיח הקהילתי סביבה מתנהלים בשפה האנגלית, מה שמקשה על הציבור דובר העברית להיחשף לתוכן זה.

אבן־דרך מרכזית בחשיפת הקהל הישראלי להגותו של יודקובסקי התרחשה בשנת 2020, עם תרגומו והדפסתו של הפאנפיק הארי פוטר והשיטה הרציונלית על ידי חברים מקהילת "רציונליות ישראל".* בספר זה מונגשים רבים מעקרונות הרציונליות תוך שילובם בסיפור אלטרנטיבי על הארי פוטר כילד־פלא רציונלי. רבים מצאו בסיפור

* ששת כרכי הספר זמינים לקריאה ולהורדה באתר הקהילה - rationality.co.il, ולהאזנה בפורדקאסט הארי פוטר והשיטה הרציונלית.

צוהר לעולם חדש של חשיבה, אך מי שירצה להעמיק בשיטות הרציונליות יתקשה לעשות זאת מתוך הסיפור, בו מובאים הרעיונות הרציונליים אגב־אורחא, כחלק מהעלילה.

הספר שלפניכם הוא המקור. פה שוטח יודקובסקי באופן סדור את משנתו על אודות הרציונליות, אותה פיתח במשך עשרות שנים של הגות ומחקר. הטקסט, שפורסם במקור כקטעים קצרים העומדים בפני עצמם ונערך מאוחר יותר לכדי ספר, מכנס 333 קטעים קצרים, המאוגדים ב-26 "רציפים", בששה כרכים. כל קטע בוחן זווית אחרת של המחשבה האנושית – החל מהטיות קוגניטיביות, דרך מהותה של השיטה המדעית, וכלה בשאלות הקיומיות של המוסר והמשמעות.

על המלאכה והתרגום

השקענו עבודה רבה בהוצאה לאור של כרך זה בעברית, בהתנדבות מלאה, מתוך מטרה להנגיש רציונליות לקורא הישראלי. בעולם של זרם מידע בלתי פוסק ושיח ציבורי אקטואלי שלא נוטים להתמקד בעיקר, בעולם בו יש כל כך הרבה להבין, ולדעת, ולהחליט, ולבחור – אני מאמין שהתובנות השוכנות בין דפי הספר הזה עשויות להיות לעזר רב, ולכלי מחזיק ברכה, למי שישכיל לאמץ אותן ולהשתמש בהן. אני מאחל לכם שתפיקו מהספר הנאה ותועלת מרובה.

עבודת התרגום נעשתה ביסודיות ובכשרון רב על ידי אבנר בן זמרה. כתיבתו של יודקובסקי אמנם נגישה לרוב, אך לעתים מתחכמת ונוטה למבנים טכניים ומורכבים. היכולת לתרגם את הדברים בשפה קולחת ובהירה מבלי לאבד את הנימה המקורית ותוך שמירה על דיוק ועל כוונת המחבר, היא אמנות של ממש, ועל כך נתונה לו הערכתי העמוקה. יחד איתו עסקתי במלאכת העריכה וההגהה, אליה הצטרפו חברים נוספים מקהילת רציונליות ישראל (ומחוצה לה). תודה מיוחדת נתונה לדני בראון על עזרתו בעימוד הספר.

ביטויים ומושגים הלקוחים מעולם המושגים האמריקאי הוחלפו

לעתים במושגים מקבילים שיהיו קרובים יותר לעולמו של הקורא העברי (לדוגמה, "זה יהיה כמו להתחפש... בפורים" במקום "בליל כל הקדושים", בפרק 18).

הערות שהוספנו בגוף הספר הובאו בתחתית העמוד וסומנו בכוכביות (*). הערות השוליים המקוריות של המחבר ממוספרות, ומקומן בסוף כל פרק.

לא נמנענו מלהוסיף בהערות ציטוטים מתוך הארי פוטר והשיטה הרציונלית במקומות שמצאנו מתאימים לכך. קוראים אשר טרם קראו את הספר, יכולים להתעלם מהערות אלו באלגנטיות.

— דביר שדה, אייר תשפ"ו

הקדמה

הספר בו אתם אווזים מרכז שנתיים של פוסטים יומיים בבלוג. בדיעבד, אני מביט לאחור על הפרויקט ומזהה מספר רב של דברים שעשיתי באופן שגוי לחלוטין. זה בסדר מבחינתי. להביט לאחור ולא לראות מספר עצום של דברים שעשיתי לא נכון, פירושו שמאז 2009 לא חל שום שיפור בכתיבה שלי, או בהבנה שלי את הדברים. אופס הוא הצליל שאנו משמיעים כשאנחנו מעדכנים את האמונות ואת האסטרטגיות שלנו. להביט לאחור על תקופה ולא למצוא אף שגיאה, פירושו שלא למדת דבר או שלא שינית את דעתך מאז.

היתה זו טעות מצדי לכתוב פוסטים בבלוג במשך שנתיים שלא מתוך כוונה לעזור לאנשים להצליח יותר בחיי היומיום שלהם. כתבתי אותם מתוך כוונה לעזור לאנשים לפתור בעיות גדולות, קשות וחשובות, ובחרתי בבעיות מופשטות שנשמעות מרשימות בתור דוגמאות.

בדיעבד, זו הייתה הטעות השנייה הכי גדולה שלי. היא מתקשרת לטעות הראשונה הכי גדולה בכתיבה שלי, והיא שלא הבנתי שהבעיה העיקרית בלימוד צורת החשיבה הזו היא להבין איך לתרגל אותה, ולא איך להבין את התיאוריה שלה. לא הבנתי שהחלק הזה הוא בראש סדר העדיפויות. לגבי זה אני יכול רק לומר "אופס" ו"כאילו דה". אמת, לעתים הסוגיות הגדולות הללו אכן גדולות ואכן חשובות, אך זה לא משנה את האמת הבסיסית האומרת שכדי לשלוט במיומנויות ררוש אימון, וקשה יותר להתאמן על דברים שרחוקים מחיי היומיום.

(כיום "המרכז לרציונליות יישומית" עובד על תיקון הטעות העצומה הזו באופן יותר שיטתי.)

טעות ענקית שלישית שעשיתי היתה להתמקד יותר מדי באמונה רציונלית, ומעט מדי בפעולה רציונלית.

הטעות הרביעית הכי גדולה שעשיתי היא שהייתי צריך לארגן טוב יותר את התוכן שהצגתי ב"רצפים". בפרט, הייתי צריך ליצור אתר וויקי מוקדם הרבה יותר, ולהפוך בכך את קריאת הפוסטים ברצף לקלה יותר.

הטעות הזו לפחות ניתנת לתיקון. בקובץ הנוכחי רוב בנסינגר סידר מחדש את הפוסטים וארגן אותם ככל יכולתו מבלי לנסות לשכתב את כל החומר הממשי (אם כי הוא בכל זאת שכתב חלק קטן).

הטעות הענקית החמישית שלי הייתה הניסיון לדבר באופן ישיר (בעיני לפחות) על הטיפשות של מה שנראה לי כרעיונות טיפשיים. כן ניסיתי להימנע מהכשל הלוגי הידוע בשם "בולֶּרִיזם", בו אתה פותח את הדיון בלהצביע על כמה אנשים הם טיפשים בגלל שהם מאמינים במשהו. תמיד דנתי קודם כל בסוגיה, ורק לאחר מכן אמרתי, "ולכן זה טיפשי." אבל ב-2009 השאלה לגבי חשיבות נוכחותם של כמה אנשים בסביבה שיביעו בוז להומאופתיה היתה בעיניי שאלה פתוחה. חשבתי אז, ואני עדיין חושב, שישנה בעיה מצערת, במסגרתה התייחסות סובלנית לרעיונות מעובדת על ידי אנשים רבים ברמה מסוימת כ"שום דבר רע לא יקרה לי אם אגיד שאני מאמין בזה. מעמדי לא ייפגע אם אגיד שאני מאמין בהומאופתיה, "ושצחוק לענגי מצד קומיקאים יכול לעזור לאנשים להתעורר מהאשליה.

אני חושב שכיום הייתי כותב בצורה סובלנית יותר. חוסר הסובלנות אכן שירת תפקיד מסויים, ואני חושב שהיו אנשים שיצאו נשכרים ממנו, אבל אני לוקח היום יותר ברצינות את הסיכון שביצירת קהילות בהן התגובה הנורמלית והאוטומטית להשקפות שונות שנחשבות "פחותות" היא לעג גלוי ובוז.

למרות הטעות שלי, אני שמח לומר שעד כה ציבור הקוראים שלי היה טוב להדהים באי־השימוש ברטוריקה שלי כתירוץ לברינות

ולהקטנה של אחרים. (בהקשר הזה אני רוצה לציין ספציפית את סקוט אלכסנדר, שהוא אדם נחמד יותר ממני וכותב מדהים על הנושאים הללו, וכנראה ראוי לחלק מהקרדיט על יצירת תרבות בריאה ב-Less Wrong).

היכולת להביט לאחור ולומר ש"נכשלת" מרמזת על כך שהיו לך מטרות. אז מה ניסיתי לעשות בעצם?

ישנה דרך חשיבה מסוימת שהיא בעלת ערך, שטרם נלמדת בבתי ספר, נכון להיום. דרך החשיבה המסוימת הזו כלל לא נלמדת באופן שיטתי. היא פשוט נקלטת על ידי אנשים שגדלו על ספרים כמו אתה כטח מתלוצץ, מיסטר פינמן! או שיש להם מורה נהדר באופן חריג בתיכון.

באופן הגלוי ביותר, לדרך החשיבה המסוימת הזו יש קשר למדע, ולשיטה הניסויית. החלק במדע שבו אתה יוצא ומסתכל על היקום ולא סתם ממציא דברים. החלק שבו אתה אומר "אופס" ומוותר על תיאוריה גרועה כשהניסויים לא תומכים בה.

אבל דרך החשיבה המסוימת הזו משתרעת הרחק מעבר לכך. היא עמוקה ואוניברסלית יותר מזוג משקפיים שאתה מרכיב כשאתה נכנס למעבדה ומוריד כשאתה עוזב. היא חלה על חיי היומיום, אם כי באופן עדין ומורכב יותר. אם אתה לא יכול להגיד "אופס" ולוותר כשנראה שמשוהו לא עובד, אין לך ברירה אלא להמשיך ולירות לעצמך ברגל. אתה חייב לשוב ולטעון את הרובה ואתה חייב לשוב ולסחוט את ההדק. אתם מכירים אנשים כאלה. ואיפשהו, במקום כלשהו בחייכם שאתם מעדיפים שלא לחשוב עליו, אתם אנשים כאלה. יכול היה להיות נחמד אם הייתה דרך חשיבה מסוימת שתעזור לנו להפסיק לעשות את זה.

למרות גודלן של הטעויות שלי, נראה שאותן שנתיים של כתיבה בבלוג עזרו במידה מפתיעה למספר מפתיע של אנשים. זה לא עבד בצורה עקבית, אבל זה עבד לפעמים. כה מעט נלמד בחברה המודרנית ממיומנויות האמונה הרציונלית וקבלת ההחלטות הרציונליות, כה מעט מהמתמטיקה והמדעים שבבסיסן... עד שמסתבר שעצם

הקריאה של אסופה גולמית ועצומה של מחשבות הגדושה בבעיות בפילוסופיה ובמדע, יכולה – כן כן – להועיל באופן מפתיע. התעיה בתוך כל זה, מתריסר זוויות שונות, עשויה אולי להעניק הצצה לקצב הכללי של הדברים.

כי בסופו של יום, מדובר באותו הדבר. כתבתי רבות על בעיות גדולות ומרוחקות והזנחתי את החיים היום-יומיים, אבל החוקים השולטים בהם אינם שונים, בסופו של דבר. יש פערים עצומים בחלקים בהם התמקדתי, ובחרתי את כל הדוגמאות הלא נכונות, אבל בסופו של יום מדובר באותו הדבר. אני גאה להביט לאחור ולומר, שאפילו אחרי כל השגיאות שביצעתי, וכל שאר הפעמים בהן אמרתי "אופס"...

אפילו חמש שנים לאחר מכן, עדיין נראה לי שזה טוב יותר מכלום.

– אליעזר יודקובסקי, פברואר 2015

הטיות: מבוא

מאת רוב בנסינגר

זה לא סוד. אולם מסיבה כלשהי, הנושא כמעט שלא עולה בשיחות, ורק מעט אנשים שואלים מה יש לעשות לגביו. זהו דפוס נסתר, הנח כסמוי מן העין מאחורי כל הצלחותינו וכשלונותינו, סמוי מאחורי עינינו. מה זה?

דמיינו שאתם מכניסים את ידכם לתוך כד המכיל שבעים כדורים לבנים ושלושים כדורים אדומים, ושולפים עשרה כדורים מסתורין. אולי שלושה מתוך עשרת הכדורים יהיו אדומים, ואתם תנחשו נכונה כמה כדורים אדומים יש בכד בסך הכל. או שאולי תשלפו ארבעה כדורים אדומים ושישה לבנים, או צירוף אחר. במקרה כזה סביר להניח שתטעו בניחוש המספר הכולל של כדורים אדומים בכד.

הטעות האקראית הזו היא המחיר של ידע חלקי, ויחסית לטעויות אחרות, היא לא כל כך גרועה. ההערכות שלכם בנוגע למספר הכדורים האדומים בכד לא יהיו שגויות בממוצע, וככל שתלמדו יותר, כך הטעות שלכם תקטן.

מצד שני, אם נניח שהכדורים הלבנים כבדים יותר, ונוטים להיאסף בתחתית הכד, אז המדגם שלכם עלול להיות לא מייצג באופן עקבי. סוג כזה של טעות נקרא "הטיה סטטיסטית". כשהשיטה בה אתה משתמש כדי ללמוד על העולם מוטית מלכתחילה, למידה נוספת ככל הנראה לא תעזור. איסוף של נתונים נוספים עשוי אף להחמיר את התחזית המוטת.

אם אתם מאמינים שמחקר ואיסוף ידע הם בעלי ערך, אז זוהי אפשרות מפחידה. אם ברצוננו לוודא שלמידה על העולם אכן תועיל לנו, ולא תחמיר את מצבנו, עלינו לאתר הטיות בנתונים שלנו ולתקן אותן.

הרעיון של הטיות קוגניטיביות בפסיכולוגיה עובד באותו האופן. הטיה קוגניטיבית היא טעות מערכתית בצורה בה אנו חושבים, לעומת

טעות אקראית, או טעות הנובעת מבורות פשוטה. בעוד שהטיה סטטיסטית מעוותת את המדגם כך שייצג בצורה פחות אמינה אוכלוסיה גדולה יותר, הטיה קוגניטיבית מעוותת את האמונות שלנו כך שייצגו בצורה פחות אמינה את העובדות, ומעוותת את קבלת ההחלטות שלנו כך שיקטינו את סיכויינו להשיג את המטרות שלנו. אולי אתם לוקים בהטיית האופטימיות, וגיליתם שהכדורים האדומים יכולים לרפא מחלה טרופית נדירה בה לוקה אחיכם. עקב כך, אולי תעריכו ביתר את מספר הכדורים האדומים בכד מכיוון שאתם מקווים שרוב הכדורים בכד אדומים. במקרה הזה, המדגם אינו מוטא. אתם אלה שמוטים.

מכל מקום, עלינו להיזהר כשאנו מדברים על אנשים מוטים. בדרך כלל כשאנו אומרים על אנשים או קבוצות שהם "מוטים", אנו עושים זאת כנזיפה על כך שהם לא הוגנים או משוחדים. הטיה קוגניטיבית היא חיה אחרת לגמרי. הטיות קוגניטיביות הן חלק יסודי מהצורה בה בני אדם חושבים, לא מין סוג של פגם שצומח מתוך ילדות עשוקה או אישיות נוראית.¹

הטיה קוגניטיבית היא צורה שיטתית בה דפוסי החשיבה המולדים שלך נכשלים בהשגת האמת (או מטרה ברת-השגה אחרת, כמו אושר). כמו הטיות סטטיסטיות, גם הטיות קוגניטיביות עשויות לעוות את תפיסת המציאות שלנו, גם אותן לא תמיד ניתן לתקן על ידי איסוף של מידע נוסף, וגם להן יש אפקט מצטבר. אך כשמכשיר המדידה אותו אתה מנסה לכייל הוא אתה, תיקון-ההטיות הוא אתגר אמיתי.

ובכל זאת, זה מקום טוב להתחיל בו, שהרי אם לא תוכל לבטוח במוחך שלך, במה תוכל לבטוח?

יהיה זה מועיל לתת שם לפרוייקט ההתגברות על ההטיות הקוגניטיביות, ולהתגברות על כל המקרים בהן מוחנו שלנו משטה בנו.

אנו יכולים לכנות את הפרוייקט הזה במגוון שמות, אך לעת עתה אני מניח ש"רציונליות" נשמע טוב כמו כל אפשרות אחרת.

רגשות רציונליים

בסרטים אמריקאיים, להיות "רציונלי" בדרך כלל אומר שאתה מין טיפוס נוקשה, סטואי והיפראינטלקטואל. חשבו על ספוק ממסע בין כוכבים, שמדכא את רגשותיו "באופן רציונלי", מסרב להסתמך על אינטואיציה או דחפים פנימיים "מסיבות רציונליות" ושבסוף מתומרן ומופתע ע"י יריבים בלתי צפויים או "לא רציונליים".²

ישנו מושג שונה לחלוטין של "רציונליות" אליו מתייחסים מתמטיקאים, פסיכולוגים וחוקרי מדעי החברה. בגדול, זה הרעיון של לעשות את הטוב ביותר האפשרי עם מה שיש. אדם רציונלי מגבש את האמונות הטובות ביותר שהוא יכול עם הראיות שיש לו, לא משנה עד כמה הוא אבוד בסיטואציה. אדם רציונלי מקבל את ההחלטות הטובות ביותר שהוא יכול כדי לשפר את סיכויי ההצלחה שלו, לא משנה עד כמה נורא המצב בו הוא נתון.

בעולם האמיתי, אין משמעה של הרציונליות התעלמות מרגשות או מאינטואיציות. פעמים רבות, משמעה של רציונליות עבור בני אדם היא להיות יותר מודעים לרגשות שלהם, על מנת שיוכלו לשקלל אותם בקבלת ההחלטות שלהם.

רציונליות יכולה אפילו לומר לנו מתי לא לחשוב יותר מדי. במקרים של בחירת פוסטר לתלות על הקיר, או ניבוי של תוצאה של משחק כדורסל, נחקרים שחשבו לעומק על ההחלטה הראויה ביצועים גרועים יותר.^{3,4} ישנם מקרים בהם חשיבה מכוונת ומודעת משרתת אותנו היטב, ומקרים אחרים בהם שיפוט מהיר משרת אותנו טוב יותר.

פסיכולוגים העובדים על תאוריות של עיבוד דואלי מבחינים בין תהליכי "מערכת 1" של המוח (קוגניציה מהירה, מרומזת, אוטואיטיבית, אוטומטית) לבין תהליכי "מערכת 2" (קוגניציה איטית, מפורשת, אינטלקטואלית, נשלטת).⁵ הסטריאוטיפ הוא שרציונליסטים מסתמכים לחלוטין על מערכת 2, תוך התעלמות מרגשותיהם ודחפיהם. אם נביט מעבר לסטריאוטיפ, מישהו שמתנהג

באמת באופן רציונלי – באמת משיג את המטרות שלו, באמת ממזער את הנזק מההטיות הקוגניטיביות שלו – יסתמך במידה רבה על הרגלים ואינטואיציות של מערכת 1 במקרים בהם הם אמינים. לרוע המזל, מערכת 1 היא כנראה מדריך נוראי לגבי השאלה: "מתי עלי להשתמש במערכת 1?" האינטואיציות הלא מבוססות שלנו לא מסוגלות לומר לנו מתי להפסיק לסמוך עליהן. להיות מוטה ולהיות לא־מוטה מרגיש בדיוק אותו הדבר.⁶ מצד שני, כפי שהעיר דן אריאלי: אנחנו לא רציונליים באופן צפוי. אנחנו מפשלים באותן דרכים, שוב ושוב, באופן שיטתי. אם לא נוכל להשתמש בתחושת בטן כדי להבין מתי אנו נכנעים להטיה קוגניטיבית, ייתכן שנוכל להשתמש במדעי התודעה.

פניה הרבים של ההטיה

כדי לפתור בעיות, המוחות שלנו התפתחו להשתמש בהיוריסטיקות קוגניטיביות – קיצורי דרך גסים שמשיגים את התשובה הנכונה לעיתים קרובות, אך לא תמיד. הטיות קוגניטיביות מתעוררות כאשר עיגול הפינות של היוריסטיקות אלו מביא לטעות עקבית ומובחנת יחסית.

היוריסטיקת הייצוגיות, למשל, היא הנטייה שלנו להעריך הסתברות של תופעות על פי עד כמה הן נראות לנו כמייצגות של קטגוריות שונות. היוריסטיקה זו עשויה להוביל להטיות כמו כשל החיתוך. טֶבֶרְסְקִי וְפֶהֶנְמַן מצאו שנבדקים החשיבו כסביר פחות ששחקן טניס חזק "יפסיד את המערכה הראשונה" מאשר שהוא "יפסיד את המערכה הראשונה אך ינצח במשחק." קאמבק נראה כמו דבר אופייני יותר לשחקן חזק, אז אנו נוטים להעריך ביתר את ההסתברות של התרחיש המסובך־אך־נשמע־הגיוני הזה בהשוואה להסתברות של תרחיש שהוא בהכרח פשוט יותר. היוריסטיקת הייצוגיות יכולה לגרום גם להזנחת שיעור הבסיס,

הטיה בה אנו מבססים את השיפוט שלנו על רמת ה"נורמליות" לה זוכה צירוף של תכונות, לדעתנו, תוך הזנחת השאלה: "עד כמה נפוצה כל אחת מהתכונות האלו בכל האוכלוסייה?"⁸ האם סביר יותר שסטיב הוא ספרן ביישן, או שהוא איש מכירות ביישן? רוב האנשים עונים על שאלה מהסוג הזה בעזרת ניסיון מנטלי להתאים את התואר "ביישן" לכל אחד מהמקצועות האלה. הם לא לוקחים בחשבון עד כמה אנשי מכירות נפוצים יותר מספרנים – פי שבעים וחמישה בארצות הברית.⁹

דוגמאות נוספות להטיות הינן *הזנחת משך הזמן* (הערכת חוויות מבלי להתייחס לכמות הזמן שהן נמשכו), *כשל העלות השקועה* (תחושת מחויבות לדברים בהם הושקעו משאבים בעבר, במקום לחתוך הפסדים ולהמשיך הלאה), *והטיית האישוש* (מתן משקל רב יותר לראיות המאששות את מה שאנו כבר מאמינים בו).¹⁰⁻¹¹

לרוע המזל, הכרת ההטיה לרוב אינה מספקת חיסון מפניה. במחקר על *עיוורון להטיות*, נבדקים חזו שאם הם יידעו שצירור מסוים הוא עבודה של אמן מפורסם, יהיה להם קשה יותר להעריך באופן אובייקטיבי את איכות הצירור. ואכן, נבדקים שנאמר להם מי האומן והתבקשו להעריך את איכות הצירור הציגו בדיוק את אותה ההטיה שהם עצמם חזו, ביחס לקבוצת הביקורת. אולם כשנשאלו *לאחר מכן*, אותם נבדקים ממש טענו שההערכות שהם נתנו לצירורים היו אובייקטיביות ולא מושפעות כלל מההטיה – בשתי הקבוצות!¹²⁻¹³

אנו מתעבים במיוחד את המחשבה לפיה ההשקפות שלנו מדוייקות פחות משל אחרים. אפילו כשאנו מזהים נכונה את ההטיות של אחרים, יש לנו *נקודה עיוורת* מיוחדת להטיות בכל הנוגע לפגמינו שלנו. כשאנו מביטים פנימה, אנחנו לעולם לא מזהים "מחשבות-שמרגישות-מוטות", ובשל כך מסיקים את המסקנה הבלתי נמנעת שאנחנו כנראה פשוט אובייקטיביים יותר מכל השאר.¹⁴

למעשה, למידה על הטיות יכולה להפוך אתכם לפגיעים יותר לביטחון-יתר ולהטיית האישוש, כיוון שאתם מתחילים לראות את ההשפעה של הטיות קוגניטיביות בכל מקום סביבכם – אצל כולם חוץ מכם. ובשונה מהטיות רבות, הנקודה העיוורת להטיות *חמורה*

יותר בקרב אנשים אינטליגנטיים, מלאי מחשבה ופתוחים מחשבתית
כמיוחד. ^{15,16}

זו סיבה לדאגה.

ובכל זאת... נדמה שאנחנו אמורים להיות מסוגלים ליותר. ידוע שאנחנו יכולים להפחית את הזנחת שיעור הבסיס על ידי חשיבה על הסתברויות כתדירויות של אובייקטים או אירועים בזמן. אנחנו יכולים למזער את הזנחת משך הזמן על ידי הפניית תשומת לב רבה יותר למשך הזמן בו התרחש אירוע והצגתו באופן גרפי.¹⁷ מידת העוצמה בה הטיות קוגניטיביות באות לידי ביטוי משתנה מאדם לאדם, כך שאמורות להיות מגוון דרכים שטרם התגלו היכולות להשפיע על הרמה בה אנו מוטים.

עם זאת, אם ברצוננו להשתפר, אין די בכך שנעמיק לחקור ברשימות של הטיות קוגניטיביות. הגישה לניטרול-הטיות ברציונליות: מ-א' ועד ת' היא לתווך הבנה שיטתית של הסיבות בעטיין חשיבה רציונלית עובדת, ושל הדרך בה המוח חושב באופנים שאינם רציונליים. ככל שהכרך הזה עושה את עבודתו, ניתן להשוות את גישתו לזו המתוארת אצל סֶרְפֶס, המציין ש"שנים של ניסיון בפנינסים" לא השפיעו על רגישותם של אנשים להטיית העלות השקועה, בעוד ש"מספר הקורסים שנלמדו בחשבונאות" כן עזר.

כתוצאה מכך, ייתכן שיש צורך להבחין בין ניסיון לבין מומחיות, כאשר משמעות המומחיות היא "פיתוח של עיקרון סכמטי המערב הבנה מושגית של הבעיה", שבתורו מאפשר למקבל ההחלטות לזהות הטיות מסוימות. עם זאת, השימוש במומחיות כאמצעינגד דורש יותר מאשר היכרות עם התוכן המצבי או מומחיות בתחום מסוים. הוא דורש שהאדם יבין במלואו את הרציונל הבסיסי של ההטיה המדוברת, שיהיה מסוגל לזהות אותה בסביבה הספציפית, ושיחזיק בידיו את הכלים המתאימים על מנת לנטרל את ההטיה.¹⁸

מטרתו של ספר זה היא להניח את היסודות ליצירת "מומחיות" ברציונליות. משמעות הדבר היא היכולת להבין לעומק את המבנה

של בעיה כללית מאוד: ההטיה האנושית, ההונאה העצמית, ואלף הדרכים בהן מחשבה מתוחכמת יכולה להביס את עצמה.

מילה על הטקסט הזה

רציונליות: מ-א' ועד ת' החל את חייו כסדרת מאמרים מאת אליעזר יודקובסקי, שפורסמו בין 2006 ל-2009 בבלוג הכלכלה *Overcoming Bias* ובבלוג הקהילה שהתפצל ממנו, *Less Wrong*. עבדתי עם יודקובסקי בשנה האחרונה במכון לחקר בינה מלאכותית (MIRI), מוסד ללא כוונת רווח שהוא ייסד בשנת 2000 כדי לחקור את הדרישות התיאורטיות לבינה מלאכותית חכמה-מאדם (AI)*. קריאת הפוסטים שלו בבלוג גרמה לי להתעניין בעבודתו. התרשמתי מיכולתו לתקשר בתמציתיות תובנות שלקח לי שנים של לימודי פילוסופיה אנליטית להפנים. בניסיונו ליישב את הרוח האנרכיסטית והסקפטית של המדע עם גישה קפדנית ושיטתית לחקירה, יודקובסקי לא רק מנסה להפריך את הצעדים השגויים הרבים והמבוי הסתום שפילוסופיה גרועה (וחוסר-פילוסופיה גרוע) יכולים לייצר, אלא גם להביך אותם. תקוותי היא כי בעזרת העריכה של חיבורים אלו לכדי ספר, הקלתי על הצלילה לתוכם ועל הערכתם כמכלול קוהרנטי.

מדריך הרציונליות שנוצר הוא לעתים קרובות אישי וחסר יראת כבוד. הוא שואב דוגמאות, בין היתר, מחוויותיו של יודקובסקי עם אימו (פסיכיאטרית) ואביו (פיזיקאי) היהודים-אורתודוקסים, משיחות בחדרי צ'אט ומרשימות תפוצה במייל. קוראים המכירים

* יש לציין כי ספר זה נכתב לפני מהפכת מודלי השפה הגדולים (LLM) שהחלה בשנת 2017 עם המצאת הטרנספורמר ופרצה לתודעה הציבורית בשנת 2022 עם השקת ChatGPT. המושג בינה מלאכותית (AI) לאורך הספר מתייחס למחקר התיאורטי של מידול תהליכי חשיבה והסקה. (כל הערות השוליים המסומנות בכוכביות הן של המתרגמים.)

את יודקובסקי מהארי פוטר והשיטה הרציונלית, הפאן־פיקשן המדעי שכתב לספרי הארי פוטר של ג'יי קיי רולינג, יזהו את אותו ניפוך מוסכמות חצוף, ורבים מאותם מושגי ליבה.

מבחינה סגנונית, החיבורים בספר זה נעים על המנעד שבין "ספר לימוד מלא חיים" דרך "אוסף אפיזודות מלאות מחשבה" ועד "מניפסט מתפרע", והתוכן מגוון בהתאם. רציונליות: מ-א' ועד ת' מאגד מאות פוסטים מהבלוג של יודקובסקי לכדי עשרים ושישה "רצפים", סדרות דמויות־פרקים של פוסטים בעלי קשר נושאי. הרצפים מקובצים לשישה ספרים, העוסקים בנושאים הבאים:

ספר I - מפה ושטח. מהי אמונה, ומה גורם לאמונות מסוימות לעבוד טוב יותר מאחרות? ארבעת הרצפים הללו מסבירים את המושגים הבייסיאניים של רציונליות, אמונה וראיות. נושא מרכזי: הדברים שאנו קוראים להם "הסברים" או "תיאוריות" עשויים שלא לתפקד כמו מפות לניווט בעולם. כתוצאה מכך, אנו עשויים לבלבל את המפות המנטליות שלנו עם שאר החפצים בארגז הכלים שלנו.

ספר II - איך לשנות את דעתך הלכה למעשה. עניין ה'אמת' הזה נראה די שימושי. למה, אם כן, אנחנו ממשיכים לקפוץ למסקנות, להתחפר בעמדותינו, ולחזור על אותן טעויות? למה אנחנו כל כך גרועים ברכישת אמונות מדויקות, ואיך אנו יכולים להצליח יותר? שבעת הרצפים הללו דנים בחשיבה מגמתית והטיית האישוש, עם דגש מיוחד על סוגים קשים־לזיהוי של הונאה עצמית והמלכודת של "שימוש בטיעונים בתור חיילים".

ספר III - המכונה שברוח. מדוע לא התפתחנו להיות רציונליים יותר? אפילו כשלוקחים בחשבון מגבלות משאבים, נראה שהיינו יכולים להרוויח הרבה יותר ידע תמורת אותה כמות ראיות. כדי לקבל תמונה ריאליסטית של איך ולמה המוח שלנו מממש את הפונקציות הביולוגיות שלו, עלינו להרים את מכסה המנוע ולראות איך האבולוציה עובדת, ואיך המוח שלנו עובד, באופן מדויק יותר. שלושת הרצפים הללו מדגימים איך אפילו פילוסופים ומדענים יכולים ליפול בפח כשהם רותמים את האבולוציה או הפסיכולוגיה לטובת

הסברים אינטואיטיביים ולא־טכניים. על ידי מיקום המוח שלנו בתוך מרחב גדול יותר של מערכות מוכוונות־מטרה, אנו יכולים לזהות כמה מהמוזרויות של החשיבה האנושית ולהעריך כיצד מערכות כאלו יכולות "לאבד את מטרתן".

ספר IV - מציאות כהווייתה. באיזה מין עולם אנו חיים? מה מקומנו בעולם הזה? בהתבסס על הדוגמאות של הרצפים הקודמים לאופן שבו מודלים אבולוציוניים וקוגניטיביים עובדים, ששת הרצפים הללו חוקרים את טבע התודעה ואת אופיו של החוק הפיזיקלי. בנוסף ליישום והכללה של שיעורי עבר על תעלומות מדעיות וחסכון בהנחות מוצא, חיבורים אלו מעלים שאלות חדשות לגבי התפקיד שהמדע צריך לשחק ברציונליות הפרטית.

ספר V - טוב כהווייתו. מה הופך משהו לבעל ערך – מוסרית, אסתטית, או תועלתנית? שלושת הרצפים הללו שואלים כיצד אנו יכולים להצדיק, לשנות או לנטרל את הערכים והתשוקות שלנו. המטרה תהיה למצוא דרך להבין את המטרות שלנו מבלי להתפשר על המאמצים שלנו להשיג אותן בפועל. האתגר הגדול ביותר כאן הוא לדעת מתי לסמוך על הדחפים האיפולסיביים, המסובכים והספציפיים שלנו לגבי מה נכון ומה לא נכון, ומתי להחליף אותם בעקרונות פשוטים ומוחלטים.

ספר VI - להפוך לחזק יותר. איך אנשים וקהילות יכולים ליישם את כל העקרונות הללו הלכה למעשה? שלושת הרצפים הללו מתחילים בתיאור אוטוביוגרפי של הטעויות הפילוסופיות הגדולות ביותר של יודקובסקי עצמו, כולל עצות לגבי איך הוא חושב שאחרים עשויים להימנע מאותם מוקשים עליהם הוא דרך. הספר נחתם בהמלצות לפיתוח תוכניות לימודים לרציונליות יישומית מבוססת־ראיות, וליצירת קבוצות ומסודות שייתמכו בתלמידים, מחנכים, חוקרים וחברי קהילה שיתעניינו בכך.

הרצפים מלווים גם ב"אינטרלודים", חיבורים שנלקחו מהאתר האישי של יודקובסקי. אלו קשורים לרצפים בדרכים שונות. למשל, שתיים־עשרה המעלות של הרציונליות מסכם באופן פואטי רבים

מהשיעורים של רציונליות: מ-א' ועד ת', ומצוטט לעיתים קרובות בחיבורים אחרים.

מפה ושטח

הספר הראשון, שאתם קוראים כעת, מתחיל ברצף העוסק בהטיות קוגניטיביות: *שגוי באופן צפוי*. שאר הספר לא ייצמד רק לנושא הזה. להרגלים רעים ולרעיונות רעים יש חשיבות, אפילו כשהם נובעים מתוך תוכן התודעה שלנו, ולא ממבנה התודעה שלנו. לפיכך, ברצפים הבאים יוצגו גם טעויות שהתפתחו וגם טעויות שהומצאו, החל מהדיון ברצף *אמונות מזויפות* על הדרכים שבהן ציפיותיו של האדם עשויות להיות שונות מאמונותיו המוצהרות.

תיאור של אירציונליות יהיה בהכרח לוקה בחסר אם לא יספק שום תאוריה לגבי האופן שבו עובדת רציונליות – או אם ה"תאוריה" שלו תהיה מורכבת רק מאמיתות נדרשות ומעורפלות, ללא שום מנגנון תיאורי מדויק. הרצף *לשים לב לבלבול* שואל מדוע יהיה זה מועיל לבסס את ההתנהגות על ציפיות "רציונליות", ואיך אנו עשויים להרגיש בעשותנו זאת.

בהמשך, *תשובות מסתוריות* שואל האם המדע פותר את הבעיות הללו עבורנו. מדענים מבססים את המודלים שלהם על ניסויים ניתנים לשחזור, ולא על ספקולציות או שמועות. ולמדע יש רקורד מצוין בהשוואה לאנקדוטות, דת, ו... כל דבר אחר, בגדול. האם עדיין עלינו לחשוש מפני אמונות "מזויפות", הטיית האישוש, הטיית החוכמה בדיעבד וכיוצא באלה, בזמן שאנו עובדים עם קהילה של אנשים המנסים להסביר תופעות, ולא רק לספר סיפורים שנעים לשמוע? לאחר מכן מגיע האינטרלוד *האמת הפשוטה*, משל העומד בפני עצמו על טבעם של הידע והאמונה.

עם זאת, ההטייה הקוגניטיבית היא זו אשר מספקת את ההצצה הברורה והישירה ביותר אל תוך החומר שממנו עשויה הפסיכולוגיה

שלנו, אל צורת ההיוריסטיקות שלנו. אל הלוגיקה של המגבלות שלנו. את הספר הזה נפתח בהטיות.

ישנו קטע בג'ואנג ז'זה, טקסט פילוסופי פרוטרדאואיסטי, שאומר: "מלכודת הדגים קיימת בגלל הדג. ברגע שהשגת את הדג, אתה יכול לשכוח מהמלכודת."¹⁹

אני מזמין אתכם לחקור את הספר הזה ברוח זו. השתמשו בו כפי שהייתם משתמשים במלכודת דגים, מודעים כל העת למטרה עבורה אתם קוראים בו. שאו עמכם את מה שיהיה לכם לעזר, כל עוד יש לו שימוש. לכל השאר הניחו. מי ייתן ומטרתכם תשרת אתכם נאמנה.

תודות

אני אסיר תודה באופן יוצא מגדר הרגיל לנייט סוארס, אליזבת טרלטון, פול קראולי, בריאן סטרול, אדם פריז, הלן טונר, ועשרות מתנדבים נוספים על ההגהה של חלקים מספר זה.

תודה מיוחדת וכנה לאלכס ורמיר, שניווט את הספר הזה לידי סיום, ולצבי בנסון־טילסן, שעבר על הספר כולו כדי להבטיח את קריאותו ואחידותו.

-
1. המושג של אישיות מוטה, תקשורת מוטה וכדומה, דומה להטיה סטטיסטית בכך שמדובר בשגיאה. לעומת זאת, דרכים אחרות להכללת המושג "הטיה" מתמקדות דווקא בזיקה שלו לחוסר אקראיות. בלמידת מכונה למשל, הטיה אינדוקטיבית היא פשוט אוסף ההנחות שהלומד משתמש בהן כשהוא מבצע תחזיות מתוך מערך נתונים. כאן, הלומד "מוטה" במובן זה שהוא מצביע לכיוון מסוים. אך מכיוון שכיוון זה עשוי להיות נכון, אין זה דבר רע שלסוכן תהיה הטייה אינדוקטיבית. היא בעלת ערך ואף הכרחית. בכך נבדלת בבירור ה"הטיה" האינדוקטיבית מסוגי הטיות אחרים.
 2. צירוף מקרים מצער: לנרד נימוי, השחקן שגילם את ספוק, הלך לעולמו ימים ספורים בלבד לפני יציאת הספר הזה לאור (באנגלית). אף על פי שאנו מצטטים את דמותו כדוגמה קלאסית ל"רציונליות הוליוודית" מזויפת, אין בכוננתנו להביע חוסר כבוד לזכרו של נימוי.
 3. Timothy D. Wilson et al., "Introspecting About Reasons Can Reduce Post-choice Satisfaction," *Personality and Social Psychology Bulletin* 19 (1993): 331-331.
 4. Jamin Brett Halberstadt and Gary M. Levine, "Effects of Reasons Analysis on the Accuracy of Predicting Basketball Games," *Journal of Applied Social Psychology* 29, no. 3 (1999): 517-530.
 5. Keith E. Stanovich and Richard F. West, "Individual Differences in Reasoning: Implications for the Rationality Debate?," *Behavioral and Brain Sciences* 23, no. 5 (2000): 645-665, http://journals.cambridge.org/abstract_S0140525X00003435.
 6. Timothy D. Wilson, David B. Centerbar, and Nancy Brekke, "Mental Contamination and the Debiasing Problem," in *Heuristics and Biases: The Psychology of Intuitive Judgment*, ed. Thomas Gilovich, Dale Griffin, and Daniel Kahneman (Cambridge University Press, 2002).
 7. Amos Tversky and Daniel Kahneman, "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment," *Psychological Review* 90, no. 4 (1983): 293-315, doi:10.1037/0033295X.90.4.293.
 8. Richards J. Heuer, *Psychology of Intelligence Analysis* (Center for the Study of Intelligence, Central Intelligence Agency, 1999).
 9. Wayne Weiten, *Psychology: Themes and Variations, Briefer Version, Eighth Edition* (Cengage Learning, 2010).

Raymond S. Nickerson, "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises," *Review of General Psychology* 2, no. 2 (1998): 175.

11. הזנחת הסתברות היא הטיה קוגניטיבית נוספת. בחודשים ובשנים שלאחר פיגועי ה-11 בספטמבר, אנשים רבים בחרו לנהוג מרחקים ארוכים במקום לטוס. חטיפת מטוס לא הייתה אירוע סביר, אך כעת התחושה הייתה שהיא משהו שעשוי לקרות. עצם האפשרות של חטיפה השפיעה במידה דרמטית על קבלת ההחלטות. על ידי הסתמכות על חשיבה בינארית (מכוניות ומטוסים הם או "בטוחים" או "לא בטוחים", סוף פסוק), אנשים למעשה העמידו את עצמם בסכנה גדולה בהרבה. במקום לשקול את ההסתברות למות בנסיעה ברכב מקצה אחד של המדינה לקצה השני אל מול ההסתברות למות בטיסה כזו – כשהראשונה סבירה פי כמה מאות – הם הסתמכו על תחושות כלליות של דאגה וחרדה (היוריסטיקה הרגש). ניתן להבחין בדפוס התנהגות דומה בקרב ילדים אשר נחשפים לטיעונים בעד ונגד הבטיחות של חגורות בטיחות. אלו נוטים לנוע כמטוטלת בין התפיסה שחגורות בטיחות הן רעיון טוב לחלוטין או רעיון גרוע לחלוטין, במקום לנסות להשוות בין עוצמת השיקולים לכאן ולכאן.²¹ דוגמאות נוספות להטיות הן: כלל "שיא-סוף" (הערכת אירועים מהזיכרון על סמך רגע השיא שלהם ועל סמך האופן שבו הסתיימו), עיגון (ביסוס החלטות על מידע שנתקלנו בו לאחרונה, גם כשהוא לא רלוונטי)²² ועיגון עצמי (שימוש בעצמך כמודל למאפיינים הסבירים של אחרים, מבלי להקדיש מחשבה מספקת לדרכים שבהן אתה עצמך יוצא דופן),²³ והטיית הסטטוס קוו (העדפה מוגזמת של הקיים והצפוי על פני החדש והשונה).²⁴

12. Katherine Hansen et al., "People Claim Objectivity After Knowingly Using Biased Strategies," *Personality and Social Psychology Bulletin* 40, no. 6 (2014): 691–699.

13. באופן דומה, פרונין כותבת על עיוורון להטיה מגדרית:

באחד המחקרים, משתתפים בחנו מועמד ומועמדת לתפקיד פיקודי במשטרה, ולאחר מכן נדרשו להעריך איזו מהתכונות חשובה יותר לתפקיד: "ניסיון מהשטח" או "השכלה פורמלית". התוצאה הייתה שהמשתתפים העדיפו את הרקע שהיה למועמד הגבר, כפי שתואר להם (למשל, אם נאמר להם שהוא "בעל ניסיון מהשטח", הם העריכו תכונה זו כחשובה יותר). המשתתפים היו עיוורים לחלוטין להטיה המגדרית הזו. למעשה, ככל שהם האמינו שהם אובייקטיביים יותר, כך הם הפגינו הטיה חזקה יותר בפועל.²⁵

פרונין מציינת כי גם כאשר אנו מודעים להטיות, אנו נותרים "ריאליסטים נאיביים" בנוגע לאמונות שלנו. אנו נסוגים באופן עקבי לדפוס של התייחסות לאמונות שלנו כאל ייצוגים נטולי עיוותים של המציאות כפי שהיא באמת.²⁶

14. בסקר שנערך בקרב 76 אנשים הממתינים בנמלי תעופה, המשתתפים דירגו את

עצמם כמושפעים הרבה פחות מהטיות קוגניטיביות מאשר האדם הממוצע
בנמל התעופה. בפרט, אנשים נוטים לחשוב על עצמם כחסינים במיוחד מפני
הטיות כאשר ההטיה אינה רצויה מבחינה חברתית, או כאשר השלכותיה
קשות להבחנה.²⁷ מחקרים אחרים מצאו כי אנשים בעלי נגיעה אישית לנושא
כלשהו תופסים זאת כגורם המעצים את ההבנה והאובייקטיביות שלהם. אך
כאשר הם רואים אנשים אחרים המציגים את אותה נגיעה אישית בדיוק, הם
סבורים כי אותם אנשים נוגעים בדבר יתר על המידה ועל כן מוטים בשיקול
דעתם.

Joyce Ehrlinger, Thomas Gilovich, and Lee Ross, "Peering Into the .15
Bias Blind Spot: People's Assessments of Bias in Themselves and
Others," *Personality and Social Psychology Bulletin* 31, no. 5 (2005):
680-692.

Richard F. West, Russell J. Meserve, and Keith E. Stanovich, .16
"Cognitive Sophistication Does Not Attenuate the Bias Blind Spot,"
Journal of Personality and Social Psychology 103, no. 3 (2012): 506.

.17 ... אין לבלבל זאת עם אנשים החושבים שהם נבונים או רגישים במיוחד וכו', עקב
הטיית העליונות המדומה.

Michael J. Liersch and Craig R. M. McKenzie, "Duration Neglect by .18
Numbers and Its Elimination by Graphs," *Organizational Behavior
and Human Decision Processes* 108, no. 2 (2009): 303-314.

Sebastian Serfas, *Cognitive Biases in the Capital Investment Context: .19
Theoretical Considerations and Empirical Experiments on Violations
of Normative Rationality* (Springer, 2010).

Zhuangzi and Burton Watson, *The Complete Works of Zhuangzi* .20
(Columbia University Press, 1968).

Cass R. Sunstein, "Probability Neglect: Emotions, Worst Cases, and .21
Law," *Yale Law Journal* (2002): 61-107.

Dan Ariely, *Predictably Irrational: The Hidden Forces That Shape Our .22
Decisions* (HarperCollins, 2008).

Boaz Keysar and Dale J. Barr, "Self-Anchoring in Conversation: .23
Why Language Users Do Not Do What They 'Should,'" in
*Heuristics and Biases: The Psychology of Intuitive Judgment: The
Psychology of Intuitive Judgment*, ed. Griffin Gilovich and Daniel
Kahneman (New York: Cambridge University Press, 2002), 150-166,
doi:10.2277/0521796792.

Scott Eidelman and Christian S. Crandall, "Bias in Favor of the Status .24

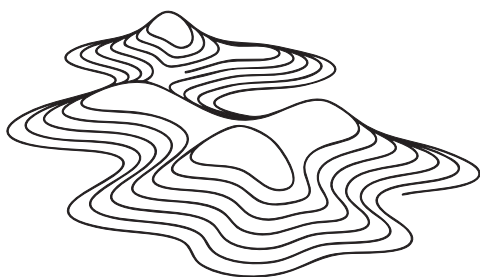
Quo," *Social and Personality Psychology Compass* 6, no. 3 (2012): 270-281.

Eric Luis Uhlmann and Geoffrey L. Cohen, "I think it, therefore it's true': Effects of Self-perceived Objectivity on Hiring Discrimination," *Organizational Behavior and Human Decision Processes* 104, no. 2 (2007): 207-223.

Emily Pronin, "How We See Ourselves and How We See Others," *Science* 320 (2008): 1177-1180, <https://www.cs.ucdavis.edu/~franklin/ecs188/2015/fall/pronin.pdf>.

Emily Pronin, Daniel Y. Lin, and Lee Ross, "The Bias Blind Spot: Perceptions of Bias in Self versus Others," *Personality and Social Psychology Bulletin* 28, no. 3 (2002): 369-381.

מפה ושטח



חלק א'

שגוי באופן צפוי

... [ספר זה] מתחיל ב'רצף' העוסק בהטיה קוגניטיבית: "שגוי באופן צפוי". שאר הספר לא ייצמד רק לנושא הזה. להרגלים רעים ולרעיונות רעים יש חשיבות, אפילו כשהם נובעים מתוך תוכן התודעה שלנו, ולא ממבנה התודעה שלנו.

— רוב בנסינגר,

הטיות: מבוא

1

למה אני מתכוון כשאני אומר 'רציונליות'?

אני מתכוון ל:

1. רציונליות הכרתית: דיוק שיטתי של האמונות שלך.
 2. רציונליות מעשית: הגשמה שיטתית של הערכים שלך.
- כשאתה פוקח את עיניך ומסתכל על החדר שמסביבך, אתה מאתר את המחשב הנייד ביחס לשולחן, ואת כוננית הספרים ביחס לקיר. אם משהו בעיניים שלך, או במוח שלך, משתבש, אזי המודל המנטלי שבנית עשוי לומר שיש כוננית במקום בו אין כוננית כלל, וכשתלך לקחת ספר, נכונה לך אכזבה.
- זו התחושה שעולה מהחזקה באמונה שגויה. מפה של העולם שאינה תואמת את השטח. רציונליות הכרתית עוסקת בשרטוט של מפות מדויקות. ההתאמה הזו בין אמונה לבין המציאות נקראת בדרך כלל "אמת", ואני שמח לקרוא לה כך.
- רציונליות מעשית, לעומת זאת, עוסקת בהכוונת המציאות – ניווט העתיד למקום אליו אתה רוצה שהוא יגיע. זוהי האומנות של בחירת הפעולות שמובילות לתוצאות המדורגות גבוה יותר בסולם ההעדפות שלך. לפעמים אני קורא לזה "לנצח".
- רציונליות, אם כן, עוסקת בגיבוש אמונות אמיתיות ובקבלת החלטות מנצחות.

החתיירה ל"אמת" במקרה הזה אין משמעה ביטולן של ראיות חלקיות או עקיפות. הפניית המבט אל החדר שסביבך ובניית מפה מנטלית שלו, אינה שונה באופן מהותי מלהאמין שלכדור הארץ יש ליבה מותכת, או שיויליוס קיסר היה קירח. השאלות הללו, שהן רחוקות ממך במרחב ובזמן, עשויות להיראות קלושות ומופשטות יותר משאלות לגבי כוננית הספרים שלך, אך למרות הכל, ישנן עובדות לגבי מצב ליבת כדור הארץ בשנת 2015 לספירה ולגבי מצב ראשו של קיסר בשנת 50 לפנה"ס. עובדות אלו עשויות להשפיע עליך באופן אמיתי לחלוטין, גם אם לעולם לא תמצא דרך לפגוש את קיסר או את הליבה פנים-אל-פנים.

"לנצח" במקרה הזה לא חייב לבוא על חשבונם של אחרים. מסע החיים של כל אדם יכול לעסוק בשיתוף פעולה או אפילו הקרבה עצמית, ולא דווקא בתחרות. בהקשר הזה, "הערכים שלך" פירושו כל דבר שאכפת לך ממנו, כולל אנשים אחרים. ערכים אלו אינם מוגבלים לערכים אנוכיים או ערכים שאינם משותפים.

כשאנשים אומרים "X הוא רציונלי!", זו לרוב פשוט דרך נחרצת יותר לומר "אני חושב ש-X הוא נכון" או "אני חושב ש-X הוא טוב". אז לשם מה אנו זקוקים ל"רציונלי" בנוסף ל"נכון" ו"טוב"? ניתן להשתמש בטיעון דומה כנגד השימוש במילה "אמת": אין צורך לומר "זוהי אמת שהשלג לבן" כשאפשר פשוט לומר "השלג לבן". מה שהופך את רעיון האמת לשימושי זה שהוא מאפשר לנו לדבר על המאפיינים הכלליים של התאמת מפה-שטח. "מודלים אמיתיים מייצרים לרוב ניבויים מדויקים יותר של המציאות ממודלים שקריים" היא הכללה שימושית, שלא ניתן לעשות ללא שימוש במושגים כמו "אמת" או "דיוק".

באופן דומה, "סוכנים רציונליים מקבלים החלטות שממקסמות את התוחלת ההסתברותית של פונקציית תועלת קוהרנטית" היא סוג של מחשבה שתלויה במושג של רציונליות (מעשית), בעוד שאת המשפט "לאכול ירקות זה דבר רציונלי" ניתן כנראה להחליף ב"לאכול ירקות זה דבר מועיל" או "לאכול ירקות משרת את האינטרס שלך". אנו

זקוקים למושג כמו "רציונלי" כדי לציין עובדות כלליות על אותן דרכי חשיבה המייצרות באופן שיטתי אמת או ערך – ועל הדרכים השיטתיות בהן אנו סוטים מאותם הסטנדרטים.

לעיתים קורה שפסיכולוגים ניסויים חושפים חשיבה אנושית שנראית מוזרה מאוד. למשל, מישהו מדרג את ההסתברות ש"ביל מנגן ג'אז" כנמוכה מההסתברות ש"ביל הוא רואה חשבון שמנגן ג'אז". זה נראה כמו שיפוט מוזר, מכיוון שכל רואה חשבון שמנגן ג'אז הוא בוודאות נגן ג'אז. אולם על איזה בסיס אנו נשענים כשאנו טוענים שהשיפוט הזה מוטעה?

פסיכולוגים ניסויים משתמשים בשתי נקודות ייחוס: תורת ההסתברות ותורת ההחלטות.

תורת ההסתברות היא מערכת החוקים העומדת בבסיס האמונה הרציונלית. המתמטיקה של ההסתברות מתארת באופן שווה וללא הבחנה את (א) בירור מיקומה של כוננית הספרים שלך, (ב) בירור הטמפרטורה של ליבת כדור הארץ, ו-(ג) הערכת מספר השערות שהיו על ראשו של יוליוס קיסר. זוהי אותה בעיה של עיבוד ראיות ותצפיות על מנת לעדכן את האמונות שלנו. באופן דומה, תורת ההחלטות היא מערכת החוקים העומדת בבסיס הפעולה הרציונלית, והיא ישימה באותה מידה ללא קשר למטרותיו של האדם ולאפשרויות העומדות בפניו.

יהי P (כך יוכך) "ההסתברות שכך יוכך יקרה", ו- $P(A, B)$ "ההסתברות שגם א' וגם ב' יקרו". כיוון שזהו חוק אוניברסלי של תורת ההסתברות ש- $P(A) \leq P(A, B)$, השיפוט שקובע כי P (ביל מנגן ג'אז) נמוך מ- P (ביל מנגן ג'אז, ביל הוא רואה חשבון) מתויג כשגוי. אם לנקוט בלשון טכנית, ניתן לומר שהשיפוט ההסתברותי הזה הוא לא־בייסיאני. אמונות ופעולות שהן רציונליות במובן המוגדר היטב מתמטית באופן הזה נקראות "בייסיאניות".

מעניין להבחין כי המושג המודרני של רציונליות אינו עוסק בחשיבה מילולית דווקא. בדוגמא שנתתי אתם פוקחים עיניים, מביטים מסביב, ובונים מודל מנטלי של חדר המכיל כוננית ספרים. המושג המודרני של רציונליות הוא כללי מספיק כדי לכלול את

העיניים שלכם ואת קליפת הראיה במוח שלכם כדברים־ממפים. הוא כולל גם את האינטואיציות שלכם, הנחוות ללא מילים. למתמטיקה לא אכפת אם אנו משתמשים באותה מילה: "רציונלי", כדי לתאר את ספוק ממסע בין כוכבים וכדי לתאר חשיבה בייסיאנית. המתמטיקה ממדלת דרכים טובות להשגת מטרות או למיפוי העולם, ללא קשר לשאלה אם דרכים אלו תואמות את הדעות הקדומות והסטריאוטיפים שלנו לגבי מה ש"רציונליות" אמורה להיות.

התיאור הזה לא באמת עונה על השאלה לגבי המשמעות של "רציונליות" הלכה למעשה, משתי סיבות עיקריות:

ראשית, הפורמליזם הבייסיאני בצורתו המלאה הוא בלתי־פתיר מבחינה חישובית עבור רוב הבעיות בעולם האמיתי. אף אחד לא באמת יכול לחשב ולציית למתמטיקה, בדיוק כפי שאיש אינו יכול לחזות את שוק המניות על ידי חישוב תנועותם של קווארקים.

זו הסיבה שישנו אתר שלם בשם "Less Wrong", ולא דף בודד שפשוט מנסח את האקסיומות הפורמליות וסוגר עניין. ישנה אומנות שלמה למציאת האמת והשגת ערך מתוך המוח האנושי: עלינו ללמוד את הפגמים של עצמנו, להתגבר על ההטיות שלנו, להימנע מהונאה עצמית, להיכנס לכושר מנטלי כדי להתעמת עם האמת ולעשות את מה שצריך להיעשות, וכן הלאה וכן הלאה.

שנית, ישנם מקרים בהם משמעותה של המתמטיקה עצמה מוטלת בספק. למשל, הכללים המדויקים של תורת ההסתברות מוטלים בספק על ידי בעיות אנתרופיות, בהן מספר הצופים אינו ודאי.* למשל, הכללים המדויקים של תורת ההחלטות מוטלים בספק על ידי בעיות דמויות־ניוקום, בהן שחקנים אחרים עשויים לחזות את ההחלטה שלך עוד לפני שהיא מתרחשת.**

במקרים כאלה, יהיה זה חסר תועלת לנסות ליישב את הבעיה על

* בבעיות אלו חישוב ההסתברות קשור בעצם קיומו של הצופה. ללא נתונים על כמות הצופים, לא ניתן לקבוע אם התצפית שלנו היא מקרה מייצג או חריג סטטיסטי.

** בבעיות אלו כל החלטה משפיעה בדיעבד על מה שהאחרים חוו לגביה, ועל כן את הנתונים שלפיהם אמורה להתקבל החלטה זו עצמה.

ידי המצאת הגדרה חדשה כלשהי למילה "רציונלי" ולומר: "לכן התשובה המועדפת עליי, מעצם ההגדרה, היא המשמעות של המילה 'רציונלי'." ניסוח כזה מעלה את השאלה: למה שמישהו יתעניין בהגדרה שלך? אני לא מתעניין בתורת ההסתברות משום שהיא המילה הקדושה שנמסרה לנו על ידי לפֿלֶס. אני מתעניין בעדכון אמונות בסגנון בייסיאני (עם הנחות המוצא של אוקאם) משום שאני מצפה שסגנון החשיבה הזה יקרב אותי באופן שיטתי לדבר הזה, אתם יודעים, דיוק, המפה שמשקפת את השטח.

נוסף לכל אלה יש גם שאלות על איך כדאי לחשוב, שנראה שאינן מקבלות מענה מספק – לא על ידי תורת ההסתברות ולא על ידי תורת ההחלטות – כמו השאלה איך להרגיש לגבי האמת מרגע שהשגת אותה. גם כאן, הניסיון להגדיר "רציונליות" באופן ספציפי אינו מספק תשובה, אלא רק מניח שיש אחת כזו.

אין לי שום עניין להתווכח על משמעות של מילה, אפילו אם המילה הזו היא "רציונליות". המטרה של חיבור בין מושגים מופשטים לרצפים של אותיות היא כדי לאפשר לשני אנשים לתקשר – לעזור להעביר מחשבות ממוח אחד לאחר. אינך יכול לשנות את המציאות, או להוכיח רעיון, באמצעות שינוי ההתאמה שבין מילים למשמעויות. אז אם הבנתם לאיזה מושג אני חותר בגדול עם המילה "רציונליות", ועם תתייהמונחים שלה "רציונליות הכרתית" ו"רציונליות מעשית", הרי שתקשרנו: השגנו כל מה שיש להשיג על ידי שיחה לגבי מהי "רציונליות". מה שנותר לדון בו אינו איזו משמעות להצמיד להברות "רצ-יו-נ-לי-יות". כל שנותר לדון בו הוא, איזוהי דרך טובה שיחשוב לו האדם.

אם אתה אומר, "מבחינה רציונלית נכון עבורי להאמין ב-א, אבל האמת היא ב", אז אתה כנראה משתמש במילה "רציונלי" כדי לציין משהו אחר ממה שיש לי בראש. (למשל, "רציונליות" צריכה להיות עקבית תחת רפלקציה – התבוננות "רציונלית" בראיות, ושקילה "רציונלית" של איך המוח שלך מעבד את הראיות, לא אמורות להוביל לשתי מסקנות שונות.)

באופן דומה, אם אתה מוצא את עצמך אומר: "הדבר הרציונלי עבורי לעשות הוא א, אבל הדבר הנכון עבורי לעשות הוא ב", אז אתה כמעט בוודאות משתמש במשמעות אחרת כלשהי למילה "רציונלי" או למילה "נכון". אני משתמש במונח "רציונליות" באופן נורמטיבי, כדי לבחור את דפוסי המחשבה הרצויים.

במקרה זה – או בכל מקרה אחר בו אנשים חלוקים לגבי משמעויות מילים – עליכם להשתמש בלשון ספציפית יותר במקום לומר "רציונלי": "הדבר האגוצנטרי לעשות הוא לברוח, אבל אני מקווה שלפחות אנסה לגרור את הילד מפסי הרכבת", או "תורת ההחלטות הסיבתית כפי שהיא מנוסחת בדרך כלל אומרת שעליך לבחור בשתי הקופסאות ב'בעיית ניוקום', אבל אני הייתי בוחר במיליון דולר".

למעשה, אני ממליץ לקרוא את המאמר הזה מההתחלה, ולהחליף כל הופעה של המילה "רציונלי" במילה "פונְּלִי", ולראות אם זה משנה במשהו את הקונוטציות של מה שאני אומר. אם זה כך, אני אומר: אל תשאפו לרציונליות, אלא לפונְּלִיות.

למילה "רציונלי" יש מלכודות פוטנציאליות, אבל ישנם שפע של מקרים שאינם מקרי קצה בהם "רציונלי" עובד מצוין כדי לתקשר את מה שאני חותר אליו, כמו גם "אירציונלי". במקרים אלו אני לא חושש להשתמש בה.

ועדיין, צריך להיזהר לא להשתמש במילה הזו יותר מדי. לא מקבלים נקודות רק על הגייתה בקול רם. אם תדבר יותר מדי על הדרך, לעולם לא תלך בה.

2

להרגיש רציונלי

ישנה אמונה רווחת הגורסת כי "רציונליות" מטבעה מנוגדת לכל רגש – שכל העצב וכל השמחה שלנו הם בהגדרה אנטי־לוגיים מעצם היותם רגשות. עם זאת, באופן מוזר למדי, לא מצאתי אף משפט בתורת ההסתברות שמוכיח כי עלי להיות קר כקרח וחסר הבעה. אז האם רציונליות מנוגדת לרגש? לא. הרגשות שלנו נובעים מתפיסת המציאות שלנו. אם אני מאמין שאחי המת נמצא חי, כמובן שאשמח. אם אתעורר ואגלה שהיה זה רק חלום, אהיה עצוב. פ. כ. הודג'ל אמרה: "מה שיש ביכולתה של האמת להשמיד, ראוי שיושמד." השמחה שחשתי בחלום הייתה מנוגדת לאמת. העצב שחשתי ביקיצה הוא רציונלי. אין שום אמת שתוכל להשמיד אותו. רציונליות מתחילה בשאלה איך־העולם־הינו, אבל מתפשטת באופן ויראלי לכל מחשבה אחרת התלויה באופן בו אנו חושבים שהעולם נראה. האמונות שלך לגבי "איך־העולם־הינו" נוגעות לכל דבר שאתה תופס כנוגע למציאות, כל דבר שקיים או לא קיים, כל חבר ברשימה של "דברים שיכולים לגרום לדברים אחרים לקרות." אם אתה מאמין שיש גמד בארון שלך הקושר את שרוכי הנעליים שלך אחד לשני, אז זו אמונה לגבי איך־העולם־הינו. הנעליים שלך הן אמיתיות – אתה יכול להרים אותן. אם יש משהו שם בחוץ שיכול להושיט יד ולקשור את השרוכים שלך זה לזה, הוא חייב להיות אמיתי בעצמו, חלק מאותה רשת עצומה של סיבות ותוצאות לה אנו קוראים

“היקום”.

לכעוס על הגמד שקשר לך את השרוכים זו פעולה המערבת מצב תודעתי שאינו רק לגבי איך־העולם־הינו. נניח שבתור בודהיסט, או מטופל כרות אונה, או סתם אדם אדיש מאוד, מציאת הנעליים שלך כשהשרוכים קשורים אחד לשני לא הכעיסה אותך. זה לא היה משפיע על הציפיות שלך מהעולם – עדיין היית מצפה לפתוח את הארון ולמצוא את השרוכים קשורים. אין סיבה שהכעס או הרוגע שלך ישפיעו על הניחוש הטוב ביותר שלך לגבי המציאות, מכיוון שמה שקורה בתוך הארון שלך לא תלוי במצב הרגשי שלך, אם כי אולי יידרש מאמץ כלשהו על מנת לחשוב בצלילות כזו.

אבל תחושת הכעס כשלעצמה קשורה למצב תודעתי שכן עוסק באיך־העולם־הינו: אתה נהיה כועס בגלל שאתה חושב שהגמד קשר את השרוכים שלך. הקריטריון לרציונליות מתפשט באופן ויראלי, מהשאלה הראשונית האם גמד קשר את השרוכים שלך או לא, אל הכעס הנובע מכך.

השיטה הרציונלית, הניסיון להשיג הערכות טובות יותר לגבי איך־העולם־הינו – יכולה לעמעם רגשות או להעצים אותם. לעיתים אנו בורחים מרגשות חזקים על ידי הכחשת העובדות, על ידי הסתת מבטנו ממראה העולם שהוליד רגש כה חזק. מכאן נובע שככל שנלמד את מיומנויות הרציונליות ונאמן את עצמנו להישיר מבט לעובדות, כך נחוש את רגשותינו בעוצמה רבה יותר.

בצעירותי תמיד תהיתי האם זה בסדר להרגיש דברים בעוצמה – האם זה מותר, האם זה ראוי. הסיבה לכך לא נעוצה, לדעתי, רק בהבנה הבוסרית שהיתה לי לגבי רציונליות. זיהיתי התלבטויות דומות בקרב אנשים שכלל לא שאפו להיות רציונליים: כשהם שמחים, הם תוהים אם מותר להם לשמוח, וכשהם עצובים, הם תמיד לא לגמרי בטוחים אם לברוח מהרגש או לא. מאז ימיו של סוקרטס, וכנראה הרבה לפני כן, הדרך הבטוחה להיראות מתורבת ומתוחכם הייתה לעולם לא להראות שבאמת אכפת לך ממהו. להרגיש זה מביך – זה פשוט משהו שלא עושים בחברה מנומסת. אתם צריכים לראות את המבטים

שאני מקבל כשאנשים מבינים כמה אכפת לי מרציונליות. זה לא הנושא החריג, לדעתי, אלא שהם פשוט לא רגילים לדאות מבוגרים שפויים שאכפת להם ממהו באופן גלוי כל כך.

אבל אני יודע עכשיו, שאין שום דבר פסול בלחוות רגשות בעוצמה. מאז אימצתי את הכלל של "מה שיש ביכולתה של האמת להשמיד, ראוי שיושמד", הגעתי גם להבנה ש"מה שיש ביכולתה של האמת להצמיח, ראוי שיצמח". כשמשוהו טוב קורה, אני שמח, ואין בי כל בלבול לגבי השאלה האם זה רציונלי עבורי להיות שמח. כשמשוהו נורא קורה, אינני בורח מהעצב בעזרת חיפוש אחר נחמות מזויפות ונקודות אור מדומיינות. אני רואה בעיני רוחי את העבר של האנושות ואת עתידה, את עשרות מיליארדי מקרי המוות לאורך ההיסטוריה, את האומללות ואת הפחד, את החיפוש אחר תשובות, את הידיים הרועדות המושטות מעלה מתוך כל הדם, את מה שאולי נהיה יום אחד כשהכוכבים יהיו לעֲרִינו, את כל החשיכה הזו ואת כל האור – אני יודע שלעולם לא אוכל להבין זאת באמת, ואין בפי מילים שאוכל לומר. חרף כל הפילוסופיה שלי אני עדיין נבוך להודות ברגשות עזים, ולכם כנראה לא נעים לשמוע אותם. אבל זאת אני יודע: זה רציונלי להרגיש.

3

למה לי לדעת את האמת? ...ו

חלק מהתגובות על *Overcoming Bias** עסקו בשאלה של מדוע עלינו לחפש את האמת בכלל. (למרבה המזל, לא רבים שאלו מהי בכלל אמת.) הסיבה שבגינה אנו רודפים אחר האמת (תהא אשר תהא אותה סיבה) היא משמעותית, שכן היא קובעת את המוטיבציה שלנו לעיצוב מחשבתנראנו. עיצוב אשר יקבע אם תצורת מחשבה כלשהי היא "טובה" או "רעה".

נאמר כי "סקרנות היא הראשונה במעלות הנפש." סקרנות היא סיבה אחת לחפש אמת. היא אולי לא הסיבה היחידה, אך יש בה טוהר מיוחד ומעורר הערצה. אם אתה מחפש את האמת מתוך סקרנות, אתה תתעדף שאלות על פי הרמה בה הן מגרות את החוש האסתטי האישי שלך. במקרה כזה, אתה עשוי לבחור בשאלות המהוות אתגר מתעתע יותר, עם סיכוי גבוה יותר לכישלון על פני שאלות המהוות אתגרים פשוטים יותר, רק לטובת ההנאה שבדבר.

כפי שצינתי לעיל, אנשים חושבים על רציונליות ועל רגש בתור יריבים לעיתים קרובות. מכיוון שסקרנות היא רגש, אני חושד שהתייחסות לסקרנות כחלק מהרציונליות תעורר התנגדות אצל חלק מהאנשים. לכשעצמי, רגש נחשב עבורי כ"לא רציונלי" אם הוא נשען על אמונות שגויות, או ליתר דיוק, על התנהלות הכרתית המייצרת

* הבלוג בו פורסם הפרק הראשון.

טעויות: "אם הברזל מתקרב אל פניך, ואתה מאמין שהוא לוחט, אך למעשה הוא קר, הפחד שלך סותר את הדרך. אם הברזל מתקרב אל פניך, ואתה מאמין שהוא קר, אך למעשה הוא לוחט, הרוגע שלך סותר את הדרך." לעומת זאת, רגש המתעורר על ידי אמונות נכונות או חשיבה רציונלית הכרתית הוא "רגש רציונלי". התועלת בכך היא שרגש כזה מאפשר לנו להכיר ב-דוגע כמצב רגשי, ולא סתם ברירת מחדל מנטלית.

כשאנשים חושבים על "רגש" ועל "רציונליות" כערכים מנוגדים, אני חושד שבעצם הם חושבים על מערכת 1 ומערכת 2 – שיפוטים תפיסתיים מהירים לעומת שיפוטים מודעים ואיטיים. שיפוטים שקולים אינם תמיד נכונים, ושיפוטים תפיסתיים אינם תמיד שגויים. מסיבה זו חשוב מאוד להבחין בין הדיכוטומיה הזו לבין "רציונליות". שתי המערכות האלה יכולות לשרת את האמת או להביס אותה, תלוי באופן השימוש בהן.

מלבד סקרנות רגשית טהורה, אילו עוד מניעים קיימים לתשוקה לאמת? ובכן, ייתכן שתרצה להשיג מטרה כלשהי בעולם האמיתי, כמו למשל לבנות מטוס, ולשם כך אתה צריך להכיר אמת מסוימת שקשורה לאווירודינמיקה. או באופן יומיומי יותר: אתה רוצה שוקן, ולכן אתה רוצה לדעת אם במכולת השכונתית יש שוקן, כדי שתוכל לבחור אם ללכת לשם או למקום אחר. אם זו הסיבה שאתה רוצה את האמת, אז העדיפות שתקצה לשאלות שלך תשקף את התועלת הצפויה של המידע שלהן עבורך – עד כמה התשובות האפשריות ישפיעו על הבחירות שלך, עד כמה הבחירות שלך חשובות ועד כמה אתה מצפה למצוא תשובה שתגרום לך לשנות את הבחירה שלך ביחס לברירת המחדל.

חיפוש האמת עבור הערך המעשי שלה בלבד עשוי להיראות לא "טהור" – האם לא מוטל עלינו לתור אחר האמת לשם האמת? – העניין הוא שחקירות כאלו הן חשובות ביותר, מכיוון שהן יוצרות קריטריון חיצוני לאישוש השערות: אם המטוס שלך נופל מהשמיים, או אם אתה מגיע לחנות ולא מוצא שוקן, יהיה זה רמז לכך שעשית משהו

לא נכון. אתה מקבל משוב לגבי אילו דפוסי חשיבה עובדים, ואילו לא. סקרנות טהורה היא דבר נפלא, אך היא עשויה שלא להתעכב זמן רב מדי על אימות התשובות שלה, ברגע שהמסתורין המרגש בא על סיפוקו. סקרנות כרגש אנושי התקיימה הרבה לפני היוונים הקדמונים, אבל מה שהציב את האנושות על נתיב המדע בבטחה היתה ההבחנה בכך שדפוסי חשיבה מסוימים חשפו אמונות שאפשרו לנו להשפיע על העולם. בכל הנוגע לסקרנות גרידא, טוויית מעשיות סביב המדורה על אלים וגיבורים סיפקה את התשוקה הזו באותה מידה, ואף אחד לא חשב שמהו בזה לא בסדר.

האם יש מניעים נוספים לחיפוש האמת מלבד סקרנות ופרגמטיזם? הסיבה השלישית שאני יכול לחשוב עליה היא מוסריות: אתה מאמין שחיפוש אחר האמת הוא דבר אצילי וחשוב וראוי.* למרות שאידיאל כזה מייחס גם הוא ערך פנימי לאמת, זהו מצב תודעתי שונה מאוד מסקרנות. להיות סקרן לגבי מה שנמצא מאחורי הוילון זה דבר אחד. להאמין שיש לך חובה מוסרית להסיט אותו ולהישיר מבט זה דבר אחר לחלוטין. במצב התודעה השני, סביר הרבה יותר שתאמין שגם אנשים אחרים צריכים להסתכל מאחורי הוילון, או שתוקיע אותם אם יבחרו לעצום את עיניהם במכוון. מסיבה זו, הייתי מתייג כ"מוסרי" גם את האמונה שחיפוש האמת הוא חשוב באופן פרגמטי לחברה, ולכן הוא מוטל כחובה על הכלל. סדרי העדיפויות שלך, תחת מוטיבציה זו, ייקבעו על פי האידיאלים שלך לגבי אילו אמיתות הן החשובות ביותר (ולא השימושיות ביותר או המסקרנות ביותר), או לגבי מתי, ובאילו נסיבות, החובה לחפש את האמת היא החזקה ביותר.

אני נוטה לחשוד במוסר כמוטיבציה לרציונליות, לא מפני שאני דוחה את האידיאל המוסרי, אלא מפני שהוא מזמין סוגים מסוימים של צרות. קל מדי לאמץ, כחובות מוסריות נרכשות, דפוסי חשיבה שהם זיופים צורמים במנגינה. חיטבו על מיסטר ספוק מ"מסע בין כוכבים", ארכיטיפ נאיבי של רציונליות. המצב הרגשי של ספוק

* "אתה לא רייבנקלו," אמר הארי בטון מכבד ורגוע, "ולכן ייתכן שלא חשבת שלכבד את האמת, ולחפש אותה בכל ימי חייך, יכול גם להיות מעשה אצילי."

מכוון תמיד ל"רגוע", גם כשזה ממש לא מותאם. לעיתים קרובות הוא מוסיף ספרות עשרוניות להסתברויות שאינן מכילות בעליל. (למשל, "קפטן, אם תנווט את האנטרפרייז ישירות לתוך החור השחור הזה, ההסתברות שלנו לשרוד היא רק 2.234%". ובכל זאת תשע מתוך עשר פעמים האנטרפרייז אינה מושמדת. איזה מין דביל טראגי נותן שלוש ספרות אחרי הנקודה למספר ששגוי בשני סדרי גודל?) ועדיין, הדימוי הפופולרי הזה הוא האופן שבו אנשים רבים תופסים את החובה להיות "רציונליים" – ממש פלא שהם לא מאמצים אותה בלב שלם. להפוך את הרציונליות לחובה מוסרית פירושו לתת לה את כל דרגות החופש האיומות של מנהג שבטי שרירותי. אנשים מגיעים לתשובה הלא נכונה, ואז מוחים בזעם שהם פעלו כנדרש, במקום ללמוד מהטעות של עצמם.

למרות כל זאת, אם ברצוננו לשפר את המיומנויות הרציונליות שלנו וללכת מעבר לסטנדרטים הביצועיים שהציבו ציידים-לקטים, נזדקק לעיצוב מודע של אמונות לגבי איך לחשוב כשורה. כשאנו כותבים לעצמנו תוכנות מנטליות חדשות, הן מתחילות במערכת 2, המערכת המחושבת, ורק לאט לאט – אם בכלל – מאומנות לתוך המעגלים העצביים העומדים בבסיס מערכת 1. אז אם יש סוגים מסוימים של חשיבה שנראה שהיינו רוצים להימנע מהם – כמו הטיות, למשל – התהליך הזה ייוצג בסופו של דבר בתוך מערכת 2, כהוראה מפורשת שלא לחשוב בדרך זו. חובת הימנעות מוצהרת.

אם האמת היא משאת נפשנו, נוכל להשיגה באופן היעיל ביותר על ידי חשיבה בדרכים מסוימות, ולא באחרות. אלו הן הטכניקות של הרציונליות. וכמה מהטכניקות של הרציונליות כרוכות בהתגברות על סוג מסוים של מכשולים: ההטיות...

4

...תזכיר לי מה זו הטיה?

הטיה היא סוג מסוים של מכשול אשר מונח בדרכנו אל האמת. (הגדרתה כ"מכשול" נובעת מהמטרה שלנו – להגיע לאמת.) עם זאת, לא כל מכשול הוא "הטיה".

אם נתחיל מיד בשאלה "מהי הטיה?", נימצא ניגשים אל הבעיה בסדר הלא נכון. כמאמר הפתגם, "יש ארבעים סוגים של טירוף, אבל רק סוג אחד של שכל ישר". האמת היא מטרה צרה, אזור קטן מאוד במרחב האפשרויות. "אוהבת, לא אוהבת" זו אולי שאלה בינארית, אבל $E = mc^2$ היא נקודה זעירה במרחב של כל המשוואות האפשריות, כמו כרטיס זוכה בלוטו במרחב של כל כרטיסי הלוטו בעולם. טעות היא לא מצב חריג. דווקא ההצלחה היא זו שדורשת הסבר, שכן סבירותה נמוכה להפליא מלכתחילה.

אנחנו לא מתחילים את התהליך עם חובה מוסרית של "צמצום הטיות", מתוך מחשבה שהטיות הן רעות ונוראות ודבר שלא עושים. זו דרך חשיבה שמאפיינת את מי שספג מהחברה תחושת חובה דאונטולוגית להיות "רציונלי", גישה שמובילה אנשים לנסות להשתמש בשיטות הרציונליות מבלי להבין את ההיגיון מאחוריהן. (וזה רע ונורא ודבר שלא עושים – לפי אתה בטח מתלוצץ, מיסטר פיינמן! שקראתי בתור ילד.)

יהיה נכון יותר לומר שאנחנו רוצים פשוט להגיע לאמת, מכל סיבה שהיא, ונתקלים במכשולים בדרך. המכשולים האלה אינם

שונים זה מזה לחלוטין. לדוגמה, חלקם נובעים ממחסור בכוח חישוב או מעלות גבוהה של מידע. עם זאת, ישנה קבוצה גדולה של מכשולים הניראים כחולקים תכונה משותפת כלשהי – הם מתקבצים באותו אזור במרחב-המכשולים-לאמת – וקבוצה זו קיבלה את התווית "הטיות".

אז מהי הטיה? האם אפשר להתבונן בקבוצה הזו ולמצוא מבחן קצר וקולע לחברות בה? ייתכן שלא נוכל להציע הגדרה טובה יותר מאשר להצביע על מספר דוגמאות קונקרטיות ולקוות שהקורא יבין. אם אתה מדען שמתחיל לחקור אש, יהיה זה חכם יותר להצביע על מדורה ולומר "אש זה הדבר החם והכתום-בהיר הזה שם", מאשר להכריז: "אני מגדיר אש כהתמרה אלכימית של חומרים המשחררת פְּלוֹגְיִסְטוֹן".* אין סיבה להתעלם ממשוהו רק משום שאיננו יכולים להגדיר אותו. אני לא יודע לצטט בעל פה את משוואות היחסות הכללית, אבל אם אקפוץ מצוק – בוודאי אפול מטה. נוכל לומר את אותו הדבר גם לגבי הטיות – הן לא יפגעו בנו פחות חזק רק משום שלא הצלחנו להגדיר אותן היטב. אז אולי נוכל להצביע על כשל הצירוף, ביטחון יתר, היוריסטיקת הזמינות והייצוגיות, הזנחת שיעורי בסיס, ולומר: "דברים מהסוג הזה."

על אף כל זאת, אנחנו נוטים לכנות בשם "הטיות" את אותם מכשולים לאמת שמקורם הוא לא מחיר המידע ולא מגבלות החישוב, אלא צורתו של המנגנון הקוגניטיבי שלנו. ייתכן שהמנגנון הזה התפתח אבולוציונית למטרות מסוימות הסותרות דיוק הכרתי – לדוגמה, נצחון בוויכוחים בהקשרים פוליטיים משתנים. או שאולי הלחץ האבולוציוני פעל בכלל בכיוון אחר – כמו לדחוף אותנו להאמין במה שמאמינים אחרים על מנת להשתלב חברתית. לעיתים, כמו במקרה הקלאסי של היוריסטיקות והטיות, מדובר באלגוריתם ברור של פעולה, אשר מועיל במידה מסוימת אך מוביל גם לטעויות שיטתיות: היוריסטיקת הזמינות עצמה אינה בגדר הטיה, אך היא

* ראו פרק 33.

יוצרת הטיות ברורות שקל מאוד לזהות. המוח שלנו עושה משהו שגוי, ולאחר הרבה ניסויים או חשיבה מאומצת, מישוהו מצליח לזהות את מקור השגיאה באופן ש"מערכת 2" מסוגלת להבין – ואז אנחנו מתייגים זאת כ"הטיה". גם אם זה כל אשר נוכל לדעת, עדיין ברור כי מדובר בכשל כלשהו שאינו נובע ממחסור בכוח חישוב, אלא ממכונת החישוב עצמה.

על פי הגדרתן, "הטיות" הן שונות משגיאות שמקורן בתוכן המחשבתי – כמו אמונות שאימצנו או חובות מוסריות שלקחנו על עצמנו. אלו מכונות "טעויות" ולא "הטיות", וברגע שמזהים אותן, הן קלות יותר לתיקון. (אם כי המקור לטעות – או המקור למקור לטעות – עשוי להיות הטיה.)

"הטיות" מובחנות גם משגיאות שמקורן בנזק מוחי, או בנורמות תרבותיות מסוימות – שכן הטיות נובעות ממנגנון קוגניטיבי המשותף לכל בני האדם.

אפלטון לא היה "מוטה" משום שלא הכיר את תורת היחסות – לא הייתה לו שום דרך להשיג את המידע הזה, ובורותו לא נבעה מצורת מנגנון החשיבה שלו. מצד שני, אם אפלטון האמין שפילוסופים יהיו מלכים טובים יותר משום שהוא עצמו היה פילוסוף, ואמונה זו נבעה מאינסטינקט אנושי אוניברסלי לקידום עצמי בהקשרים פוליטיים, ולא משום שאביו של אפלטון לימד אותו שחובתו המוסרית של כל אדם היא לקדם את מקצועו לעמדת שלטון, ולא משום שאפלטון הסניף יותר מדי דבק בילדותו – אז הייתה זו הטיה, גם אם איש לא הזהיר אותו מפניה.

הטיות לא תמיד פשוטות לתיקון. ייתכן שהן אפילו לא ברות-תיקון. אך כאשר אנו מתבוננים במנגנון החשיבה שלנו ורואים בו הסבר סיבתי לקבוצה ברורה של שגיאות, וכאשר מקור הבעיה הוא בצורתו של המנגנון האבולוציוני, ולא במחסור בכוח חישוב או בתוכן שגוי – אנו קוראים לזה הטיה.

כשלעצמי, אני רואה את המסע שלנו במונחים של רכישת כישורים אישיים של רציונליות, של שיפור השיטות שלנו למציאת אמת.

האתגר הוא להגיע לאמת על דרך החיוב, ולא רק להימנע מהכשלון על דרך השלילה. מרחב הכישלון הוא רחב, ומכיל אינספור טעויות באינספור צורות. קשה לתאר מרחב כה עצום: "מה שנכון לגבי תפוח אחד לא נכון בהכרח לגבי כל התפוחים. לפיכך יש יותר מה לומר על תפוח בודד מאשר על כל התפוחים שבעולם." מרחב ההצלחה צר יותר – ולכן ניתן לומר עליו יותר.

אף שאינני מתנגד (כפי שניתן לראות) לדון בהרחבה בהגדרות, עלינו לזכור שאין זה היעד העיקרי שלנו. אנו כאן כדי להמשיך את המסע האנושי הגדול במרדף אחר האמת: שכן אנו זקוקים נואשות לידע, וחוץ מזה – אנחנו סקרנים. ולמטרה זו – הבה נחתור להתגבר על כל מכשול שיעמוד בדרכנו, בין אם נקרא לו "הטיה" ובין אם לאו.

5

זמינות

היוריסטיקת הזמינות היא האופן בו אנו מעריכים את התדירות או ההסתברות של אירוע על פי הקלות בה אנו מצליחים להיזכר באירועים דומים.

מחקר מפורסם משנת 1978 שנערך על ידי ליכטנשטיין, סלוביק, פישוהוף, ליימן וקומבס, בשם "תדירות מוערכת של אירועים קטלניים", חקר טעויות בקביעת חומרתם של סיכונים, או בניסיון להעריך איזו מבין שתי סכנות מתרחשת בתדירות גבוהה יותר.¹ המחקר מצא כי נבדקים חשבו שתאונות דרכים גורמות למספר דומה של מקרי מוות כמו מחלות, ושרצח הוא סיבת מוות שכיחה יותר מאשר התאבדות. האמת היא שמחלות גורמות לפי שישה-עשר יותר מקרי מוות מאשר תאונות, והתאבדות נפוצה פי שניים יותר מרצח.

השערה מתבקשת לעיוות-האמונות הזוהי היא שסביר יותר שידברו על מקרי רצח מאשר על התאבדויות – ולכן סביר יותר שמישהו יזכור ששמע על רצח מאשר ששמע על התאבדות. תאונות הן דרמטיות יותר ממחלות – אולי זו הסיבה שלאנשים קל יותר לזכור תאונות, ולהיזכר בהן, כשנשאלים עליהן. ב-1979, מחקר המשך של קומבס וסלוביק הראה שקביעות ההסתברות המוטות הראו מתאם חזק (0.85 ו-0.89) עם תדירויות דיווח מוטות בשני עיתונים.² זה לא פותר את השאלה של "האם מקרי רצח זמינים יותר לזיכרון כיוון שמדווחים עליהם יותר, או שהעיתונים מדווחים יותר על מקרי

רצח כיוון שהם דרמטיים יותר (ולכן גם זכורים יותר)?” כך או כך, הטיית הזמינות פועלת כאן. דיווח סלקטיבי הוא מקור עיקרי להטיית הזמינות. בסביבה הקדמונית, הרבה ממה שידעת – חווית בעצמך, או ששמעת זאת ישירות מחבר שבט שראה זאת. בדרך כלל היתה לכל היותר שכבה אחת של דיווח סלקטיבי בינך לבין האירוע עצמו. עם האינטרנט של היום, אתה עשוי לקרוא דיווחים שעברו דרך שישה בלוגרים בדרך אליך – שישה מסננים רצופים. בהשוואה לאבותינו, אנו חיים בעולם גדול יותר, שקורה בו הרבה יותר, והרבה פחות ממנו מגיע אלינו – אפקט סלקציה חזק הרבה יותר, שיכול ליצור הטיות זמינות גדולות בהרבה.

בעולם האמיתי, לא סביר שתפגשו את ביל גייטס אי פעם, אך הודות לדיווח סלקטיבי של המדיה, אתם עשויים להתפתות להשוות את הצלחת חייכם לזו שלו – ולשלם קנסות ברמת האושר שלכם בהתאם. התדירות האובייקטיבית של ביל גייטס היא 0.00000000015, אך אתם שומעים עליו לעיתים קרובות בהרבה. לעומת זאת, 19% מהאנשים על פני הכדור חיים על פחות מדולר ליום, ואני מסופק אם חמישית מהפוסטים בבלוגים שאתם קוראים נכתבו על ידי אנשים.

השימוש ב"זמינות" נראה כמעורר את הטיית האבסורד : אירועים שמעולם לא קרו אינם נשלפים מהזיכרון, ולכן נחשבים כבעלי הסתברות אפס. בזמנים בהם לא התרחש שיטפון לאחרונה (ועדיין ההסתברויות בהחלט ניתנות לחישוב), אנשים מסרבים לקנות ביטוח שיטפונות גם כשהוא מסובסד עמוקות ומתומחר הרבה מתחת למחיר הריאלי מבחינה אקטוארית. קונרדו יותר ואחרים מציעים שתגובת-חסר לאיומי שיטפונות עשויה לנבוע מ"חוסר היכולת של פרטים להמשיג שיטפונות שמעולם לא קרו... נראה שאנשים החיים במישורי הצפה הם במידה רבה אסירים של ניסיונם-שלהם... נראה ששיטפונות שאירעו לאחרונה מציבים חסם עליון לגודל ההפסד שמנהלים מאמינים שראוי להיות מוטרדים לגביו."³

ברטון ואחרים מצאו כי כאשר נבנים סכרים וסוללות, הם אמנם

מפחיתים את תדירות השיטפונות, אך יוצרים מצד שני תחושת ביטחון כוזבת, המובילה לצמצום באמצעי הזהירות.⁴ אך בעוד שבניית סכרים מפחיתה את תדירותם של השיטפונות, היא גורמת לנזק ככל שיטפון להיות גדול כל כך, עד שהנזק השנתי הממוצע גדל ביחס למצב הקודם. האדם הנבון יקיש מזכרון של סכנות קטנות אל האפשרות של סכנות גדולות. במקום זאת, נראה שניסיון העבר בסכנות קטנות מציב חסם עליון על עוצמת הסכנה שאנו מסוגלים לתפוס. חברה המוגנת היטב מפני סכנות מינוריות אינה נוקטת פעולה כנגד סיכונים גדולים, ובונה בשטחים שבסכנת הצפה ברגע שהשיטפונות המינוריים הרגילים חוסלו. חברה הנתונה לסכנות מינוריות סדירות מתייחסת לאותן סכנות מינוריות כאל חסם עליון לגבי עוצמתם של הסיכונים, ומתגוננת מפני שיטפונות מינוריים סדירים אך לא מפני שיטפונות גדולים נדירים.

הזיכרון אינו תמיד מדריך טוב לחישוב הסתברויות בעבר, לא כל שכן בעתיד.

1. Sarah Lichtenstein et al., "Judged Frequency of Lethal Events," *Journal of Experimental Psychology: Human Learning and Memory* 4, no. 6 (1978): 551-578, doi:10.1037/0278-7393.4.6.551.

2. Barbara Combs and Paul Slovic, "Newspaper Coverage of Causes of Death," *Journalism & Mass Communication Quarterly* 56, no. 4 (1979): 837-849, doi:10.1177/107769907905600420.

3. Howard Kunreuther, Robin Hogarth, and Jacqueline Meszaros, "Insurer Ambiguity and Market Failure," *Journal of Risk and Uncertainty* 7 (1 1993): 71-87, doi:10.1007/BF01065315.

4. Ian Burton, Robert W. Kates, and Gilbert F. White, *The Environment as Hazard*, 1st ed. (New York: Oxford University Press, 1978).

6

פרטים מכבידים

“פרט מאמת, שנועד להעניק מראית-עין אמנותית לנרטיב עירום
וחסר-שכנוע...”

– פּוֹרְבָּה, ב"המיקאדו" של גילברט וסאליבן.¹

כשל החיתוך מתרחש כאשר בני אדם מעריכים את ההסתברות $P(A,B)$ כגבוהה יותר מההסתברות $P(B)$, למרות שהמשפט המתמטי קובע ש- $P(A,B) \leq P(B)$. לדוגמה, בניסוי משנת 1981, 68% מהנבדקים דירגו כסביר יותר ש"רייגן יספק תמיכה פדרלית לאמהות לא נשואות ויקצץ בתמיכה הפדרלית לשלטון המקומי" מאשר ש"רייגן יספק תמיכה פדרלית לאמהות לא נשואות".

סדרה ארוכה של ניסויים מתוכננים בחוכמה, אשר ניפו השערות חלופיות וביססו את הפרשנות הסטנדרטית, איששה את הטענה שכשל החיתוך מתרחש מכיוון שאנו "מחליפים הערכה של ייצוגיות בהערכה של הסתברות". על ידי הוספת פרטים, ניתן לגרום לתוצאה להישמע סבירה יותר מאשר התהליך שיצר אותה. אפשר לגרום לכך ש"רייגן יתמוך באמהות לא נשואות" יישמע תרחיש סביר יותר על ידי הוספת הטענה שבנוסף, רייגן גם "יקצץ בתמיכה לשלטון המקומי". הסבירות שבטענה אחת מפצה על חוסר הסבירות שבטענה האחרת.

הן "מאזנות זו את זו".

רוצה לומר: הוספת פרטים יכולה לגרום לתרחיש להישמע סביר יותר, למרות שהאירוע הופך לסביר פחות בהכרח.

במקרה כזה אנו עשויים למצוא, באופן היפותטי, עתידנים שטוויים היסטוריות עתידיות הגיוניות ומפורטות באופן בלתי מתקבל על הדעת, או למצוא אנשים שבולעים ערימות ענקיות של טענות לא מבוססות הארוזות סביב כמה קביעות שנשמעות מבוססות במרכזן. אם כשל החיתוך מוצג בפניך בהשוואה ישירה ועירומה, ייתכן שתצליח להתגבר על הבעיה הספציפית הזו בכך שתתקן את עצמך באופן מודע. אך זה יהיה בבחינת הנחת פלסטר על הבעיה, ולא תיקון יסודי שלה באופן כללי.

בניסוי אחר, משנת 1982, מספקייתחזיות מקצועיים הקצו באופן שיטתי הסתברויות גבוהות יותר ל"רוסיה פולשת לפולין, ולאחר מכן ניתוק היחסים הדיפלומטיים בין ארה"ב לבריה"מ" מאשר ל"ניתוק היחסים הדיפלומטיים בין ארה"ב לבריה"מ", כאשר לכל קבוצת ניסוי הוצג רק אחד מההיגדים.² באיזו אסטרטגיה שהייתה מחסלת את כשל החיתוך יכלו החוזים הללו לנקוט, כאשר אף אחד מהם לא ידע על ההשוואה באופן ישיר? כאשר איש מהם אפילו לא ידע שהניסוי עוסק בכשל החיתוך? כיצד הם יכלו לבצע הערכות הסתברות טובות יותר? התגברות על מלכודת אחת כמקרה פרטי אינה מתקנת את הבעיה הכללית. המלכודת היא הסימפטום, לא המחלה עצמה.

מה יכלו החוזים לעשות כדי להימנע מכשל החיתוך, מבלי לראות את ההשוואה באופן ישיר, או מבלי לדעת אפילו שמישהו בוחן אותם על כשל החיתוך? נראה לי שהם היו צריכים להבחין במילית "ו-" (and). היא היתה אמורה להדליק אצלם נורה אדומה – לא רק נורה, הם היו צריכים לזנק לאחור בבהלה. אפילו בלי לדעת שחוקרים עומדים לבחון אותם ספציפית על כשל החיתוך. הם היו צריכים להבחין בחיתוך של שני אירועים שלמים, ולהיות מזועזעים מהתעוזה של מי שמבקש מהם לספק ניבוי מסובך באופן מטורף כל כך. והם היו צריכים "לקנוס" את ההסתברות להתממשות התרחיש

באופן משמעותי – בפקטור של ארבע לפחות, על פי פרטי הניסוי. דבר נוסף שהיה עשוי לעזור להם, היה לנסות לחשוב על סיבות אפשריות לכך שארה"ב וברה"מ ישהו יחסים דיפלומטיים. התרחישי אינו "ארה"ב וברה"מ מנתקות לפתע פתאום יחסים דיפלומטיים ללא כל סיבה", אלא "ארה"ב וברה"מ מנתקות יחסים דיפלומטיים מסיבה כלשהי".

ומה לגבי הנבדקים שדירגו גבוה יותר את "רייגן יספק תמיכה פדרלית לאמהות לא נשואות ויקצץ בתמיכה לשלטון המקומי"? גם במקרה הזה, הם היו אמורים להודעזע מהמילית "ו-". יתרה מכך, הם היו צריכים לחבר אבסורדים – כאשר האבסורד הוא הלוגריתם של ההסתברויות, כך שניתן לחבר אותם, במקום לחשב ממוצע. הם היו צריכים לחשוב: "רייגן אולי יקצץ ואולי לא יקצץ תמיכה לשלטון המקומי (ביט אחד), אבל נראה מאוד לא סביר שהוא יתמוך באמהות לא נשואות (4 ביטים). סך האבסורד: 5 ביטים."* או אולי, "רייגן לא יתמוך באמהות לא נשואות. זה פשוט לא יקרה. הטענה השנייה רק הופכת את זה לפחות סביר".

באופן דומה, חשבו על קובייה בעלת שש פאות, עם ארבע פאות ירוקות (י') ושתי פאות אדומות (א'). הנבדקים היו צריכים להמר על אחד הרצפים הבאים: (1) איאאא, (2) יאיאאא, או (3) יאאאאא שיופיע בשלמותו במקום כלשהו במהלך עשרים הטלות רצופות.³ 65% מהנבדקים בחרו ב-יאיאאא, אפשרות שנשלטת חזק** על ידי איאאא, מכיוון שכל רצף המכיל את יאיאאא מכיל בהכרח גם את יאיאאא. מה יכלו הנבדקים לעשות על מנת להצליח יותר? על ידי כך ששימו לב להכלה של רצף אחד באחר? אולי, אבל שוב, זה רק פלסטר, לא תיקון של הבעיה היסודית. על ידי חישוב מפורש של ההסתברויות? זה בהחלט היה מתקן את הבעיה היסודית, אבל חישוב מדויק של ההסתברויות הוא לא תמיד אפשרי.

* להסבר אודות 'ביטים' מתמטיים ראו פרק 23.
** בתורת המשחקים, אסטרטגיה א' שולטת חזק על אסטרטגיה ב' אם התועלת מאסטרטגיה א' תמיד תהיה טובה יותר מאסטרטגיה ב'.

הנבדקים הפסידו מבחינה היוריסטית בכך שחשבו: "אהה! לרצף 2 יש את היחס הגבוה ביותר של ירוק לאדום! אני צריך להמר עליו!" כדי לנצח היוריסטית, הנבדקים היו צריכים לחשוב: "אהה! רצף 1 הוא קצר! אני צריך ללכת על רצף 1!"

"תערו של אוקאם" היה צריך לעורר אצלם אפקט רגשי חזק יותר – כזה שיגרום להם להרגיש כל פרט נוסף כנטל, אפילו הטלה אחת של קוביה.

לפני זמן רב שוחחתי עם מישהו שהוקסם על ידי עתידן לא זהיר (אחד כזה שמוסיף המון פרטים שנשמעים מגניבים). ניסיתי להסביר מדוע אני לא מוקסם כמוהו מהתיאוריות המדהימות והנהדרות האלה. אז הסברתי לו על כשל החיתוך, וספציפית על ניסוי "ניתוק יחסים ± פלישה לפולין". והוא אמר, "אוקיי, אבל מה זה קשור ל-" ואמרתי, "סביר יותר שיקומים משתכפלים מסיבה כלשהי, מאשר שהם משתכפלים באמצעות חורים שחורים כי תרבויות מתקדמות מייצרות חורים שחורים כי יקומים מתפתחים כדי לגרום להם לעשות את זה." והוא אמר, "אה."

עד אותו רגע, הוא לא הרגיש את הפרטים הנוספים הללו כנטל נוסף. במקום זאת הם היו פרט מאשש, ששיווה לנרטיב חזות של אמת. מישהו מציג בפניך חבילה של רעיונות מוזרים, שאחד מהם הוא שיקומים משתכפלים. ואז הוא מציג תמיכה לטענה שיקומים משתכפלים. אבל זו אינה תמיכה לחבילה, למרות שהכל מסופר כסיפור אחד.

עליך להפריד בין הפרטים. עליך להתבונן בכל אחד מהם בנפרד ולשאול, "איך אנחנו יודעים את הפרט הזה?" מישהו משרטט תמונה של הידרדרות האנושות ללוחמה ננוטכנולוגית, שבה סין מסרבת לציית להסכם פיקוח בינלאומי, מה שמוביל למירוץ חימוש... רגע אחד – מניין לך שזו תהיה סין? האם זה כדור בדולח בכיס שלך או שאתה פשוט שמח להיות עתידן? מאיפה כל הפרטים האלה מגיעים? מאיפה הפרט הספציפי הזה הגיע?

שהרי נכתב:

אם בכוחך להקל את העול שאתה נושא, עליך לעשות כן.
בכל קש טמון הכוח לשבור את גבך.

1. William S. Gilbert and Arthur Sullivan, *The Mikado*, Opera, 1885.
2. Tversky and Kahneman, "Extensional Versus Intuitive Reasoning."
3. Amos Tversky and Daniel Kahneman, "Judgments of and by Representativeness," in *Judgment Under Uncertainty: Heuristics and Biases*, ed. Daniel Kahneman, Paul Slovic, and Amos Tversky (New York: Cambridge University Press, 1982), 84-98.

7

כשל התכנון

נמל התעופה הבינלאומי של דנבר נפתח באיחור של 16 חודשים, עם חריגה תקציבית של 2 מיליארד דולר. (ראיתי גם טענות ל-3.1 מיליארד.) ה"יורופייטר טייפון", פרויקט הגנה משותף של מספר מדינות אירופיות, נמסר באיחור של 54 חודשים ובעלות של 19 מיליארד דולר במקום 7 מיליארד. בית האופרה של סידני הוא אולי פרוייקט הבנייה עם החריגה האגדית מכולן: תאריך הסיום המקורי תוכנן להיות ב-1963 עם עלות מוערכת של 7 מיליון דולר. הפרוייקט הושלם לבסוף ב-1973, בעלות של 102 מיליון דולר.¹

האם מדובר באסונות חריגים שהובאו לתשומת ליבנו על ידי זמינות סלקטיבית? האם הם תוצאה של בירוקרטיה או כשל בתמריצים ממשלתיים? כן, סביר מאוד להניח. אך ישנה גם הטיה קוגניטיבית מתאימה, ששוחזרה בניסויים עם מתכננים יחידים.

פְּיוֹלר ואחרים ביקשו מהסטודנטים הנבדקים במחקר לספק הערכות לגבי המועד שבו הם (הסטודנטים) מאמינים שיימו את הפרוייקטים האקדמיים שלהם.² ספציפית, החוקרים ביקשו מהסטודנטים לספק הערכות לתאריך סיום לפרוייקט שלהם בהסתברות של 50%, 75% ו 99%. תרצו לנחש כמה סטודנטים הצליחו לעמוד בהערכה שלהם עצמם?

• 13% מהנבדקים סיימו את הפרוייקט עד המועד שהוקצתה לו הסתברות של 50%.

- 19% סיימו עד המועד שהוקצתה לו הסתברות של 75%.
- ורק 45% (פחות מחצי!) סיימו עד המועד שהוקצתה לו הסתברות של 99%.

כפי שכתבו החוקרים: "התוצאות עבור רמת הסתברות של 99% הן בולטות במיוחד: אפילו כשהתבקשו לתת תחזית שמרנית ביותר, תחזית שהוודאות שלהם בקשר ליכולתם לעמוד בה היא כמעט מוחלטת, הביטחון של הסטודנטים בהערכות הזמנים שלהם עלה בהרבה על ההישגים שלהם בפועל."³

באופן כללי יותר, תופעה זו ידועה בשם "כשל התכנון". משמעו של כשל התכנון הוא שאנשים חושבים שהם מסוגלים לתכנן, חה חה. רמז לבעיה המונחת בשורש מנגנון התכנון שלנו נחשף על ידי ניופי-קלארק ואחרים, שמצאו כי:

- בקשת תחזיות מנבדקים המבוססות על תרחישים ריאליסטיים של "הניחוש הטוב ביותר שלך", ו-
- בקשת תחזיות מנבדקים המבוססות על "התרחיש הטוב ביותר" שניתן לקוות שיתרחש...
...הניבו תוצאות זהות לחלוטין.⁴

כשאנשים מתבקשים לתת תרחיש "ריאליסטי", הם מדמיינים שהכל הולך בדיוק כמתוכנן, ללא עיכובים בלתי צפויים או אסונות בלתי ניתנים לחיזוי – בדיוק אותו תרחיש כמו ב"תרחיש הטוב ביותר" שלהם.*

למציאות, לרוע המזל, יש נטיה לספק תוצאות גרועות מעט יותר מ"התרחיש הגרוע ביותר".

שלא כמו רוב ההטיות הקוגניטיביות, קיימת שיטה די טובה להתמודד עם כשל התכנון. היא לא תעבוד עבור כאוס בקנה מידה של נמל התעופה הבינלאומי של דנבר, אבל היא כן תעבוד עבור מקרים רבים של תכנונים אישיים, ואפילו עבור כמה עניינים ארגוניים בקנה מידה קטן. הרעיון הפשוט הוא להשתמש ב"מבט מבחוץ" במקום

* "הוא הולך להשתלט על העולם... דברים כאלה אכן דורשים לעיתים יותר מיוחדים." (הארי ג'יימס פוטר אוונס רוס)

ב"מבט מבפנים".

אנשים נוטים לחשב את התחזיות שלהם בעזרת חשיבה על המאפיינים הספציפיים והייחודיים של המשימה שעל הפרק, ובניית תרחיש לאופן שבו הם מתכוונים להשלים את המשימה – שזו, לרוב, פשוט ההגדרה של תכנון.

כשברצונך לבצע משהו, עליך לתכנן איפה, מתי, ואיך. עליך להבין כמה זמן ומהם המשאבים הנדרשים. עליך לדמיין את הצעדים מההתחלה ועד לסיום המוצלח. כל התהליך הזה הוא ה"מבט מבפנים", והוא לא לוקח בחשבון עיכובים בלתי צפויים ואסונות לא מתוכננים. כפי שראינו קודם, לבקש מאנשים לדמיין את "המקרה הגרוע ביותר" עדיין לא מספיק כדי לנטרל את האופטימיות שלהם – הם פשוט לא מדמיינים מספיק תקלות.

המבט מבחוץ הוא כשאתה נמנע במכוון מלחשוב על המאפיינים המיוחדים והספציפיים של הפרויקט שלך, ופשוט בודק כמה זמן לקח לסיים פרויקטים דומים באופן כללי במקרים קודמים. זו פעולה הנוגדת את האינטואיציה, שכן המבט מבפנים הוא כל כך הרבה יותר מפורט – אנו מתפתים לחשוב שתחזית המחושבת בקפידה, הלוקחת בחשבון את כל הנתונים הזמינים, תניב תוצאות טובות יותר.

אמנם זו האינטואיציה, אך ניסויים הראו שכל שדימיונם של הנבדקים מפורט יותר, כך הם הופכים לאופטימיים יותר (ומדויקים פחות). ביולר ואחרים ביקשו מקבוצת ניסוי של נבדקים לתאר תוכניות מפורטות ככל הניתן לקניות חג המולד שלהם – איפה, מתי ואיך.⁵ בממוצע, קבוצה זו ציפתה לסיים את הקניות יותר משבוע לפני חג המולד. נבדקים בקבוצה אחרת נשאלו בפשטות מתי הם מצפים לסיים את קניות חג המולד שלהם, עם תשובה ממוצעת של ארבעה ימים לפני החג. שתי הקבוצות סיימו בממוצע שלושה ימים לפני חג המולד.

באופן דומה, ביולר ואחרים מצאו בדיווח על מחקר בין־תרבותי, שסטודנטים יפנים ציפו לסיים את החיבורים שלהם עשרה ימים לפני הדר־ליין. בפועל הם סיימו יום אחד לפני הדר־ליין. כשנשאלו מתי הם

השלימו משימות דומות בעבר, הם ענו: "יום אחד לפני הדרדליין".⁶ זהו כוחו של המבט מבחוץ על פני המבט מבפנים. ממצא דומה הוא שמשקיפים חיצוניים מנוסים, כאלה שיודעים פחות פרטים על הפרוייקט הנוכחי, אך יש להם זיכרון רלוונטי ללמוד ממנו, נוטים להיות הרבה פחות אופטימיים והרבה יותר מדויקים מהמתכננים והמבצעים בפועל. אז ישנה שיטה אמינה למדי להתמודד עם כשל התכנון, אם אתם עושים משהו שדומה באופן כללי ל"מקרי הייחוס" של פרויקטים קודמים. פשוט בדקו כמה זמן פרויקטים דומים לקחו בעבר, מבלי לשקול אף אחד מהמאפיינים המיוחדים של הפרוייקט הנוכחי. ואפילו טוב מכך – שאלו אדם חיצוני בעל ניסיון כמה זמן לקחו פרויקטים דומים. אתם תקבלו תשובה שתישמע ארוכה להחריד, ומשקפת בבירור חוסר הבנה מוחלט של הסיבות המיוחדות שבגללן המשימה הספציפית הזו תיקח הרבה פחות זמן. התשובה שקיבלתם נכונה. תתמודדו.

-
1. Roger Buehler, Dale Griffin, and Michael Ross, "Inside the Planning Fallacy: The Causes and Consequences of Optimistic Time Predictions," in Gilovich, Griffin, and Kahneman, *Heuristics and Biases*, 250-270.
 2. Roger Buehler, Dale Griffin, and Michael Ross, "Exploring the 'Planning Fallacy': Why People Underestimate Their Task Completion Times," *Journal of Personality and Social Psychology* 67, no. 3 (1994): 366-381, doi:10.1037/0022-3514.67.3.366; Roger Buehler, Dale Griffin, and Michael Ross, "It's About Time: Optimistic Predictions in Work and Love," *European Review of Social Psychology* 6, no. 1 (1995): 1-32, doi:10.1080/14792779343000112.
 3. Buehler, Griffin, and Ross, "Inside the Planning Fallacy."
 4. Ian R. Newby-Clark et al., "People Focus on Optimistic Scenarios and

Disregard Pessimistic Scenarios While Predicting Task Completion Times," *Journal of Experimental Psychology: Applied* 6, no. 3 (2000): 171-182, doi:10.1037/1076-898X.6.3.171.

Buehler, Griffin, and Ross, "Inside the Planning Fallacy." .5

.6 ש.ם

8

אשליית השקיפות: למה אף אחד לא מבין אותך

כהטיית החוכמה בדיעבד, אנשים שיודעים את התוצאה של מצב מסוים, מאמינים שאמור להיות קל לחזות את אותה תוצאה מראש. בהכירנו את התוצאה הסופית, אנו מפרשים מחדש את המצב לאור אותה תוצאה. אפילו כשמפנים את תשומת ליבנו לכך, איננו יכולים "לבטל את הידיעה" כדי להזדהות עם מישהו שלא יודע את מה שאנו יודעים.

אשליית השקיפות קשורה לכך באופן הדוק: אנחנו תמיד יודעים למה אנו מתכוונים כשאנו מדברים, ולכן אנו מצפים שגם אחרים ידעו זאת. בקריאת טקסטים שכתבנו, הפרשנות המיועדת נוחתת במקומה בקלות, מובלת על ידי הידע שלנו לגבי כוונותינו האמיתיות. קשה להזדהות עם מישהו שצריך לפרש בצורה עיוורת, מובל ידי מילים בלבד.

גיין ממליצה למארק על מסעדה. מארק אוכל במסעדה ומגלה (א) אוכל לא מרשים ושירות בינוני או (ב) אוכל טעים ושירות ללא דופי. לאחר מכן מארק משאיר את ההודעה הבאה במשיכון של גיין: "גיין, סיימתי הרגע ארוחת ערב במסעדה שהמלצת לי עליה, ואני חייב לומר, זה היה נפלא, פשוט נפלא." קיטר הציג לקבוצת נבדקים את תרחיש א' (אוכל גרוע), ו-59% חשבו שההודעה של מארק

הייתה סרקטית ושג'יין תבחין בסרקזם.¹ בקרב הנבדקים ששמעו את תרחיש ב' (אוכל טעים), רק 3% חשבו שג'יין תבין את ההודעה של מארק כסרקטית. משתמע מדבריהם של קיסר ופאר שהנבדקים שמעו הודעה קולית אמיתית.² קיסר הראה שאם נאמר לנבדקים שהמסעדה הייתה נוראית אך גם שמארק רצה להסתיר את אכזבתו, הם האמינו שג'יין לא תבין את אותה ההודעה כסרקזם:³

הסבירות שישערו שהיא תבין סרקזם כשהוא מנסה להסתיר את החוויה השלילית שלו הייתה זהה לסבירות לכך כשהיתה לו חוויה חיובית והוא דיבר בכנות. כלומר, המשתתפים תפסו את כוונת התקשורת של מארק כשקופה. היה זה כאילו הם הניחו שג'יין תבין כל כוונה שמארק רצה שהיא תבין.⁴

"The goose hangs high" (האווז תלוי גבוה) הוא ביטוי אנגלי ארכאי שאינו בשימוש בשפה המודרנית. קיסר ופליי אמרו לקבוצת נבדקים אחת שמשמעות הביטוי היא שהעתיד נראה טוב, בעוד שקבוצת נבדקים אחרת למדה שמשמעותו היא שהעתיד נראה קודר.⁵ לאחר מכן שאלו את הנבדקים איזו משתי המשמעויות הללו סביר יותר שתיוחס לביטוי על ידי מאזין שאינו יודע את משמעותו. שתי הקבוצות האמינו כי מאזינים ייחסו לביטוי את המשמעות שהוצגה כ"סטנדרטית".

אגב, ביטויים אחרים שנבדקו כללו:

- "Come the uncle over someone"
 - "To go by the board"
 - "To lay out in lavender"
- אח, אנגלית, איזו שפה מקסימה.

קיסר והנלי בדקו את האבחנה העצמית של דוברים: האם דוברים יעריכו בחסר, יעריכו ביתר, או יעריכו נכונה באיזו תדירות המאזינים הבינו אותם.⁶ לדוברים ניתנו משפטים דרמטיים (כמו "האיש רודף אחרי אישה על אופניים") ותמונות שמראות משמעות אחת של המשפט (איש רץ אחרי אישה רוכבת), ואז ביקשו מהדוברים

לבטא את המשפט בפני מאזינים, ולהעריך כמה מהמאזינים הבינו את המשמעות המיועדת. ב-72% מהמקרים הדוברים חשבו שהם הובנו כראוי, כשבפועל הם הובנו ב-61% מהמקרים. מבין המאזינים שלא הבינו, הדוברים חשבו שהם כן הבינו ב-46% מהמקרים. כאשר המאזינים כן הבינו, הדוברים חשבו שהם לא הבינו רק ב-12% מהמקרים.

נבדקים נוספים ששמעו מהצד את ההסבר לא הראו הטיה כזו, וציפו שהמאזינים יבינו רק ב-56% מהמקרים.

כפי שקיסר ובאר מציינים, יומיים לפני ההתקפה של גרמניה על פולין, שיגר צ'מברליין להיטלר מכתב שנועד להבהיר כי בריטניה תילחם אם תתרחש פלישה כלשהי⁷. המכתב, שנוסח בשפה דיפלומטית-מנומסת, נשמע להיטלר כמפויס למדי – והטנקים שעטו קדימה.

אל תמהרו להאשים את אלו שלא מבינים את המשפטים הברורים-לחלוטין שלכם, בין בכתב ובין בעל פה. סביר להניח שהמילים שלכם הרבה פחות חדיש-שמעיות ממה שנדמה לכם.

1. Boaz Keysar, "The Illusory Transparency of Intention: Linguistic Perspective Taking in Text," *Cognitive Psychology* 26 (2 1994): 165-208, doi:10.1006/cogp.1994.1006.

2. Keysar and Barr, "Self-Anchoring in Conversation."

3. Boaz Keysar, "Language Users as Problem Solvers: Just What Ambiguity Problem Do They Solve?," in *Social and Cognitive Approaches to Interpersonal Communication*, ed. Susan R. Fussell and Roger J. Kreuz (Mahwah, NJ: Lawrence Erlbaum Associates, 1998), 175-200.

4. Keysar and Barr, "Self-Anchoring in Conversation."

5. Boaz Keysar and Bridget Bly, "Intuitions of the Transparency of Idioms: Can One Keep a Secret by Spilling the Beans?," *Journal of Memory*

and Language 34 (1 1995): 89-109, doi:10.1006/jmla.1995.1005.

Boaz Keysar and Anne S. Henly, "Speakers' Overestimation of .6
Their Effectiveness," *Psychological Science* 13 (3 2002): 207-212,
doi:10.1111/1467-9280.00439.

Keysar and Barr, "Self-Anchoring in Conversation." .7

9

ציפייה למרחקי היסק קצרים

סביבת ההסתגלות האבולוציונית (הידועה גם כ-EEA או "הסביבה הקדמונית") של ההומו סאפיינס הייתה מורכבת מציידיים-לקטים שחיו בחבורות של 200 איש לכל היותר וללא כתב. כל הידע הביין דורי הועבר באמצעות דיבור ובזיכרון.

בעולם שכזה, כל ידע מקדים הוא ידע אוניברסלי. הכל ידוע לכולם למעט מה שפרטי במובהק, נקודה.

בסביבה הקדמונית, לא סביר שהיית מוצא את עצמך במרחק של יותר מצעד היסק אחד מכל אדם אחר. אם גילית נווה מדבר חדש, אתה לא צריך להסביר לחברי השבט מהו נווה מדבר, או למה זה רעיון טוב לשתות מים, או איך ללכת. רק אתה יודע איפה נווה המדבר נמצא. זהו ידע פרטי. אך לכולם יש את הידע המקדים כדי להבין אותך כשאתה מתאר את נווה המדבר, את המושגים הדרושים כדי לחשוב על מים. זהו ידע אוניברסלי. כשאתה מסביר דברים בסביבה קדמונית, אתה כמעט אף פעם לא צריך להסביר את המושגים שלך. לכל היותר אתה צריך להסביר מושג חדש אחד, לא שניים או יותר ברזמנית.

בסביבה הקדמונית לא היו דיסציפלינות מופשטות עם גופי ידע עצומים של ראיות שנאספו בקפידה והוכללו לכדי תאוריות אלגנטיות המועברות בספרים כתובים, שמסקנותיהן נמצאות במרחק של מאה צעדי היסק מהנחות היסוד המשותפות לכולם.

בסביבה הקדמונית, כל מי שאומר משהו ללא ביסוס ברור הוא בודאי שקרן או אידיוט. לא סביר שתחשוב: "היי, אולי לאדם הזה יש ידע מקדים מבוססהיטב שאף אחד בשבט שלי אפילו לא שמע עליו", משום שהיתה זו הנחת יסוד ברורה של הסביבה הקדמונית שדבר כזה לא קרה.

לעומת זאת, אם אתה אומר משהו ברור לחלוטין והאדם השני לא מבין את זה, הוא האידיוט, או שהוא מתעקש בכוונה על מנת להכעיס אותך.

ומעל הכל, אם מישהו אומר משהו חסר בסיס ברור ומצפה ממך להאמין לו – ומתנהג כאילו הוא מזועזע לחלוטין כשאתה לא מאמין – אז הוא בטוח משוגע.

בשילוב עם אשליית השקיפות ועיגון עצמי, אני חושב שזה מסביר הרבה מהקושי האגדי שיש לרוב המדענים המנסים לתקשר עם קהל של הדיוטות – או אפילו לתקשר עם מדענים מתחומים אחרים. כשאני צופה במישהו מנסה להסביר משהו ונכשל, אני בדרך כלל שם לב שהמסביר לוקח צעד אחד אחורה, כשלמעשה הוא צריך לקחת שני צעדים אחורה או יותר. או שהמאזינים מניחים שדברים צריכים להיות מובנים בעזרת צעד אחד, כאשר נדרשים שני צעדים לפחות כדי להסבירם. שני הצדדים פועלים כאילו הם מצפים למרחקי היסק קצרים מאוד שיובילו מהידע האוניברסלי אל כל ידע חדש.

ביולוג, כשהוא מדבר עם פיזיקאי, יכול לטעון לטובת האבולוציה באומרו כי זהו ההסבר הפשוט ביותר. אבל לא כל אחד על כדור הארץ גדל על ברכי המסורת האגדית הזו של המדע, מניוטון ועד איינשטיין, המעניקה לביטוי "ההסבר הפשוט ביותר" את החשיבות האדירה שלו: ביטוי של כוח, הנאמר בלידתן של תאוריות וחרוט על מצבותיהן. למישהו אחר, הביטוי "אבל זה ההסבר הפשוט ביותר!" עשוי להישמע כמו טיעון מעניין אך ממש לא מוחץ. זה לא נותן תחושה של כלי רביעוצמה להבנת פוליטיקה משרדית או לתיקון מכונת מקולקלת. ברור שהביולוג מוקסם מרעיונותיו שלו עצמו, והתנהגותו היהירה מצביעה על כך שאין ביכולתו להיות פתוח

להסברים אלטרנטיביים שנשמעים סבירים באותה מידה. (אם לי זה נשמע סביר, זה אמור להישמע סביר לכל חבר שפוי בשבט שלי.) מנקודת מבטו של הביולוג, הוא יכול להבין איך אבולוציה עשויה להישמע קצת מוזרה בהתחלה, אבל כשמישהו דוחה אותה אפילו אחרי שהביולוג מבהיר שזה ההסבר הפשוט ביותר, טוב, ברור שהלא-מדענים האלה הם פשוט מפגרים ואין טעם לדבר איתם בכלל. טיעון ברור חייב לפרוש נתיב־היסק, המתחיל ממה שהקהל כבר יודע או מקבל. אם אתה לא מתחיל מנקודה מוקדמת מספיק אתה פשוט מדבר לעצמך.

אם בנקודה כלשהי אתה טוען טענה ללא הצדקה ברורה הנובעת מהטיעונים שביססת קודם לכן, הקהל פשוט חושב שאתה משוגע. דבר דומה יקרה גם כשאתה מרשה לעצמך להיראות כמי שמייחס לטיעון משקל רב יותר באופן גלוי ממה שמוצדק בעיני המאזינים באותו רגע. לדוגמה, לדבר כאילו אתה חושב ש"ההסבר הפשוט ביותר" הוא טיעון מוחץ לאבולוציה (מה שנכון), ולא סתם רעיון שהוא סוגי־של־מעניין (כך הוא נשמע למישהו שלא חונך על ברכי התער של אוקאם).

אה, ועדיף שלא לפזר רמזים לגבי כך שאתה חושב שאתה מדבר על דברים הרחוקים תריסר צעדי היסק ממה שהקהל יודע, או שיש לך איזה ידע מקדים מיוחד שאינו זמין להם. הקהל לא יודע שום דבר על טיעון אבולוציוני־פסיכולוגי להטייה קוגניטיבית הגורמת להערכת־חסר של מרחקי היסק המובילה לפקקי תנועה בתקשורת. הם פשוט יחשבו שאתה מתנשא.

ואם אתה חושב שאתה יכול להסביר את המושג "הערכת־חסר שיטתית של מרחקי היסק" בתמצית, ממש בכמה מילים, יש לי משהו עצוב לספר לך...

10

העדשה שרואה את פגמיה שלה

אור קורן מהשמש, פוגע בשרוכים שלך ומוחזר מהם. חלק מהפוטונים חודרים לאישוני העיניים שלך ופוגעים ברשתית. האנרגיה של הפוטונים מעוררת תגובות עצביות. התגובות העצביות מועברות לאזורי עיבוד הראייה במוח שלך, שם המידע האופטי מעובד ומורכב מחדש לכדי מודל תלת־ממדי שמתוייג כשרוך לא קשור, וכך אתה מאמין שהשרוכים שלך פרומים.

זהו הסוד של רציונליות מודעת – אין שום דבר קסום בתהליך הזה, ואתה מסוגל להבין אותו. אתה מסוגל להבין איך אתה רואה את השרוכים שלך. אתה מסוגל להבדיל בין תהליכי חשיבה שיפיקו אמונות המשקפות את המציאות, לבין תהליכי חשיבה שאינם משקפים את המציאות.

עכברים מסוגלים לראות, אך הם לא מסוגלים להבין ראייה. אתה מסוגל להבין ראייה, ובזכות יכולת זו אתה מסוגל לעשות דברים שעכברים לא יכולים לעשות. הקדש רגע להתפעל מכך, שכן אכן פלא הוא.

עכברים מסוגלים לראות, אך הם לא מודעים לכך שיש להם קליפת ראייה, ולכן הם לא מסוגלים לתקן אשליות אופטיות. עכבר חי בעולם מנטלי הכולל חתולים, חורים, גבינה ומלכודות עכברים – אך הוא לא כולל מוחות של עכברים. המצלמה שלהם לא מסוגלת לצלם את העדשה שלה־עצמה. אך אנו בני האדם, לעומת זאת, מסוגלים

להביט בתמונה שנראית מוזרה, ולהבין שחלק ממה שאנו רואים הוא בעצם העדשה עצמה. לא תמיד אנו חייבים להאמין למה שרואות עינינו, אך אנו מוכרחים להבין שיש לנו עיניים – אנו חייבים להחזיק "קופסאות" מנטליות שונות עבור המפה ועבור השטח, עבור הקלט החושי ועבור המציאות. אל תטעה לחשוב שזו יכולת טריוויאלית, זכור כמה היא נדירה בממלכת החיות.

אם לפשט את העניין, כל הרעיון של מדע הוא לחשוב באופן מודע על תהליכים אמינים שיגרמו לתודעה שלך לשקף את העולם באופן אמין יותר. מסוג הדברים שעכברים לעולם לא היו ממציאים. כשאנו מהרהרים בכל העסק הזה של "ביצוע ניסויים ניתנים לשחזור על מנת להפריך תיאוריות", אנו מסוגלים לראות למה זה עובד. מדע איננו מין תחום ידע נפרד המונח במגדל השן, הרחק מהעולם האמיתי ומההבנה של בני תמותה רגילים. מדע איננו משהו שתקף רק בתוככי המעבדות ובין דפי המאמרים, בבחינת "לא בשמים היא, לאמר מי יעלה לנו השמימה...".* המדע, כשלעצמו, הוא הלך מחשבה אמיתי בעולם, כזה הניתן להבנה, המתווך בין מוחות לבין מציאות. "...כי קרוב אליך הדבר מאוד, בפיך ובלבבך לעשותו."*

מדע הוא הגיוני, כשחושבים על כך. לעכברים אין מדע, משום שהם לא מסוגלים לחשוב על חשיבה. אל לנו להתעלם מהפלא שבכך – או מהכוח הפוטנציאלי שיכולת זו מעניקה לנו, לא רק כחברה שוחרת מדע, אלא גם כאנשים פרטיים.

אכן, הבנת התהליכים של מנוע החשיבה עשויה להיות מעט מורכבת יותר מהבנת אופן פעולתו של מנוע קיטור – אך זו לא משימה שונה באופן יסודי.

לפני שנים רבות, נכנסתי לחדר הצ'אט #philosophy של EFNet ושאלתי, "האם אתם מאמינים שתפרוץ מלחמה גרעינית ב-20 השנים הקרובות? אם לא, למה לא?" אחד האנשים שענו על השאלה כתב שהוא לא צופה מלחמה גרעינית במשך 100 שנים,

* תוספת המתרגם.

משום ש"כל השחקנים המעורבים בהחלטות לגבי מלחמה גרעינית אינם מעוניינים בה כרגע." "אבל למה שזה יהיה נכון ל-100 שנים קדימה?" שאלתי. "תקווה טהורה," הייתה תשובתו.

כשאנו מתבוננים בתהליך החשיבה הזה, אנו יכולים לראות איך המחשבה על מלחמה גרעינית גורמת לאותו אדם אי נוחות, וניתן לראות כיצד המוח שלו דוחה בשל כך את אותה מחשבה. אך אם תדמינו מיליארד עולמות – ענפי אָוּרְט, או כפילי טַגְמָארק* – תהליך החשיבה הזה לא יקצה אופטימיסטים באופן שיטתי לענפים שבהם לא פורצת מלחמה גרעינית. (חוכמולוג כלשהו ודאי יאמר בשלב זה: "אה, אבל מכיוון שיש לי תקווה, אני אעבוד קשה יותר בעבודה שלי, אתרום לכלכלה העולמית, וכך אעזור למנוע ממדינות להידרדר למצבים של אנרכיה וחוסר תקווה המהווים כר פורה למלחמות גרעיניות. כך ששני האירועים קשורים זה לזה אחרי הכל." בנקודה זו, עלינו לגרור פנימה את חוק בייס ולמדוד את הקשר באופן כמותי. לטבע האופטימי שלך לא יכולה להיות השפעה כזו גדולה על העולם. הוא לא יכול, כשלעצמו, להפחית את ההסתברות למלחמה גרעינית ב-20%, או בכמה שהאופטימי שלך הסיט את אמונותיך. הסטת האמונות שלך בכמות גדולה, עקב אירוע שמגדיל רק במעט את הסיכוי שלך להיות צודק, עדיין תעוות את המיפוי שלך.)

לשאול אילו אמונות גורמות לך אושר משמעו לפנות פנימה במקום החוצה. זו שאלה שמספרת לך משהו על עצמך, אבל לא מספקת לך שום ראייה שקשורה לסביבה החיצונית שלך. אין לי שום דבר נגד אושר, אבל הוא צריך לנבוע מתמונת העולם שצייירת בנפשך, ולא על ידי כך ששינית את הצבעים.

אם תוכל להבין את כל זאת – אם תוכל להבחין כי התקווה מסיטה את מחשבותיך מסדר-ראשון במידה גדולה מדי – אם תוכל להבין את מודעותך שלך כמנוע מיפוי בעל פגמים – אזי תוכל ליישם עבורו תיקון מודע. המוח הוא עדשה פגומה, דרכה אנו רואים את המציאות.

* אורט (Everett) הציע לראשונה את פרשנות העולמות המרובים למכניקת הקוונטים. טגמארק (Tegmark) כתב על סוגים שונים של יקומים מקבילים.

קביעה זו נכונה עבור מוחות של עכברים כמו גם מוחותיהם של בני אדם. אבל מוח אנושי הוא עדשה פגומה המסוגלת להבין את פגמיה שלה – את הטעויות השיטתיות שלה, את ההטיות שלה – ולהפעיל עליהם תיקונים מסדר-שני. היכולת הזו, כפועל, הופכת את העדשה לחזקה בהרבה. לא מושלמת, אבל חזקה בהרבה.

1. Max Tegmark, "Parallel Universes," in *Science and Ultimate Reality: Quantum Theory, Cosmology, and Complexity*, ed. John D. Barrow, Paul C. W. Davies, and Charles L. Harper Jr. (New York: Cambridge University Press, 2004), 459–491.

חלק ב'

אמונות מזויפות

...ברצפים הבאים יוצגו גם טעויות שהתפתחו וגם טעויות שהומצאו, החל מהדיון ברצף אמונות מזויפות על הדרכים שבהן ציפיותיו של האדם עשויות להיות שונות מאמונותיו המוצהרות.

— רוב בנסינגר,
הטיות: מבוא

גם אמונות צריכות לשלם שכירות (בציפיות לעתיד)

כך מתחיל המשל העתיק:

אם עץ נופל ביער ואיש לא שמע, האם הוא משמיע קול? אחד אומר, "הוא משמיע קול, משום שהוא יוצר תנודות באוויר." אחר אומר, "הוא לא משמיע קול, משום שלא בוצע עיבוד קולי באף מוח." נניח כי לאחר שהעץ נפל, השניים נכנסים יחד ליער. האם אחד יצפה לראות את העץ מוטל לימין, והשני יצפה לראותו מוטל לשמאל? נניח כי לפני שנפל העץ, השאירו השניים מכשיר הקלטה על ידו. כשישמיעו את ההקלטה, האם יצפה אחד מהם לשמוע משהו שונה מהאחר? נניח שהשניים מחברים מכשיר EEG (מכשיר המודד פעילות חשמלית במוח) לכל מוח בעולם. האם יצפה האחד לתוצאות מדידה שונות מהאחר? למרות שהשניים מתווכחים – אחד אומר "לא" והשני אומר "כן", הם לא צופים שום חוויות שונות. השניים אולי חושבים שיש להם מודלים שונים של העולם, אך אין ביניהם שום מחלוקת לגבי מה שהם מצפים שיקרה להם.

מפתה לנסות לחסל את סוג הטעויות הזה באמצעות התעקשות על כך שהסוג הלגיטימי היחיד של אמונה הוא כזה שכולל ציפייה לחוויה חושית. אך האמת היא שדברים רבים בעולם אינם נחווים באופן ישיר. אנו לא רואים את האטומים המרכיבים את הלְבָנָה, אך האטומים אכן

שוכנים שם. יש רצפה מתחת לרגליכם, אבל אינכם חווים את הרצפה באופן ישיר. אתם רואים את האור המוחזר מהרצפה, או ליתר דיוק, אתם רואים את העיבוד שהופק על ידי הרשתית וקליפת הראיה במוח שלכם. להסיק מראיית הרצפה את קיומה של רצפה פירושו לקחת צעד אחורה אל הסיבות הלא־נראות של החוויה. זה אולי נראה כמו צעד קצר וישיר מאוד, אבל זה עדיין צעד.

אתם עומדים בראש בניין גבוה, כשלידכם שעון אורולוגין עם מחוג שעות, מחוג דקות ומחוג שניות מתקתק. בידכם כדור באולינג, ואתם מפילים אותו מראש הגג. כמה תקתוקי שעון יחלפו עד שתשמעו את כדור הבאולינג מתרסק אל הקרקע?

כדי לענות במדויק, עליכם להשתמש באמונות כמו הכבידה של כדור הארץ היא 9.8 מטר לשנייה בריבוע, ו-גובהו של הבניין הוא בערך 120 מטר. האמונות הללו אינן ציפיות חסרות־מילים לחוויה חושית. הן בעלות אופי מילולי, הטוען משהו לגבי המציאות. בהחלט אפשר לחשוב על שתי האמונות האלה פשוט כמשפטים המורכבים ממילים, אך מהאמונות הללו נובעת מסקנה שהיא כן ציפייה חושית ישירה – אם מחוג השניות של השעון מורה על 12 כשאתם מפילים את הכדור, אתם מצפים לראות אותו על 1 כשאתם שומעים את ההתרסקות חמש שניות לאחר מכן. כדי לצפות מראש חוויות חושיות בצורה מדויקת ככל האפשר, עלינו לעבד אמונות שאינן ציפיות לחוויה חושית.

יכולתנו, כהומו סאפיינס, למדל את הלא־נראה טוב יותר מכל מין אחר בעולם, היא אחת מסגולותינו הכבירות ביותר. זוהי גם אחת מחולשותינו הגדולות ביותר. בני אדם מאמינים לעיתים קרובות בדברים שהם לא רק בלתי־נראים, אלא גם בלתי־אמיניים.

אותו מוח שטווה רשת של צעדי־היסק מאחורי חוויה חושית, יכול באותה מידה לטוות רשת של היסקים שאינה מחוברת לחוויה חושית, או מחוברת ברישול. האלכימאים האמינו בזמנם שפְּלוֹזִיסְטוֹן הוא הגורם לאש. נוכל לשרטט תמונה פשטנית של התודעה שלהם בעזרת תרשים של צומת (node) קטן עם התיוג "פְּלוֹזִיסְטוֹן", ושל חץ

המצביע מהצומת הזו לחוויה החושית של מדורה מתפצצת, אבל האמונה הזו לא הכילה שום תחזיות מוקדמות. הקישור מפלוגיסטון לחוויה הוגדר תמיד אחרי החוויה, במקום להגדיר מראש את מגבלותיה.

דמיינו שהמרצה שלכם לספרות אנגלית פוסט־מודרנית מלמד אתכם שהסופר המפורסם וולקי וילקינסן הוא למעשה "פוסט־אוטופיסט". מה זה אומר לגבי הציפיות שצריכות להיות לכם מהספרים שלו? לא כלום. האמונה הזו, אם אפשר לקרוא לה כך, לא מתחברת לשום חוויה חושית. מכל מקום, מוטב שתזכרו את הקביעה הטוענת ש"וולקי וילקינסן" מקיים את התכונה "פוסט־אוטופי", כדי שתוכלו להקיא את המידע הזה בבוהן הקרוב. אותו השיפוט חל גם אם "פוסט־אוטופיסטים" מראים "ניכור קולוניאלי". אם הבוחן שואל האם וולקי וילקינסן מראה ניכור קולוניאלי, מוטב שתענו בחיוב. האמונות מחוברות זו לזו, אם כי עדיין אינן מחוברות לשום חוויה צפויה.

אנו יכולים לטוות רשתות שלמות של אמונות שמחוברות אך ורק זו לזו – נקרא להן אמונות "צפות". זהו פגם אנושי ייחודי בקרב כל ממלכת החי, עיוות יכולתו של ההומו סאפיינס לטוות רשתות אמונה כלליות וגמישות יותר.

המעלה הרציונליסטית של השיטה הניסויית מורכבת משאלה מתמדת של השאלה: אילו חוויות האמונות שלנו חוזות – או טוב מכך, אוסרות. אתם מאמינים שפלוגיסטון הוא הגורם לאש? מה תצפו לראות בעקבות כך? אתם מאמינים שוולקי וילקינסן הוא פוסט־אוטופיסט? מה תצפו לראות בעקבות כך? לא, לא "ניכור קולוניאלי". איזו חוויה חושית תצפו לחוות? אתם מאמינים שאם עץ נופל ביער, ואף אחד לא שמע, הוא עדיין משמיע קול? מה אתם בהכרח מצפים לחוות?

מוטב אף יותר לשאול: איזו חוויה לא ייתכן שתחו? האם אתם מאמינים שכוח החיים (élan vital)* מסביר את החיות המסתורית

* ראו פרק 35.

של יצורים חיים? אם כן, מה האמונה הזו לא מאפשרת שיקרה – מה יפריך את האמונה הזו באופן ודאי? תשובה ריקה ("שום דבר") משמעה שהאמונה שלכם לא מגבילה את החוויה. היא מתירה לכל דבר לקרות לכם. היא צפה.

כשאתם מתווכחים על שאלה עובדתית לכאורה, נסו לשים לב להבדלים בציפיות עליהם אתם מתווכחים. אם אינכם מצליחים למצוא את ההבדל בציפיות, אתם כנראה מתווכחים על תוויות ברשת האמונות שלכם – או גרוע מכך, על אמונות צפות, חזונית הצומחת על רשת האמונות שלכם. אם אינכם יודעים אילו חוויות משתמעות מכך שוולקי וילקינסן הוא פוסט־אוטופיסט, אתם יכולים להמשיך ולהתווכח לנצח.

מעל לכל, אל תשאלו במה להאמין – שאלו למה לצפות. כל שאלה לגבי אמונות צריכה להתבסס על שאלה לגבי ציפיות, ושאלת הציפיות היא זו שצריכה לעמוד במוקד החקירה. כל השערה של אמונה צריכה להוביל להשערה ספציפית של ציפייה, והיא צריכה להמשיך לשלם שכירות ע"י יצירה של ציפיות עתידיות. אם אמונה מפסיקה לשלם שכירות, פנו אותה מהדירה.

12

משל על מדע ופוליטיקה

בתקופת האימפריה הרומית, התפלגו החיים האזרחיים בין שני מחנות: הכחולים והירוקים. הכחולים והירוקים רצחו זה את זה בקרבות יחיד, במארבים, בקרבות קבוצתיים, במהומות. פרוקופיוס אמר על המחנות הלוחמים: "כך צומחת בהם כנגד זולתם עוינות שאין לה סיבה, שאינה פוסקת או נעלמת לעולם, שכן אין היא נסוגה אף בפני קשרי נישואין, קרבת משפחה או חברות. כך הדבר אף אם החלוקים ביניהם על אודות הצבעים הללו הם אחים או כל שאריבשר אחרים."¹ אדוארד גיבון כתב: "תמיכת ה'מחנה' הפכה הכרחית לכל מועמד למשרה ציבורית או דתית."²

מי היו הכחולים והירוקים? הם היו אוהדי ספורט – תומכיהן של קבוצות מרוצי המרכבות הכחולות והירוקות.

דמיינו ציוויליזציה עתידית הנמלטת אל רשת מערות תת־קרקעית עצומה, ואוטמת את הכניסות מאחוריה. אין זה משנה אם הם נמלטו ממחלה, ממלחמה או מקרינה. נניח שראשוני שוכני המערות מצליחים לגדל מזון, למצוא מים, למחזר אוויר, להפיק אור ולשרוד, שצאצאיהם משגשגים ובסופו של דבר מקימים ערים. מהעולם שמעל נותרות רק אגדות הכתובות על קרעי נייר. אחד מקרעי הנייר הללו מתאר את השמיים: מרחב פתוח ועצום של אוויר מעל רצפה אדירה וחסרת גבולות. השמיים הם בצבע תכול, ומרחפים בהם עצמים מזורים הדומים לציצות ענק של כותנה לבנה. אך משמעות המילה

"תכול" שנויה במחלוקת – יש האומרים שהיא מתייחסת לצבע הידוע כ"כחול", ואחרים טוענים שהיא מתייחסת לצבע הידוע כ"ירוק". בימיה הראשונים של החברה התת־קרקעית, נאבקו זה בזה הכחולים והירוקים באלימות גלויה, אך כיום שוררת בחברה הפסקת אש – שלום שנולד מתוך תחושה גוברת של חוסר טעם. הנורמות החברתיות השתנו. התפתח מעמד ביניים רחב ומשגשג שגדל לצדה של אכיפת חוק יעילה והחל נוטש את האלימות. בתי הספר מעניקים פרספקטיבה היסטורית מסוימת: כמה זמן נמשך הקרב בין הכחולים והירוקים, כמה אנשים קיפחו את חייהם בשמו, כמה מעט השתנה בעקבות כך. מוחות נפתחו לרעיון החדש והמוזר, הגורס כי אדם הוא אדם, בין אם הוא כחול או ירוק.

הסכסוך לא נעלם. החברה נותרה משוסעת בין המחנה הכחול למחנה הירוק, וישנה עמדה "כחולה" ועמדה "ירוקה" כמעט בכל נושא פוליטי או תרבותי שעל סדר היום. הכחולים מצדדים במסע על הכנסה אישית, הירוקים מצדדים במיסוי על מסחר; הכחולים מצדדים בחוקי נישואים נוקשים יותר, בעוד הירוקים מבקשים להקל על הגירושיין; בסיס התמיכה של הכחולים הוא מרכזי הערים, בעוד החקלאים ומוכרי־המים מהפריפריה נוטים להיות ירוקים; הכחולים מאמינים שכדור הארץ הוא סלע כדורי ענק במרכז היקום, הירוקים מאמינים שהוא סלע שטוח ענק המקיף עצם אחר הנקרא שמש. לא כל תושב כחול או ירוק מחזיק בעמדה ה"כחולה" או ה"ירוקה" בכל סוגיה, אך יהיה זה נדיר למצוא סוחר עירוני המאמין שהשמיים כחולים, ועם זאת מצדד במיסוי על הכנסה אישית ובחוקי נישואים ליברליים יותר.

תת־הקרקע עדיין מקוטב. שלום מתוח. יש אנשים שמאמינים באמת ובתמים שכחולים וירוקים צריכים להיות חברים, ובימים אלה, ירוק שקונה בחנות כחולה או כחול שמבקר בפונדק ירוק כבר אינם בגדר מחזה נדיר. וכך, מתוך הפסקת־אש שנולדה במקור מתשישות, צומחת אט־אט רוח של סובלנות, ואפילו חברות.

יום אחד, רעידת אדמה קלה פוקדת את תת־הקרקע. קבוצה של

שישה מטיילים נקלעת לרעש בזמן בחינת ההריסות של מבני מגורים עתיקים במערות העליונות. הם חשים בתנודת הסלע תחת רגליהם, ואחת המטיילות מועדת ושורטת את ברכה. הקבוצה מחליטה לחזור על עקבותיה מחשש לרעידות אדמה נוספות. בדרכם חזרה, אחד המטיילים קולט שמץ של ריח מוזר באוויר, ניחוח המגיע ממעבר שלא היה בשימוש זמן רב. תוך התעלמות מאזהרותיהם של חבריו למסע, הוא שואל פנס ופוסע אל תוך המעבר. מסדרון האבן מתפתל מעלה... ומעלה... ולבסוף מסתיים בחור שנפער מתוך העולם, מקום שבו כל האבן נגמרת. מרחב, מרחב אינסופי, נמתח הרחק אל תוך הנצח. חלל המסוגל להכיל בתוכו אלף ערים. למעלה, גבוה באופן בלתי נתפס, בהיר מכדי להישיר אליו את המבט, ניצוץ יוקד מטיל אור על כל המרחב הנראה, חוט להט חשוף של איזו נורת ענק. באוויר, תלויות ללא כל תמיכה, ציצות ענק תמוהות של כותנה לבנה. והתקרה הזוהרת העצומה מעל... הצבע... הוא...

בנקודה זו ההיסטוריה מתפצלת, בהתאם לשאלה מי מבין חברי קבוצת המטיילים החליט לצעוד לאורך המעבר אל פני השטח. אדיטיה הכחולה עמדה תחת הכחול הנצחי, וחיכה לאיטה. לא היה זה חיוך נעים. היו בו שנאה, וגאווה פגועה. היא נזכרה בכל ויכוח שהיה לה אי פעם עם ירוק, כל יריבות, כל מאבק על קידום. "צדקת לכל אורך הדרך," לחשו לה השמים מלמעלה, "ועכשיו את יכולה להוכיח זאת." אדיטיה עמדה שם לרגע, סופגת את המסר, מתענגת עליו, ואז פנתה חזרה אל מעבר האבן כדי לבשר לעולם. בעודה הולכת, אדיטיה קמצה את ידה לאגרוף קפוץ. "הפסקת האש," היא אמרה, "הגיעה אל קיצה."

בֶּאֱרוֹן הירוק בהה באנדרלמוסיית הצבעים במשך שניות ארוכות, מבולבל. ההבנה, כשהגיעה לבסוף, הכתה בבטנו כמהלומת פטיש אוויר. דמעות פרצו מעיניו. בארון חשב על טבח קתאי, שבו טבח צבא כחול בכל תושביה של עיירה ירוקה, עד אחרון הילדים. הוא חשב על הגנרל הכחול המיתולוגי, אנאס רָל, שהכריז כי הירוקים הם "קן של מחלות. נגע שיש לבער." הוא חשב על מבטי השנאה שראה

בעיניים כחולות, ומשהו בתוכו נסדק. "איך אתם יכולים להיות בצד שלהם?" צרח בארון אל השמיים, ואז החל להתייפח. מפני שידע, בעומדו תחת הכוהק הכחול והזדוני, שהיקום, מאז ומתמיד, היה מקום של רוע.

צ'ארלס הכחול בחן את התקרה הכחולה, מוכה תדהמה. כפרופסור במכללה מעורבת, הדגיש צ'ארלס בקפדנות כי שתי נקודות המבט, הכחולה והירוקה, תקפות במידה שווה וראויות לסובלנות: השמיים היו הבניה מופשטת, ותכול הוא צבע שניתן לראותו ביותר מדרך אחת. לרגע קצר תהה צ'ארלס שמא אדם ירוק, אילו עמד כאן, היה רואה דווקא תקרה ירוקה, או שמא התקרה תהיה ירוקה בשעה זו למחרת. אך הוא לא יכול היה להמר על המשך הישרדותה של הציוויליזציה על סמך השערה כזו. הייתה זו בסך הכל תופעת טבע כלשהי, שאין לה דבר וחצי דבר עם הפילוסופיה של המוסר או עם מדעי החברה... אך בהחלט כזו שעלולה להתפרש שלא כהלכה, חשש צ'ארלס. הוא נאנח, וסב על עקבותיו אל תוך המעבר. מחר יחזור לכאן לבדו ויחסום את הפתח.

דריה, ירוקה לשעבר, נאבקה לנשום בין שרידי עולמה שהפך לאפר. אני לא ארתע, אמרה דריה לעצמה, אני לא אסיט את מבטי. כל חייה הייתה ירוקה, וכעת נגזר עליה להיות כחולה. חבריה, משפחתה, כולם יפנו לה עורף. אמרי את האמת, גם אם קולך רועד, אמר לה פעם אביה. אך כעת אביה כבר מת, ואמה לעולם לא תבין. דריה הישירה את מבטה אל תוך המבט הכחול והרגוע של השמיים, מנסה להשלים איתו, ולבסוף נרגעה נשימתה. טעיתי, היא אמרה לעצמה בעצב. זה לא כל כך מסובך, אחרי הכל. היא תמצא חברים חדשים, ואולי משפחתה תסלח לה... או, היא תהתה עם שמץ של תקווה, אולי גם הם יעמדו במבחן הזה, כשיעמדו תחת אותם השמיים? "השמיים כחולים", אמרה דריה, מגלגלת את המילים בפיה, ושום דבר נורא לא קרה, אך היא לא מצאה כוח בנפשה לחייך. דריה הכחולה פלטה אנחה עצובה וחזרה אל תוך העולם, תוהה מה תוכל לומר. אֶדִין, ירוק, הביט מעלה אל השמיים הכחולים והחל לצחוק

במרירות. כל מהלך ההיסטוריה של העולם בו חי נפרש לפניו בבהירות. אפילו הוא לא יכול היה להאמין כמה שוטים הם היו. "מטומטם, " אמר אדין, "מטומטם, מטומטם, וזה היה פה כל הזמן, ממש כאן." שנאה, רצח, מלחמות, ולאורך כל הדרך היה זה בסך הכל דבר כלשהו, שמישהו כתב עליו כמו שכותבים על כל דבר סתמי אחר. בלי שירה, בלי יופי, שום דבר שלאדם שפוי יהיה אכפת ממנו אי פעם, סתם דבר חסר טעם שנופח מעבר לכל פרופורציה. אדין נשען בלאות כנגד פתח המערה, מנסה לחשוב על דרך למנוע מהמידע הזה להחריב את העולם, ותוהה אם לא מגיע לכולם שזה בדיוק מה שיקרה. נשימתו של פריס נעתקה, והוא עמד משותק מרוב פליאה ועונג צרוף. עיניו של פריס התרוצצו סביב ברעב, נאחזות בכל מראה בתורו לפני שעברו בחוסר רצון למראה הבא. השמיים הכחולים, העננים הלבנים, הכחול העצום והלא־נודע, מלא במקומות ודברים (ואנשים?) שאף שוכן־תתיקרקע לא ראה מעולם. "הו, אז זה הוא פירושו של הצבע," אמר פריס, ויצא לחקור.

1. Procopius, *History of the Wars*, ed. Henry B. Dewing, vol. 1 (Harvard University Press, 1914).

2. Edward Gibbon, *The History of the Decline and Fall of the Roman Empire*, vol. 4 (J. & J. Harper, 1829).

13

אמונה באמונה

קרל סייגן סיפר פעם משל על מישהו שבא אלינו וטוען: "יש לי דרקון במחסן." מרתק! אנו משיבים שאנו רוצים לראות את הדרקון הזה – הבה נצא מיד אל המחסן! "אבל רגע," אומר האיש, "זה דרקון בלתי נראה."

נשים לב כי, כפי שסייגן מציין, אין בכך בכדי להפוך את הטענה לבלתי ניתנת להפרכה. אולי נלך למחסן של אותו אדם, ולמרות שלא נראה דרקון, נשמע נשימות כבדות ממקור בלתי נראה, עקבות יופיעו באופן מסתורי על הקרקע, או שמכשירי מדידה יראו שמהו במחסן צורך חמצן ופולט פחמן דרחמצני.

עם זאת, נניח שאנו אומרים לבעל המחסן, "אוקיי, נבקר במחסן ונראה אם נוכל לשמוע נשימות כבדות," והוא ממהר להשיב – לא, זה דרקון בלתי נשמע. אנו מציעים למדוד את רמות הפחמן הדרחמצני באוויר, והוא אומר שהדרקון אינו נושם. אנו מציעים להשליך קמח לאוויר כדי לראות אם הקמח יתאסף סביב קווי המתאר של דרקון בלתי נראה, והאדם אומר מיד: "הדרקון חדיר לקמח."

קרל סייגן השתמש במשל הזה כדי להמחיש את מוסר ההשכל הקלאסי, לפיו השערות גרועות נדרשות ללהטוטנות זריזה כדי להימנע מהפרכה, אבל אני מספר את המשל הזה כדי להעביר נקודה אחרת: בעל המחסן חייב להחזיק במודל מדויק של המצב במקום כלשהו בתודעתו, שהרי הוא יודע לחזות מראש בדיוק אילו תוצאות

של הניסוי יהיה עליו לתרץ.

פילוסופים מסוימים נבוכו רבות מתרחישים מעין אלה, ותהו: "האם אותו אדם באמת מאמין שיש דרקון בנמצא, או לא?" כאילו למוח האנושי יש שטח אחסון שמספיק לייצוג של אמונה אחת בלבד בכל פעם! מוחות אמיתיים הם סבוכים הרבה יותר. ישנם סוגים שונים של אמונה – לא כל האמונות הן ציפיות ישירות. אותו אדם בבירור אינו מצפה לראות שום דבר יוצא דופן עם פתיחת דלת המחסן. אחרת, הוא לא היה ממציא תירוצים מראש. ייתכן גם שמאגר האמונות ההצהרתיות שלו מכיל את המשפט יש דרקון במחסן שלי. רציונליסט עשוי לחשוב ששתי האמונות הללו אמורות להתנגש וליצור סתירה, על אף שהן מסוגים שונים. ועדיין, עובדה פיזיקלית היא שניתן לכתוב "השמיים ירוקים!" ליד תמונה של שמיים כחולים מבלי שהנייר יעלה בלהבות.

סגולתה הרציונליסטית של השיטה הניסויית אמורה למנוע מאיתנו לעשות טעויות מסוג זה. אנו אמורים לשאול את האמונות שלנו ללא הרף אילו חוויות הן חוזות, ולגרום להן לשלם שכירות במטבע של ציפיות מהעתיד. אך הבעיה של הטוען לקיומו של הדרקון עמוקה יותר, ולא ניתן לרפא אותה בעצה פשוטה שכזו. זה לא בדיוק קשה לקשר בין אמונה בדרקון לבין החוויה החושית הצפויה במחסן. אם אתה מאמין שיש דרקון במחסן שלך, אתה יכול לצפות לפתוח את הדלת ולראות דרקון. אם אינך רואה דרקון, משמע שאין דרקון במחסן שלך. זה די פשוט. נסו בעצמכם במחסן הקרוב אליכם. לא, הקטע הזה של בלתינראות הוא סימפטום של משהו גרוע בהרבה.

אם גדלת במקום ובזמן הרלוונטיים, ייתכן שאתה זוכר תקופה בה התחלת לראשונה לפקפק בקיומו של סנטה קלאוס, אבל עדיין האמנת שאתה אמור להאמין בסנטה קלאוס, אז ניסית להכחיש את הספקות. כפי שמציין דניאל דָנְט, כשקשה להאמין בדבר מה, לעיתים קרובות קל הרבה יותר להאמין שעליך להאמין בו. מה זה אומר להאמין שהשמיים הקוסמיים המוחלטים הם כחולים לגמרי

וגם ירוקים לגמרי' בעת ובעונה אחת? זו הצהרה מבלבלת. לא ברור אפילו מה המשמעות של להאמין בזה – מה בדיוק יאומן, לו היית מאמין. קל הרבה יותר להאמין שזה ראוי, שזה טוב ומוסרי ומועיל, להאמין שהשמיים הקוסמיים המוחלטים הם כחולים לגמרי וגם ירוקים לגמרי. דנט קורא לזה "אמונה באמונה".¹

וכאן העניינים מתחילים להסתבך, כדרכם של מוחות אנושיים – אני חושב שאפילו דנט מפשט יתר על המידה את האופן שבו הפסיכולוגיה הזו עובדת בפועל. ראשית, אם אתה מאמין באמונה, אתה לא יכול להודות בפני עצמך שאתה רק מאמין באמונה, כי להאמין הינו דבר בעל ערך, אך לא כן להאמין באמונה, ולכן אם אתה רק מאמין באמונה, במקום להאמין, האמונה שלך חסרת ערך. אף אחד לא יודה בפני עצמו, "אני לא מאמין שהשמיים הקוסמיים המוחלטים הם כחולים וגם ירוקים, אבל אני מאמין שאני צריך להאמין בזה" – אלא אם כן הוא בעל יכולת חריגה להכיר בחוסר המוסריות שלו עצמו. אנשים לא מאמינים ב"אמונה באמונה", הם פשוט מאמינים באמונה.

(אלו שמוצאים זאת מבלבל עשויים להפיק תועלת מלימוד לוגיקה מתמטית, המאמנת את האדם לעשות אבחנות חדות מאוד בין הטענה P , הוכחה של P , והוכחה לכך ש- P הוא ברהוכחה. באופן דומה, ישנן הבחנות חדות בין P , לרצות את P , להאמין ב- P , לרצות להאמין ב- P , ולהאמין שאתה מאמין ב- P .)^{*}

ישנם סוגים שונים של אמונה באמונה. אתה עשוי להאמין באמונה באופן מפורש. אתה עשוי לדקלם בזרם התודעה המודע שלך את המשפט "זה ראוי להאמין שהשמיים הקוסמיים המוחלטים הם כחולים לגמרי וגם ירוקים לגמרי" (תוך שאתה גם מאמין שאתה אכן מאמין בכך, אלא אם כן אתה ניחן ביכולת יוצאת הדופן להודות בכך שאינך עומד בסטנדרט המוסרי של עצמך). אך אמונה באמונה יכולה להופיע גם בצורות מרומזות יותר. אולי הטוען לקיומם של דרקון חושש מהלעג הציבורי שהוא מדמיין שייווצר אם יודה בטעותו

* לידוקבסקי יש תפיסה לא שגרתית לגבי פשטות.

בפומבי (אם כי, למעשה, רציונליסט יוריד בפניו את הכובע, בעוד סביר שאחרים ילעגו לו יותר דווקא אם ימשיך לטעון בעקשנות שיש דרקון במחסן שלו). אולי אותו אדם נרתע מעצם האפשרות להודות בפני עצמו שהדרקון לא קיים, כי הדבר יתנגש עם הדימוי העצמי שלו כמגלה המהולל של הדרקון, שראה במחסן שלו את מה שכל האחרים כשלו מלראות.

לו כל המחשבות שלנו היו משפטים מילוליים מנוסחים היטב, כמו אלו של הפילוסופים, המוח האנושי היה קל הרבה יותר להבנה עבור בני אדם. הימנעות מתמונות מנטליות מסוימות, רתיעות סמויות, תשוקות הפועלות מתחת לפני השטח – כל אלו מרכיבים את עצמיותנו לא פחות ממילים.

בעוד שאני חולק עליו בנוגע לכמה פרטים ומורכבויות, אני עדיין סבור שמושג האמונה באמונה של דנט הוא המפתח להבנת הטוען לקיומם של דרקון. אבל אנו זקוקים למושג רחב יותר של אמונה, שאינו מוגבל למשפטים מילוליים. "אמונה" צריכה לכלול מנגנוני בקרה לא־מדוברים של ציפיות. "אמונה באמונה" צריכה לכלול מכווני־התנהגות קוגניטיביים לא־מדוברים. לא יהיה זה מציאותי מבחינה פסיכולוגית לומר: "טוען־הדרקון אינו מאמין שיש דרקון במחסן שלו. הוא מאמין שזה מועיל להאמין שיש דרקון במחסן שלו." לעומת זאת, יהיה זה מציאותי לומר שטוען־הדרקון מצפה שהמציאות תתנהג כאילו אין דרקון במחסן שלו, ומתרץ תירוצים כאילו הוא מאמין באמונה.

אתה יכול להחזיק בתמונה מנטלית רגילה של המחסן שלך, ללא דרקונים בתוכו, שחווה נכונה את החוויה החושית שלך כשתפתח את הדלת, ובשום שלב לא יעבור בראשך הביטוי המילולי אין דרקון במחסן שלי. אני אפילו מוכן להתערב שזה קרה לך – שכאשר שאתה פותח את דלת המחסן או דלת חדר השינה שלך או כל דלת אחרת, ומצפה לראות אפס דרקונים, שום ביטוי מילולי כזה לא עובר במוחך. וכדי להירתע מהוויתור על אמונתך בדרקון – או מהדימוי העצמי שלך כאדם שמאמין בדרקון – אין זה הכרחי לחשוב במפורש: אני

רוצה להאמין שיש דרקון במחסן שלי. כל שנדרש הוא להירתע מעצם
האפשרות להודות בפני עצמך שאינך מאמין.

כדי לחזות מראש אילו תוצאות ניסוי יזדקקו לתירוצים, טוען
הדרקון חייב:

(א) להחזיק במקום כלשהו במוחו מודל שולט־ציפייה מדויק, ו-

(ב) לבצע פעולה מנטלית על מנת להגן על:

(1ב) האמונה ההצהרתית הצפה שלו על אודות הדרקון, או על

(2ב) הדימוי העצמי שלו כאדם המאמין בדרקון.

אם מישוהו מאמין באמונה שלו בדרקון, וגם מאמין בדרקון עצמו,
הבעיה חמורה הרבה פחות. הוא יהיה מוכן להסתכן בתחזיות
ניסוייות, ואולי אפילו יסכים לזנוח את אמונתו אם התחזית הניסויית
תוכח כשגויה – אם כי אמונה באמונה עדיין עלולה להפריע לכך,
אם האמונה עצמה אינה ודאית לחלוטין. כשמישהו ממציא תירוצים
מראש, זה מעיד על כך שהאמונה והאמונה־באמונה שלו אינן
מתואמות.

Daniel C. Dennett, *Breaking the Spell: Religion as a Natural Phenomenon* (Penguin, 2006).

14

ג'ודו בייסיאני

לעיתים אפשר להשתעשע קצת עם אנשים שהציפיות שלהם יוצאות מתיאום עם מה שהם מאמינים שהם מאמינים.

הייתי פעם באיזו מסיבה, בה ניסיתי להסביר לאחד האורחים מה אני עושה למחייתי, והוא אמר: "אני לא מאמין שבינה מלאכותית היא אפשרית, משום שרק אלוהים יכול ליצור נשמה."
בשלב זה זכיתי כנראה להשראה אלוהית, כי אמרתי מיד: "אתה מתכוון שאם אוכל ליצור בינה מלאכותית, הדבר יוכיח שהדת שלך שקרית?"

הוא אמר, "מה?"

אמרתי, "ובכן, אם הדת שלך מנבאת שאין שום אפשרות שאוכל ליצור בינה מלאכותית, פירוש הדבר שאם אצליח ליצור בינה מלאכותית, תהיה זו הוכחה לכך שהדת שלך שקרית. או שהדת שלך מכירה בכך שזה אפשרי עבורי לבנות AI, או שאם אבנה AI, יהווה הדבר הפרכה של הדת שלך."

השתררה שתיקה קצרה, כשהבחור הבין שהרגע הוא חשף את ההשערה שלו להפרכה, ואז הוא אמר, "ובכן, לא התכוונתי שלא תוכל ליצור אינטליגנציה, רק שהיא לא תוכל להיות רגשית באותו האופן כמו בני אדם."

אמרתי, "אז אם אצור בינה מלאכותית שתתחיל לדבר על רגשות שנשמעים דומים לשלנו, מבלי שתוכנתה לכך מראש, פירוש הדבר

יהיה שהדת שלך שגויה.”
הוא אמר, “ובכן, אממ, אני מניח שנצטרך להסכים שלא להסכים
בעניין הזה.”
אמרתי: “לא, למעשה לא נוכל. יש משפט ברציונליות שנקרא
‘משפט ההסכמה של אומן’ המראה כי שני רציונליסטים לא יכולים
להסכים שלא להסכים. אם שני אנשים לא מסכימים זה עם זה, לפחות
אחד מהם עושה משהו לא נכון.”
המשכנו עוד קצת עם הפינג פונג הזה. לבסוף הוא אמר, “ובכן,
אני מניח שמה שבאמת ניסיתי לומר זה שאני לא חושב שאתה יכול
ליצור משהו נצחי.”
אמרתי, “ובכן, גם אני לא חושב כך! אני שמח שהצלחנו להגיע
להסכמה על כך, כפי שדורש משפט ההסכמה של אומן.” הושטתי את
ידי, הוא לחץ אותה, ואז הלך לדרכו.
אישה שעמדה בסמוך והקשיבה לשיחה, אמרה לי בכובד ראש,
“זה היה יפהפה.”
“תודה רבה לך,” אמרתי.

15

להעמיד פני חכם

המקום החם ביותר בגיהנום שמור לאלה שבזמן משבר נשארים נייטרליים.

— דנטה אליגיירי, מומחה גיהנום מפורסם
ג'ון פ. קנדי, מצטט שגוי

זהו מנהג נפוץ שאדם יעטה מצג של נייטרליות או השהיית שיפוט על מנת לאותת על היותו בוגר, חכם, חסר פניות, או פשוט בעל נקודת מבט נעלה.

דוגמה טובה לכך היא המקרה של הוריי, שהיו מגיבים לשאלות תיאולוגיות כמו "למה למצרים העתיקה, שהיה לה תיעוד טוב בנושאים רבים אחרים, אין כל תיעוד לכך שהעברים היו שם אי פעם?" ב: "הו, גם אני נהגתי לשאול שאלות כאלה כשהייתי בגילך, אבל התבגרתי מאז."

דוגמה נוספת תהיה אותו מנהל, הניצב מול שני ילדים שנתפסו רבים בחצר, ואומר בתקיפות: "לא חשוב מי התחיל, אנחנו מסיימים את זה עכשיו." כמובן שזה חשוב מי התחיל את הריב. למנהל אולי אין גישה למידע אמין לגבי העובדה הקריטית הזו, אבל אם כך, המנהל צריך לומר זאת, לא לבטל את החשיבות של מי נתן את

האגרוף הראשון. אולי כדאי שהורה ינסה לתת אגרוף למנהל, ונראה כמה רחוק יגיע הטיעון של "זה לא משנה מי התחיל" בפני שופט. אבל מבוגרים פשוט מרגישים לא נוח כשילדים רבים, וזה כלל לא משנה לנוחות שלהם איזה ילד התחיל. זה פשוט נוח שהריב יסתיים מהר ככל האפשר.

דינמיקה דומה, אני מאמין, שוררת באירועים בדיפלומטיה בינלאומית, בהם מעצמות גדולות מורות בתקיפות לקבוצות קטנות יותר להפסיק את הלחימה הזו תכף ומיד. לא משנה למעצמה מי התחיל בסכסוך – מי התגרה, או מי הגיב באופן לא מידתי לפרובוקציה – מכיוון שאי־הנוחות המתמשכת של המעצמה היא רק פונקציה של עצם הקונפליקט המתמשך. אוף, למה ישראל וחמאס לא יכולים פשוט להסתדר כבר?

לתופעה זו אני קורא "להעמיד פני חכם". ישנן כמובן דרכים רבות לנסות לשדר חוכמה. אבל הניסיון לשדר חוכמה על ידי סירוב להעריך, סירוב לסכם ראיות, סירוב לשפוט, סירוב לבחור צד, להישאר מעל המאבק ולהביט מטה במבט רם ומתנשא – הווה אומר, לשדר חוכמה על ידי אמירה ועשייה של שום דבר – ובכן, אני מוצא שזה יזמרני במיוחד.

פאולו פֶּרִיֵּיֶרָה אמר, "לרחוץ בניקיון כפיך בזמן של קונפליקט בין החזקים לבין חסרי הכוח משמעו לצודד בחזקים, לא להיות ניטרלי."¹ חצר משחקים היא מקום נהדר לביריונים ומקום נורא לקורבנות, אם למורים לא אכפת מי התחיל. כך גם בפוליטיקה הבינלאומית: עולם בו המעצמות מסרבות לנקוט צד ורק דורשות הפסקות אש מיידיות הוא עולם נהדר לתוקפנים ומקום נורא למותקפים. אך, כמובן, זהו עולם נוח מאוד להיות בו מעצמה גדולה או מנהל בית ספר.

חלק מההתנהגות הזו, אם כן, ניתן לייחס לאנוכיות פשוטה מצדם של אותם חכמים.

אבל חלק מזה קשור גם לאיתות על נקודת מבט נעלה. אחרי הכל – מה יחשבו המבוגרים האחרים על מנהל שבאמת נראה כמי שלוקח

צד בסתם ריב בין ילדים? דבר שכזה הרי ישנמך את מעמדו של אותו מנהל לכזה של משתתף שווהמעמד בעימות!
 באופן דומה גם זקן-השבט הנערץ – שעשוי להיות מנכ"ל, אקדמאי בעל יוקרה, או מייסד של רשימת תפוצה – שמוניטין ההוגנות שלו תלוי בסירובו לשפוט בעצמו, כשכולם מסביב בוחרים צד. שני הצדדים פונים אליו לקבלת תמיכה, אך כמעט תמיד לשווא, שכן החכמים הם שופטים נערצים רק בתנאי שהם לא שופטים בפועל כמעט אף פעם – שהרי אז הם יהיו פשוט עוד צד במאבק, לא נעלים יותר מכל מתווכח אחר.

(באופן מוזר, שופטים במערכת המשפט האמיתית יכולים לפסוק פסקי דין אמיתיים שוב ושוב מבלי לאבד באופן אוטומטי את המוניטין חסר הפניות שלהם. אולי בזכות הנורמה המובנת לפיה הם מוכרחים לשפוט, שזו העבודה שלהם. או אולי כי שופטים לא צריכים לפסוק שוב ושוב בסוגיות שפיצלו את אותו השבט לו הם חייבים את המוניטין שלהם.)

ישנם מקרים שבהם השהיית השיפוט היא הדבר הרציונלי לעשות, מקרים בהם הטיות קונגטיביות גורמות לאנשים לקפוץ להכרעות. כפי שאמר מייקל רוני:

הטעות כאן דומה לזו שאני רואה כל הזמן אצל סטודנטים מתחילים לפילוסופיה: כשהם מתמודדים עם סיבות להיות ספקנים, הם הופכים במקום זאת לרלטיביסטים. כלומר, כשהמסקנה הרציונלית היא להשהות את השיפוט לגבי סוגיה מסויימת, יותר מדי אנשים מסיקים במקום זאת שכל שיפוט יהיה סביר באותה מידה.

אם כן, איך נוכל להימנע מאותה התנהגות כאילררציונליסטית (הקשורה אך נפרדת) של איתות על חוסר הפניות הלא-מוטה שלנו על ידי הטענה השקרית שמאזן הראיות הנוכחי הוא ניטרלי? "טוב, כמובן שמסתובבים בעולם הרבה דרוויניסטים נלהבים, אבל אני חושב שהראיות שיש לנו לא באמת מאפשרות לנו לפסוק באופן ברור לטובת הברירה הטבעית על פני תכנון תבוני."

בנקודה זו הייתי מציע לזכור שניטרליות היא שיפוט מוחלט. היא לא מהווה ריחוף מעל שום דבר. זו בעצם הצגת העמדה המוחלטת והמסוימת לפיה מאזן הראיות במקרה הספציפי הזה מתיר רק סיכום ראיות אחד, שהוא במקרה ניטרלי. גם זה יכול להיות שגוי. צידוד בעמדה ניטרלית הוא בהגדרה חשוף לביקורת בדיוק כמו צידוד בכל עמדה אחרת.

כך גם לגבי שאלות של מדיניות. אם מישהו אומר שלשני הצדדים, מתנגדי הפלות ותומכי הפלות, יש נקודות טובות ושהם באמת צריכים לנסות להתפשר ולכבד זה את זה, הוא לא נוקט עמדה שהיא מעל שני הצדדים הסטנדרטיים בדיון על הפלות. הוא מציג שיפוט מוחלט, פרטני בדיוק באותה מידה כמו לומר "נגד הפלות!" או "בעד הפלות!"

אם אתה שואף לשפר את היכולת הכללית שלך לגבש אמונות מדויקות יותר, ייתכן שיהיה זה יעיל להימנע מלהתמקד בסוגיות טעונות רגשית כמו הפלות או הסכסוך הישראלי-פלסטיני. אך הסיבה לכך היא לא שרציונליסט הוא בוגר מכדי לדבר על פוליטיקה. גם לא שרציונליסט נמצא הרחק מעל המאבק המטופש הזה שבו רק פעילים פוליטיים פשוטים וצעירים נלהבים ישפילו את עצמם כדי להשתתף בו.

כפי שרובין הנסון מתאר זאת, היכולת לנהל שיחות בנושאים נפיצים היא משאב מוגבל. אם אתה יכול למצוא דרך שלא להתמודד ישירות עם השאלה, תצא נשכר בכך שהשקעת את המשאבים המוגבלים שלך על סוגיות רחבות פחות באופן יחסי, שבהן דיון שולי עשוי להניב תשואות שוליות גבוהות יחסית.

אך במקרה כזה תחומי האחזיות שאתה מוריד בסדר העדיפויות שלך הם עניין של ניהול משאבים מוגבלים. לא עניין של ריחוף גבוה מעל הדיון, תוך עטיית ארשת חכמה ושלווה.

התגובה שלי להערה של פול גרהם ב-Hacker News נראית לי כמו סיכום שראוי לחזור עליו:

יש הבדל בין:

- החלטה על שיפוט ניטרלי.
- ויתור על השקעת משאבים שוליים.
- העמדת פנים כאילו כל אחד מהנ"ל הוא אות לחוכמה עמוקה, בגרות ונקודת מבט נעלה, עם הפרשנות המשתמעת שהצדדים המקוריים בדיון תופסים עמדות נחותות יותר, שבמבט מגובה שכזה – ההבדלים ביניהם נראים זניחים.

1. Paulo Freire, *The Politics of Education: Culture, Power, and Liberation* (Greenwood Publishing Group, 1985), 122.

16

תביעת הדת להיות בלתי ניתנת להפרכה

באופן אירוני, התיאור המוקדם ביותר שאני מכיר לניסוי מדעי, הוא הסיפור של אליהו ונביאי הבעל.

עם ישראל פוסח על שתי הסעיפים בבחירה בין אלוהי ישראל לבין הבעל, אז אליהו מכריז שבכונתו לערוך ניסוי על מנת להכריע בדבר – רעיון חדשני למדי באותם ימים! נביאי הבעל יעלו את הפר שלהם על מזבח, ואליהו יעלה את הפר שלו על מזבח, אך לשני הצדדים לא יותר להצית את האש. האל שיוריד אש מהשמים על קורבנו יוכרו כאל האמיתי. נביאי הבעל משמשים כקבוצת ביקורת עבור אליהו – אותו חומר בעירה מעץ, אותו הפר, ואותם נביאים הקוראים בשם האל, אך האל כוזב. בשלב זה אליהו שופך מים על המזבח שלו – ובכך הורס את הסימטריה הניסויית, אך היה זה אז, בימים ההם – כדי להחמיר על עצמו במופגן את נטל ההוכחה, כמו הצורך ברמת מובהקות של 0.05. האש יורדת על המזבח של אליהו, ומהווה את התצפית הניסויית. בני ישראל הצופים צועקים "ה' הוא האלוהים!" – ביקורת עמיתים.

מיד לאחר מכן גוררים אנשי העם את 450 נביאי הבעל אל נחל קישון ומספסים את גרונם. זה אמנם קשוח, אך הכרחי. עליך להשליך בתקיפות את ההשערה שהופרכה, ולעשות זאת במהירות, לפני שהיא תתחיל לתרץ תירוצים כדי להגן על עצמה. אם יורשה לכוהני הבעל

לחיות, הם יתחילו לקשקש על כך שהדת היא "תחום ידע נפרד" שלא ניתן להוכיחו או להפריכו.

בימים עברו, אנשים באמת האמינו בדת שלהם ולא רק היו מאמינים של הדת שלהם. הארכיאולוגים המקראיים שיצאו לחפש את תיבת נוח לא חשבו שהם מבזבזים את זמנם. הם באמת ציפו לכך שהם עשויים להתפרסם. רק לאחר שנכשלו במציאת ראיות מאששות – ומצאו במקומן ראיות מפריכות – ביצעו אנשי הדת את מה שוויליאם פֶּארטְלִי כינה הנסיגה אל המחויבות, "אני מאמין כי אני מאמין".

בימים עברו, הרעיון של דת כתחום ידע נפרד לא היה קיים. התנ"ך הוא זרם תודעה של מידע תרבותי: היסטוריה, חוק, משלי מוסר, וכן – מודלים של אופן פעולת היקום. באף פסוק אחד בתנ"ך לא תמצאו מישהו מדבר על פליאה נשגבת ממורכבותו של היקום.* אך בהחלט תמצאו בו שפע של טענות מדעיות, כמו על כך שהיקום נברא בשישה ימים (שהם מטאפורה למפץ הגדול), או ארנבות שמעלות גרה** (שהן מטאפורה ל...).

בימים עברו, לומר על הדת המקומית שהיא "לא ניתנת להוכחה" היה מביא אותך להישרף על המוקד. אחת מאמונות הליבה של היהדות האורתודוקסית היא שאלוהים הופיע בהר סיני ואמר בקול רועם, "כן, הכל אמת." מנקודת מבט בייסיאנית זו חתיכת ראייה חד משמעית לישות בעלת עוצמה על-אנושית (אם כי זה לא מוכיח שאותה ישות היא אלוהים כשלעצמו, או שהיא שוחרת טוב – אלו יכולים להיות גם חייזרים בגיל ההתבגרות). רובן המכריע של הדתות בהיסטוריה האנושית – להוציא רק את אלו שהומצאו ממש לאחרונה – מספרות סיפורים על אירועים שהיו מהווים ראיות חסינות לביקורת

* פסוקים רבים כאלה ודאי מצויים בתנ"ך, ונראה שיש להתייחס לאמירות מסוג זה בפרק הנוכחי בהקשר הרטורי שלהן ולא כאל טענות עובדתיות.

** על פי המקרא (ויקרא י"א ו') הארנבת נמנית עם מעלי הגרה. מבחינה מדעית, הארנבת אינה מעלה גרה במובן המקובל, אך היא מבצעת תהליך דומה שנקרא צקוטרוֹפִיָה (Cecotrophy) – אכילה מחדש של גללים רכים ועשירים בחומרי מזון (צקוטרוֹפִיָים), המאפשרת לה לספוג רכיבים תזונתיים חיוניים שעברו התססה במעי.

בעליל לו היו מתרחשים בפועל. מערכת היחסים המרוחקת שבין דת לבין שאלות עובדתיות היא רעיון חדשני ומערבי לחלוטין. האנשים שכתבו את התנ"ך כלל לא העלו על דעתם שיש הבדל.

האימפריה הרומית ירשה פילוסופיה מהיוונים הקדמונים, כפתה חוק וסדר בפרובינציות שלה, שמרה על רישומים בירוקרטיים ואכפה סובלנות דתית. הברית החדשה, שנכתבה בתקופת האימפריה הרומית, נושאת עקב כך כמה מאפיינים של מודרניות. אי אפשר היה להמציא סיפור על כך שאלוהים משמיד לחלוטין את העיר רומא (בסגנון סדום ועמורה), כי ההיסטוריונים הרומים היו שמים לב לשקר, ולא יכולת פשוט לסקול אותם באבנים.

האנשים שהמציאו את סיפורי התנ"ך, לעומת זאת, יכלו להמציא פחות או יותר כל מה שרצו. אגיפטולוגים מוקדמים באמת הופתעו כשלא הצליחו למצוא שום עקבות לכך ששבטים עבריים היו אי פעם במצרים – הם לא ציפו למצוא תיעוד מפורט של עשרת המכות, אך הם ציפו למצוא משהו. כפי שהסתבר, משהו אכן נמצא. הם גילו שבמהלך התקופה המשוערת של יציאת מצרים, מצרים שלטה ברוב השטח של ארץ כנען. זו חתיכת טעות היסטורית עצומה, אבל בהיעדרן של ספריות, אין מי שיחשוף את הבלוף.

לאימפריה הרומית, לעומת זאת, היו ספריות. אי לכך, הברית החדשה אינה טוענת לניסים גיאופוליטיים גדולים וראוותניים כפי שהתנ"ך עשה כדבר שבשגרה. במקום זאת, הברית החדשה טוענת לניסים קטנים יותר שלמרות הכל, עדיין מתאימים לאותה מסגרת של ראיות. נער נופל וקצף יוצא מפיו. הסיבה להתקף היא רוח טמאה. סביר לצפות שרוח טמאה תברח מפני נביא אמת, אך לא תברח מפני שרלטן. ישו מגרש את הרוח הטמאה. נובע מכך שישו הוא נביא אמת ולא שרלטן. זוהי חשיבה בייסיאנית רגילה לחלוטין, אם אתה מקבל את הנחת היסוד שאפילפסיה נגרמת על ידי שדים (ושסיומו של התקף אפילפטי מוכיח שהשד ברח).

בעבר, לא רק שהדת טענה טענות בנושאים עובדתיים ומדעיים, הדת טענה טענות לגבי כל נושא. הדת קבעה חוקים – לפני שהיו

גופי חקיקה; הדת עסקה בהיסטוריה – לפני שנולדו היסטוריונים וארכיאולוגים; הדת קבעה את המוסר המיני – לפני שקמה התנועה לשחרור האישה; הדת הגדירה את יחסי הכוחות בחברה – לפני שנכתבו חוקות; והדת ענתה על שאלות מדעיות החל בטקסונומיה ביולוגית וכלה בהיווצרות הכוכבים. התנ"ך לא מדבר על תחושת פליאה ממורכבות היקום – הוא עסוק מדי בקביעת עונש מוות לנשים שלבשו בגדי גברים, דבר שנחשב כתוכן דתי לגיטמי לחלוטין באותה תקופה. התפיסה המודרנית של דת כעוסקת בנושאים אתיים גרידא נובעת מכך שכל שאר התחומים נשאר מאחור עקב הופעתם של מוסדות טובים יותר. אתיקה היא מה שנשאר.

או ליתר דיוק, אנשים חושבים שאתיקה היא מה שנשאר. קחו זרם תודעה תרבותי מלפני 2,500 שנה. עם הזמן, האנושות תתקדם באופן משמעותי, וחלקים מאותה תרבות עתיקה יהפכו למיושנים באופן יותר ויותר צורם. האתיקה אינה חסינה מפני הקידמה האנושית – למשל, אנו לא רואים כיום בעין יפה פרקטיקות תנ"כיות כגון החזקת עבדים. למה אנשים חושבים שהאתיקה התנ"כית עדיין מהווה יריב שקול לזו המודרנית?

באופן מהותי, אין שום דבר זניח בבעיה המוסרית העולה מטבח של אלפי ילדים בכורים חפים מפשע ממין זכר על מנת לשכנע פרעה לא־נבחר לשחרר עבדים שבאופן לוגי יכלו באותה מידה להשתגר החוצה מהמדינה. זה אמור להיות צורם יותר מהטעות המדעית הטריטוריאליטלמדי של האמירה לפיה לחגבים יש ארבע רגליים.* ובכל זאת, אם תטען שהארץ שטוחה, אנשים יסתכלו עליך כאילו אתה משוגע. אך אם תגיד שאתה מתייחס לתנ"ך כמקור יסודי למוסר, נשים לא יסטרו לך. תפיסת הרציונליות של רוב האנשים נקבעת על פי מה שהם חושבים שיהיה בטוח עבורם. הם חושבים שביכולתם

* במקרא מתואר 'שרץ העוף' כבעל ארבע רגליים (ויקרא י"א), כשהכוונה על פי חלק מהפרשנים היא גם לחרקים בעלי שש רגליים כגון זבובים ויתושים. דווקא החגב מתואר בתור 'שרץ העוף ההולך על ארבע אשר לו כרעיים ממעל לרגליו', אך בחלק מהתרגומים לאנגלית קל לטעות ולהבין שהכרעיים הם חלק מארבע הרגליים.

לתמוך במוסר התנ"כי ולצאת בשלום, ולכן נדרש מהם רק מאמץ נסבל של הונאה עצמית כדי להתעלם מהבעיות המוסריות של התנ"ך. כולם הסכימו שלא להבחין בפיל שבחדר, ומצב עניינים שכזה יכול להתקיים למשך זמן מה.

אולי יום אחד האנושות תתקדם צעד אחד קדימה, וכל מי שיכיר בתנ"ך כמקור מוסרי יזכה לאותו יחס כמו טָרְנַט לוט התומך בקמפיין הנשיאותי של סְטְרוֹם ת'ורמונד*. ואז נאמר ש"המהות האמיתית" של הדת היתה מאז ומעולם אילנות יוחסין או משהו כזה.

הרעיון שהדת היא תחום ידע נפרד שלא ניתן להוכיחו או להפריכו הוא שקר גדול – שקר שחוזרים עליו שוב ושוב, כך שאנשים אומרים אותו מבלי לחשוב, אך בבחינה ביקורתית הוא פשוט לא נכון. זהו עיוות פרוע של הדרך בה התנהלה הדת מבחינה היסטורית, של איך כל כתבי הקודש מתארים את האמונות המיוצגות בהם, של מה שאומרים לילדים על מנת לשכנע אותם, ושל מה שרוב האנשים הדתיים על פני כדור הארץ עדיין מאמינים בו. אי אפשר שלא להעריץ את החוצפה הטהורה שבדבר, המגיעה לאותה רמה כמו "אוקיאניה תמיד הייתה במלחמה עם אי־סְטְאָסְיָה".** התובע מציג את הגרזון נוטף הדם, והנאשם, המום לרגע, חושב במהירות ואומר: "אבל אתה לא יכול להפריך את חפותי באמצעות ראיות פשוטות – זה תחום ידע נפרד!"

ואם זה לא עובד, תפוס חתיכת נייר ושרבט לעצמך כרטיס חיסם של "צא מהכלא".

* סוג מוקדם של טרוליות.

** מתוך הספר 1984, מיניסטריון האמת.

17

הצהרות וקריאות עידוד

פעם נכחתי בפאנל בנושא "האם דת ומדע יכולים לדור בכפיפה אחת?" אחת הנשים בפאנל, שהציגה את עצמה כפגאנית, הפליגה במונולוג בלתי נגמר על אמונתה בכך שכדור הארץ נברא כאשר פרה קדמונית ענקית נולדה אל תוך התהום הקדמונית, ואז ליקקה לכדי קיום אל קדמוני, שצאצאיו הרגו ענק קדמוני ויצרו מגופתו את כדור הארץ, וכך הלאה. הסיפור היה ארוך, מפורט, ומגוחך אפילו יותר מהרעיון שכדור הארץ נתמך על גבו של צב ענקי, ואין שום ספק כי הדוברת ידעה מספיק מדע כדי להבין זאת.

אני עדיין מתקשה למצוא את המילים המדוייקות שיתארו את מה שראיתי בעודה מדברת. היא דיברה ב... גאוה? שביעות רצון עצמית? התרסה מכוונת?

אותה אישה המשיכה בתיאור מיתוס הבריאה שלה במשך עוד נצח או שניים, שנמשכו כנראה חמש דקות. אותן תחושות מוזרות של גאוה/סיפוק/התרסה היו קשורות באופן ברור לכך שהיא ידעה שהאמונות שלה מקוממות מבחינה מדעית. וזה לא שהיא שנאה מדע, כחברת פאנל היא הצהירה שדת ומדע יכולים לדור בכפיפה אחת. היא אפילו דיברה על כך שזה די הגיוני שהויקינגים דיברו על תהום קדמונית, בהתחשב בארץ שבה חיו – ובכך הסבירה את הדת של עצמה! – ובכל זאת התעקשה על כך שזה הסיפור בו היא "מאמינה", באומרה את החלק האחרון בסיפוק משונה.

אני לא בטוח שניתן למתוח את הרעיון של דניאל דָּנְט על "אמונה באמונה" מספיק רחוק על מנת להסביר את האירוע הזה. זה היה מוזר יותר מזה. היא לא דקלמה את מיתוס הבריאה שלה באותה אמונה פנאטית של מישהי שמנסה לשכנע את עצמה. היא גם לא התנהגה כאילו היא מצפה מאיתנו, הקהל, להשתכנע – או כאילו היא זקוקה לאמונה שלנו כדי לתת תוקף לזו שלה.

דנט, בנוסף להצגת הרעיון של אמונה באמונה, הציע גם את הפרשנות לפיה לחלק ניכר ממה שמכונה "אמונה דתית" צריך בעצם להתייחס כאל "הצהרה דתית". נניח שחייזר אנתרופולוג היה חוקר קבוצה של סטודנטים פוסט־מודרניסטים לספרות אנגלית בה כולם מאמינים לכאורה שוולקי וילקינסן הוא סופר פוסט־אוטופי. השאלה המתאימה לא תהיה: "מדוע כל הסטודנטים מאמינים באמונה המוזרה הזו?" אלא "מדוע כולם כותבים את המשפט המוזר הזה במבחנים?" גם אם משפט כלשהו הוא חסר משמעות ביסודו, אתה עדיין יכול לדעת מתי עליך לדקלם את התשובה בקול.

אני חושב שדנט היה אולי ציני מדי כשהציע שהצהרה דתית משמעה אך ורק דקלום האמונה מן השפה ולחוץ. רוב האנשים מספיק כנים כך שאם הם ידקלמו הצהרה דתית בקול רם, הם ירגישו מחויבים לומר את המשפט המילולי גם אל תוך זרם התודעה שלהם. אבל אפילו הרעיון של "הצהרה דתית" נראה כלא מספק בתור הסבר לטענתה של האישה הפגאנית לגבי אמונתה בפרה הקדמונית. אילו היית צריך להצהיר על אמונה דתית כדי לספק כומר, או חבר לדת – לעזאזל, אפילו כדי לספק את הדימוי העצמי שלך כאדם דתי – היית צריך לפחות להעמיד פנים שאתה מאמין באופן משכנע הרבה יותר ממה שעשתה אותה אישה. כשהיא דקלמה את אותה מעשייה על פרה קדמונית, באותה התרסה מוזרה, היא אפילו לא ניסתה להיות משכנעת – אפילו לא ניסתה להראות לנו שהיא לוקחת את הדת שלה ברצינות. אני חושב שזה החלק שכל כך ערער אותי. אני מכיר אנשים שמאמינים שהם מאמינים בדברים הזויים, אבל כשהם מצהירים עליהם בפומבי, הם משקיעים הרבה יותר מאמץ בלשכנע את עצמם

שהם לוקחים את האמונות שלהם ברצינות. לבסוף הכתה בי ההבנה שהאישה הזו כלל לא ניסתה לשכנע אותנו, או אפילו לשכנע את עצמה. הנושא של כל אותו דקלום של סיפור הבריאה ההוא, לא היה קשור בכלל לבריאת העולם. למען האמת, בכך שפתחה בנאום חוצב להבות של חמש דקות על הפרה הקדמונית, היא בעצם הריעה לפגאניזם, ממש כאילו היא מניפה שלט במשחק כדורגל. שלט שמכריז יאללה מכבי אינו מהווה הצהרת עובדה, או ניסיון לשכנע; אין לו צורך להיות משכנע – הוא קריאת עידוד.

אותה התרסה מוזרה... זה היה כאילו היא צועדת עירומה במצעד הגאווה. (לא שיש משהו רע בלצעוד עירומה במצעד הגאווה. לסביות היא לא משהו שהאמת יכולה להשמיד.) זו לא הייתה רק קריאת עידוד, כמו צעידה, אלא קריאת עידוד שערורייתית, כמו צעידה בעירום – הטומנת בחובה את האמונה שלא ניתן לעצור אותה או לבקר אותה, כי היא עושה זאת למען מצעד הקהילה שלה. זו הסיבה שהיה חשוב לתאר את המיתוס בצורה כה מגוחכת. אילו הייתה מנסה לגרום לו להישמע סביר יותר, היה זה כמו ללבוש בגדים.

18

אמונה כסממן קבוצתי

עד כה הבחנתי בין אמונה כמייצרת־ציפיות, אמונה באמונה, הצהרה, וקריאת עידוד. מבין אלו, אנו עשויים לקרוא לאמונות המייצרות ציפיות "אמונות תקינות" ולצורות האחרות "אמונות לא־תקינות". גם אמונה תקינה יכולה להיות שגויה או לא רציונלית, כמו במקרה של מישהו המצפה בכנות שתינוקו החולה יחלים מכוחה של תפילה. אך ניתן לטעון שהצורות האחרות הן "לא אמונות כלל".

צורה נוספת של אמונה לא־תקינה היא אמונה כהזדהות קבוצתית – כדרך של השתייכות. רובין הנסון משתמש במטאפורה המצוינת של לבישת בגדים ייחודיים: מדים קבוצתיים כמו גלימת כומר או כיפה יהודית. לכן אני קורא לזה "אמונה כסממן קבוצתי".

במונחים של פסיכולוגיה אנושית ריאליסטית, המוסלמים שהטיסו את המטוסים אל תוך מגדלי התאומים ראו את עצמם ללא ספק כגיבורים המגנים על האמת, הצדק, והדרך האיסלאמית מפני מפלצות חייזריות מעוותות ברוח הסרט היום השלישי*. רק חנון חסר ניסיון לחלוטין, מסוג החנונים שאין להם מושג איך לא־חנונים רואים את העולם, יאמר זאת בקול רם בבר באלבמה. זה לא דבר אמריקאי לומר. הדבר האמריקאי לומר הוא שהטרוריסטים "שונאים את החופש שלנו" ושהטסת מטוס לתוך בניין היא "מעשה פחדני". הביטויים

* סרט מדע בדיוני על פלישת חייזרים בו טייס גיבור מרסק את מטוסו אל חללית הקרב החייזרית על מנת להציל את האנושות.

"הקרבה עצמית הירוואית" ו"מחבל מתאבד" לא יכולים להיאמר באותו משפט, אפילו לצורך תיאור מדויק של האופן שבו האויב רואה את העולם. עצם המושג של אומץ ואלטרואיזם כמאפיינים של מחבל מתאבד הוא סממן־אויב – קל לדעת את זה, כי הרי זה מה שהאויב אומר. פחדנות וסוציופתייה של מחבל מתאבד הן סממן־אמריקאי. לא הודפסו המרכאות בהן תוכל להשתמש כדי לדבר על הצורה בה האויב רואה את העולם; זה יהיה כמו להתחפש לנאצי בפורים.

אמונה־כסממן־קבוצתי עשויה להסביר כיצד אנשים יכולים להיות נלהבים לגבי אמונות לא־תקינות. סתם אמונה באמונה, או הצהרה דתית, לרוב לא יצליחו לייצר אפקטים רגשיים אמיתיים, עמוקים ועוצמתיים. או כך לפחות אני חושד, אני מודה שאינני מומחה לנושא. אך הרושם שלי הוא כזה: אנשים שהפסיקו לייצר ציפיות כאילו הדת שלהם נכונה, ישקיעו מאמצים כדי לשכנע את עצמם שהם נלהבים, וניתן בטעות לזהות את הייאוש הזה כתשוקה. אך זו לא אותה אש שהייתה להם בילדותם.

לעומת זאת, קל מאוד לבן אנוש להשתייך לקבוצה באופן אמיתי, מלא תשוקה, מהבטן, ולהריע לקבוצת הספורט האהובה עליו. (זה היסוד שעליו נשענת התרמית של "רפובליקנים נגד דמוקרטים" ודילמות כוזבות דומות במדינות אחרות, אבל זה נושא לפעם אחרת.)* הזדהות שבטית היא כוח רגשי חזק מאוד. אנשים ימסרו את נפשם למענו. וברגע שאתה גורם לאנשים להזדהות עם שבט, האמונות שמהוות סממן שלו ייִצְעֶקו במלוא התשוקה של השתייכות לאותו השבט.

* "...אבל כל עוד הם יכולים לומר שטהורידים הם עליונים, הם יכולים להרגיש עליונים בעצמם, הם יכולים להרגיש חלק מהמעמד השליט..."

19

נורת מחיאות הכפיים

בכנס הסינולוריות של 2007, אחד הדוברים קרא לפיתוח דמוקרטי ורב־לאומי של בינה מלאכותית. אז ניגשתי למיקרופון ושאלתי:

נניח שקבוצה של מדינות דמוקרטיות מקימה קואליציה מחקרית לפיתוח AI, והרבה פוליטיקה ותככים מעורבים בתהליך – לקבוצות אינטרס מסוימות יש השפעה גדולה באופן חריג, קבוצות אחרות נדחקות הצידה – במילים אחרות, התוצאה נראית בדיוק כמו התוצרים של דמוקרטיות מודרניות. לחלופין, נניח שקבוצה של גיקים מורדים מפתחת AI במרתף של אחד מחבריה, ומנחה את ה-AI לערוך משאל בקרב כל האנשים בעולם – תוך הצנחת טלפונים סלולריים לכל מי שאין לו – ולעשות את כל מה שיחליט הרוב. איזו אפשרות היא "דמוקרטית" יותר בעיניך, והאם תרגיש בטוח עם אחת מהן יותר מאשר עם השניה?

רציתי לגלות האם הוא מאמין ביעילות המעשית של ההליך הפוליטי הדמוקרטי או שהוא מאמין בצדק המוסרי של הצבעה ישירה. אך הדובר השיב:

התרחיש הראשון נשמע כמו מאמר מערכת במגזין *Reason**, והשני נשמע כמו עלילה של סרט הוליוודי.

* מגזין אמריקאי ליברטריאני.

מבולבל, שאלתי:

אז על איזה סוג של תהליך דמוקרטי כן חשבת?

הדובר השיב:

משהו בסגנון פרויקט הגנום האנושי – שהיה פרויקט מחקר בחסות בינלאומית.

שאלתי:

איך קבוצות אינטרס שונות יפתרו את המחלוקות ביניהן במבנה ארגוני כמו פרויקט הגנום האנושי?

והדובר אמר:

אני לא יודע.

חילופי הדברים הללו הזכירו לי ציטוט של דיקטטור כזה או אחר, שנשאל אם יש לו כוונות כלשהן להניע את מדינת המחמד שלו לכיוון של דמוקרטיה:

אנחנו מאמינים שאנו חיים במערכת דמוקרטית כבר כעת. כמה גורמים חסרים עדיין, כמו הביטוי של רצון העם.

המהות של דמוקרטיה היא המנגנון הספציפי שנועד לפתור קונפליקטים בנושאי מדיניות. אם כל הקבוצות היו מאמינות באותה מדיניות, לא היה צורך בדמוקרטיה – היינו משתפים פעולה באופן אוטומטי. תהליך הפתרון יכול להיות מסוג של הכרעת רוב ישירה, או גוף מחוקק נבחר, או אפילו התנהגות רגישה-לבוחרים של בינה מלאכותית, אבל חייב להיות משהו. מה זה אומר לקרוא לפתרון "דמוקרטי" אם אינך מדמיין שום מנגנון לפתרון קונפליקטים?

אני חושב שזה אומר שאם אמרת את המילה "דמוקרטיה", אז הקהל אמור למחוא כפיים. זו לא ממש הצהרה שמכילה טענה כלשהי, אלא פשוט המקבילה הרטורית לנורת ה"מחיאורת כפיים" שמסמנת

לקהל באולפן מתי לפצוח בתשואות. המקרה הספציפי הזה ראוי לציון רק משום שטעיתי לחשוב שנורת מחיאות הכפיים היא הצעה למדיניות, מה שגרם לאחר מכן למבוכה לכל המעורבים. רוב נורות מחיאות הכפיים הן בוטות הרבה יותר, וניתנות לזיהוי על ידי מבחן היפוך פשוט. למשל, נניח שמישהו אומר:

אנחנו צריכים לאזן בין הסיכונים וההזדמנויות של AI.

אם נהפוך את ההצהרה הזו, נקבל:

אנחנו לא צריכים לאזן בין הסיכונים וההזדמנויות של AI.

מכיוון שההיפוך נשמע לא נורמלי, ההצהרה הלא-הפוכה היא כנראה נורמלית, ומרמזת על כך שהיא לא טומנת בחובה שום מידע חדש. ישנן סיבות לגיטימיות רבות להגיית משפט חסר-אינפורמציה כשהוא עומד בפני עצמו. המשפט: "אנחנו צריכים לאזן בין הסיכונים וההזדמנויות של AI" יכול להציג נושא לדיון; הוא יכול להדגיש את החשיבות של הצעה ספציפית לאיזון; הוא יכול לבקר הצעה לא מאוזנת. קישור לטענה נורמלית יכול להעביר מידע חדש לרציונליסט מוגבל* – הקישור עצמו עשוי שלא להיות ברור מאליו. אבל אם אין שום פרטים שמגיעים בהמשך, המשפט מתפקד כנראה כנורת מחיאות כפיים.

אני מתפתה לתת הרצאה שתהיה מורכבת כולה אך ורק ממשפטים המהווים נורות מחיאות כפיים, ולראות כמה זמן יקח לקהל להתחיל לצחוק:

אני כאן כדי לטעון בפניכם היום שאנחנו צריכים לאזן בין הסיכונים וההזדמנויות של בינה מלאכותית מתקדמת. עלינו להימנע

* רציונליות מוגבלת (Bounded rationality) היא רעיון פסיכולוגי לפיו היכולת לערוך שיפטים ולקבל החלטות באופן אופטימלי, או אפילו משיע רצון, עשויה להיות מוגבלת בשל תנאים מסוימים שבהם מתקבלת ההחלטה - למשל עומס קוגניטיבי, חוסר במידע או מגבלות זמן (ויקיפדיה).

מהסיכונים, וככל שזה אפשרי, לממש את ההזדמנויות. אל לנו להתעמת שלא לצורך עם סכנות מיותרות לחלוטין. כדי להשיג מטרות אלו, עלינו לתכנן את צעדינו בתבונה ובאופן רציונלי. אל לנו לפעול מתוך פאניקה ופחד, או להיכנע לטכנופוביה, אך גם להיזהר שלא לפעול מתוך התלהבות עיוורת. עלינו לכבד את האינטרסים של כל הצדדים המביעים עניין בסינגולריות. עלינו לנסות להבטיח שהיתרונות של טכנולוגיות מתקדמות יונגשו לאנשים רבים ככל האפשר, במקום להיות מוגבלים למעטים. עלינו לנסות להימנע, ככל האפשר, מקונפליקטים אלימים המשתמשים בטכנולוגיות אלו, ועלינו למנוע מיכולות הרס מסיביות ליפול לידיהם של יחידים. עלינו להקדים ולחשוב לעומק על הסוגיות האלו מוקדם ככל הניתן, ולפני שיהיה מאוחר מדי לעשות משהו לגביהן...

לשים לב לבלבול

תיאור של אי-רציונליות יהיה בהכרח לוקה בחסר אם לא יספק שום תאוריה לגבי האופן שבו עובדת רציונליות – או אם ה"תאוריה" שלו תהיה מורכבת רק מאמיתות נדושות ומעורפלות, ללא שום מנגנון תיאורי מדויק. הרצף לשים לב לבלבול שואל מדוע יהיה זה מועיל לבסס את ההתנהגות על ציפיות "רציונליות", ואיך אנו עשויים להרגיש בעשותנו זאת.

– רוב בנסינגר,

הטיות: מבוא

20

למקד את אי-הוודאות

האם התשואות על האג"ח יעלו, ירדו, או יישארו ללא שינוי? אם אתה פרשן טלוויזיה והתפקיד שלך הוא להסביר את התוצאה בדיעבד, הרי שאין סיבה לדאגה. לא משנה איזו אפשרות תתממש, תוכל להסביר מדוע התוצאה תואמת באופן מושלם לתיאוריית השוק החביבה עליך. אין כל סיבה לחשוב על שלוש האפשרויות הללו כמנוגדות זו לזו באיזשהו מובן, או כשוללות זו את זו, כיוון שתקבל את מלוא הניקוד על הפרשנות שלך ללא תלות באפשרות שתתממש.

אבל רגע! נניח שאתה פרשן טלוויזיה מתחיל, ואתה לא מספיק מנוסה כדי לשלוף הסברים משכנעים על המקום. עליך להכין את הדברים מראש לשידור של מחר, ויש לך זמן מוגבל להתכונן. במקרה כזה, היה עוזר אילו יכולת לדעת מראש איזה תרחיש יתממש בפועל – האם תשואות האג"ח יעלו, ירדו או יישארו ללא שינוי – מכיוון שאז היית צריך להכין רק סט אחד של תירוצים.

למרבה הצער, איש אינו יכול לחזות את העתיד. אם כן, מה עליך לעשות? לא תוכל להשתמש ב"הסתברויות", כמובן. בית הספר לימד אותנו ש"הסתברויות" הן מספרים קטנים המופיעים לצד בעיה מילולית, וכאן אין שום מספרים קטנים. גרוע מכך, אתה מרגיש חוסר ודאות. אינך זוכר שהרגשת חוסר ודאות בזמן שביצעת חישובים על אותם מספרים קטנים בבעיות המילוליות. שיעורי מתמטיקה בקולג' הם מקומות מסודרים וסטרייליים, ומכאן נובע שהמתמטיקה עצמה

לא חלה על מצבי חיים שאינם מסודרים וסטרייליים. הדבר האחרון שהיית רוצה הוא לשנע כישורי חשיבה מהקשר אחד לאחר בצורה לא הולמת. ברור שזה לא עניין ל"הסתברויות".

כך או כך, יש לך רק 100 דקות להכין את התירוצים שלך. אינך יכול להשקיע את כל 100 הדקות ב"עלייה", וגם להשקיע את כל 100 הדקות ב"ירידה", וגם להשקיע את כל 100 הדקות ב"ללא שינוי". עליך לתעדף איכשהו.

לו היית צריך להצדיק את ניצול הזמן שלך בפני ועדת ביקורת, היית חייב להקדיש זמן שווה לכל האפשרויות. מכיוון שבשום מקום לא כתובים מספרים קטנים, לא היה לך שום מקור עליו תוכל להסתמך על מנת להצדיק השקעה של פרקי זמן שונים. אתה ממש יכול לשמוע את המבקרים מטיחים בך: "ולמה, מר פינקלדינגר, הקדשת בדיוק 42 דקות לתירוץ מס' 3? למה לא 41 דקות, או 43? תודה בזה – אתה לא אובייקטיבי! יש לך העדפות אישיות!"

ואז, בהבזק קטן של הקלה, אתה נזכר שאין שום ועדת ביקורת שתוכל לנזוף בך. וזה טוב, מכיוון שמחר צפוי הבנק המרכזי להוציא הודעה חשובה, ונראה לא סביר שמחירי האג"ח יישארו ללא שינוי. לא תרצה להקדיש 33 דקות יקרות לתירוץ שאינך צופה שתזדקק לו. מחשבתך נודדת שוב ושוב לאותם הסברים בהם אתה משתמש בהופעותיך בטלוויזיה, המראים תמיד שכל אירוע תואם לתיאוריית השוק שלך באופן מתקבל על הדעת. אך במהרה מתברר כי עצם סבירותם של ההסברים לא תעזור לך כאן – שלושת האירועים מתקבלים על הדעת במידה דומה. היכולת שלך להתאים כל תוצאה לתיאוריית השוק המועדפת עליך לא עוזרת לך בקשר לשאלה כיצד לחלק את הזמן שלך. ישנו פער בלתי ניתן לגישור בין 100 הדקות שלך, שהן משאב מוגבל, לבין היכולת שלך להסביר כיצד תוצאה מסויימת מתאימה לתיאוריה שלך, שהיא בלתי מוגבלת.

ובכל זאת... אפילו במצב חסר הוודאות בו אתה שרוי, נדמה שאתה מצפה להתרחשות של שלושת האירועים באופן שונה; שאתה מצפה להזדקק לתירוצים מסוימים יותר מאשר לאחרים. וזה החלק

המרתק – כשאתה חושב על משהו שגורם לעליה במחירי האג"ח להיראות סבירה יותר, אתה בהכרח מרגיש שסביב פחות שתזדקק לתירוץ למקרה שמחירי האג"ח ירדו או יישארו ללא שינוי.

נדמה שיש אפילו קשר כלשהו בין המידה שבה אתה מצפה להתממשות כל אחת משלוש התוצאות, לבין כמה זמן אתה רוצה להקדיש להכנת כל תירוץ. כמובן שלא ניתן באמת לכמת את אותו קשר. יש לך 100 דקות להכין את הנאום שלך, אבל אין 100 יחידות של שום דבר לחלק, בעסק הזה של הציפייה. (אם כי אתה כן מסיק שאם תוצאה מסוימת תתרחש, אז פונקציית התועלת שלך היא לוגריתמית ביחס לזמן שהושקע בהכנת התירוץ.)

ועדיין... אתה חוזר בראשך שוב ושוב לרעיון שהציפייה היא מוגבלת, בשונה מהיכולת לתרוץ, אך בדומה לזמן להכנת תירוצים. אולי יש להתייחס לציפייה כאל משאב מוגבל, כמו כסף. האינסטינקט הראשוני שלך הוא לנסות להשיג עוד ציפייה, אך מהר מאוד אתה מבין שגם אם תשיג עוד ציפייה, לא יהיה בידך זמן רב יותר להכין את התירוצים. לא, האסטרטגיה היחידה שלך היא להקצות את המלאי המוגבל של ציפייה בצורה הטובה ביותר.

אתה די בטוח שלא לימדו אותך שום דבר מהסוג הזה בקורסים לסטטיסטיקה. אף אחד לא אמר לך מה לעשות כשהרגשת חוסר ודאות נורא כל כך. אף אחד לא אמר לך מה לעשות כשלא ניתנו לך מספרים קטנים. שהרי, גם לו היית מנסה להשתמש במספרים, היית יכול להשתמש בכל סוג של מספרים באותה מידה – אין בנמצא שום רמז לגבי איזה סוג של מתמטיקה נחוצה פה, אם בכלל עליך להשתמש במתמטיקה! אולי היית משתמש בסוף בזוגות של מספרים, מספרים ימניים ושמאליים, שהיית קורא להם DS כקיצור של Dexter-Sinister...* או מי יודע מה עוד? (למרות שיש לך רק 100 דקות להשקיע בהכנת תירוצים.)

* צמד המילים Dexter Sinister (מלטינית: ימין ו-שמאל) משמש בתיאור של סמלי אצולה במסורת האירופית. יודקובסקי רומז כאן לתאוריית Dempster-Shafer, (ובקיצור DS) – שיטה מתמטית המשתמשת בווז מספרים, שמייצגים אמונה ואפשרויות, כדי לייצג אי-ודאות ולהעריך את סבירותה של טענה.

לו רק הייתה אמנות העוסקת במיקוד של אי-הוודאות שלך – של מיצוי ציפיייה רבה ככל הניתן אל תוך התוצאה שתתממש בפועל!
איזה שם היינו יכולים לתת לאמנות שכזו? ואילו מין חוקים
עשויים להיות לה?

21

מהי ראייה?

המשפט "השלג לבן" הוא אמת אם ורק אם השלג לבן.

— אלפרד טרסקי

לומר על מה שישנו, שהוא ישנו, או על מה שאינו, שהוא אינו, זוהי אמת.

— אריסטו, מטאפיזיקה IV

אם שני הציטוטים הללו לא נראים לכם כמו הגדרה מספקת ל"אמת", דלגו אל האינטרלוד האמת הפשוטה. בפרק זה בכוונתי לדבר על "ראיות". (כמו כן, בכוונתי לדון באמונות עובדתיות, ולא ברגשות או מוסר, כפי שהוגדרו בפרק להרגיש רצינולי.)

בעודך הולך ברחוב, שרוכי נעליך נפרמים. זמן קצר לאחר מכן, מסיבה מוזרה, אתה מתחיל להאמין ששרוכי נעליך פרומים. אור קורן מהשמש, פוגע בשרוכים שלך ומוחזר מהם. חלק מהפוטונים חודרים לאישוני העיניים שלך ופוגעים ברשתית. האנרגיה של הפוטונים מעוררת תגובות עצביות. התגובות העצביות מועברות לאזורי עיבוד הראייה במוח שלך, שם המידע האופטי מעובד ומורכב מחדש לכדי מודל תלת־ממדי שמתווייג כשרוך פרום. ישנו רצף של אירועים,

שרשרת של סיבה ותוצאה, בעולם האמיתי ובתוך מוחך, המסתיימת בכך שאתה מאמין במה שאתה מאמין. התוצאה הסופית של התהליך היא מצב תודעתי המשקף את מצבם של שרוכי נעליך בפועל.

מהי ראיה? זהו אירוע השזור, בקשרים של סיבה ותוצאה, בכל דבר שברצונך לדעת עליו. אם מושא החקירה שלך הוא, למשל, שרוכי נעליך, אזי האור החודר לאישוניך הוא ראיה השזורה בשרוכי נעליך. אין לבלבל זאת עם המושג הטכני של "שזירה" המשמש בפזיקה* – כאן אני מדבר על "שזירה" רק במובן של שני דברים המגיעים למצבים קשורים בשל קשרי הסיבה והתוצאה ביניהם.

לא כל השפעה יוצרת את סוג ה"שזירה" הנדרש לראיה. מכונה שמצפצפת כשאתה מזין לתוכה מספרי לוטו זוכים לא תועיל לך, אם המכונה מצפצפת באותו אופן גם כשאתה מזין מספרי לוטו שאינם זוכים. האור המוחזר מנעליך לא יהווה ראיה שימושית לגבי שרוכי הנעליים, אם הפוטונים היו מסיימים את מסעם באותו מצב פיזיקלי בין אם השרוכים קשורים ובין אם לאו.

אם נאמר זאת באופן מופשט: כדי שאירוע יהווה ראיה לגבי מושא חקירה, עליו להתרחש באופן שונה, בצורה שתהיה שזורה עם המצבים האפשריים השונים של המושא. (ובניסוח טכני: חייבת להתקיים אינפורמציה הדדית של שאנון** בין האירוע הראייתי לבין מושא החקירה, יחסית למצב איהוודאות הנוכחי שלך לגבי שניהם.) שזירה יכולה להיות מדבקת כאשר היא מעובדת נכון, אך לשם כך אתה זקוק לעיניים ולמוח. אם פוטונים מוחזרים משרוכי נעליך ופוגעים בסלע, הסלע לא ישתנה יותר מדי. הסלע לא ישקף את השרוכים בשום דרך בעלת משמעות. הוא לא ישתנה באופן ניכר

* שזירה קוונטית (Quantum Entanglement) היא תופעה במכניקת הקוונטים שבה פעולה על חלק אחד של מערכת מבוזרת חוצת את המצב של החלק האחר, גם כאשר שתי המערכות מרוחקות זו מזו ואין אפשרות של מעבר מידע ביניהן (ויקיפדיה).

** אינפורמציה הדדית של שאנון הוא מושג מתורת האינפורמציה, שפיתח קלוד שאנון (Claude Shannon), המודד את כמות המידע שניתן ללמוד מפרט מידע אחד על פרט אחר.

כתלות במצב השרוכים. זו הסיבה שסלעים הם עדים ממש גרועים בבתי משפט. סרט צילום לעומת זאת, 'יתפוס' את השזירה-עם-השרוך בעזרת הפוטונים הנכנסים, כך שהתמונה עצמה תוכל לשמש כראיה. אם עיניך ומוחך פועלים כשורה, אתה עצמך תהפוך לשזורעם-שרוכי הנעליים-שלך.

זו הסיבה שרציונליסטים מייחסים חשיבות רבה כל כך לטענה הפרדוקסלית-לכאורה, שאמונה היא בעלת ערך רק אם היית יכול, עקרונית, להשתכנע להאמין אחרת. אם הרשתית שלך מגיעה תמיד לאותו מצב ללא תלות באור שנכנס אליה, אתה למעשה עיוור. מערכות אמונה מסוימות טוענות, תוך שימוש בטריק שקוף למדי, שישנן אמונות שהן בעלות ערך אמיתי רק אם אתה מאמין בהן ללא תנאי – ללא קשר למה שאתה רואה, ללא קשר למה שאתה חושב. המוח שלך אמור להגיע לאותו מצב ללא תלות בדבר. מכאן הביטוי – "אמונה עיוורת". אם מה שאתה מאמין בו אינו תלוי במה שאתה רואה, הרי שאתה מתהלך באפילה ממש כאילו ניקרת את עיניך שלך. אם מוחך ועיניך פועלים כשורה, אמונותיך יהפכו בסופו של דבר לשזורות בעובדות. חשיבה רציונלית מייצרת אמונות המהוות ראיות בפני עצמן.

אם פיך דובר אמת, אזי האמונות הרציונליות שלך, המהוות ראיות בפני עצמן, יכולות לשמש כראיות עבור אדם אחר.* שזירה יכולה לעבור דרך שרשראות של סיבה ותוצאה – ואם אתה מדבר, ואחר שומע, גם אלו הן סיבה ותוצאה. כשאתה אומר את המילים "השרוכים שלי פרומים" אל תוך הטלפון, אתה חולק עם חבר את השזירה-עם-השרוכים.

אשר על כן, אמונות רציונליות הן מדבקות, בקרב אנשים ישרים המאמינים זה ליושרו של זה. וזו הסיבה שהטענה לפיה אמונותיך אינן מדבקות – שאתה מאמין מסיבות פרטיות שאינן ניתנות להעברה

* "האם עליי להיות משוכנע?" שאל הארי. "רק משום שאתה מאמין בכך? האם אתה רציונליסט חזק מספיק כעת כדי שהאמונה שלך תהיה ראייה חזקה בעיני, משום שעד כדי כך לא סביר שתאמין בכך אם זה לא נכון?"

— היא חשודה כל כך. אם אמונותיך שזורות במציאות, הן אמורות להיות מדבקות בקרב אנשים ישרים.

אם מודל המציאות שלך גורס כי המסקנות של תהליכי החשיבה שלך לא אמורים להיות מדבקים עבור אחרים, אז המודל שלך בעצם אומר שאמונותיך לא מהוות ראיות בפני עצמן, כלומר אינן שזורות במציאות. עליך להחיל תיקון רפלקטיבי, ולהפסיק להאמין.

כמובן, אם תרגיש, ברמת תחושת הבטן, מה כל זה אומר, ממילא תחדל מלהאמין באופן אוטומטי. מכיוון ש"האמונה שלי אינה שזורת במציאות" פירושו "האמונה שלי אינה מדויקת". ברגע שאתה מפסיק להאמין ש"השלג לבן" הוא אמת, עליך להפסיק להאמין (באופן אוטומטי!) ש"השלג לבן", או שמהו מאוד יסודי לא בסדר.

אז קדימה, הסבר מדוע סוג תהליכי החשיבה שבהם אתה משתמש מייצרים באופן שיטתי אמונות המשקפות את המציאות. הסבר מדוע אתה חושב שאתה רציונלי. מדוע אתה חושב שבאמצעות שימוש בתהליכי חשיבה כמו אלו בהם אתה נוקט, תודעות אחרות יאמינו לבסוף ש"השלג לבן" אם ורק אם השלג לבן. אם אינך מאמין שהמסקנות של תהליכי החשיבה שלך שזורות במציאות, מדוע אתה מאמין לאותן מסקנות? זה אותו הדבר, או לפחות אמור להיות.

22

ראיה מדעית, ראיה משפטית, ראיה רציונלית

נניח שחברך הטוב, מפכ"ל המשטרה, מספר לך בסודי סודות כי ראש ארגון הפשע בעיר שלך הוא וולקי וילקינסן. כרציונליסט, האם יהיה זה מוצדק מצדך להאמין להצהרה שכזו? אומר זאת כך: אם תחליט לפתע להתגרות בוולקי, הייתי אומר שזה יהיה מעשה פזיז מצדך. מכיוון שזה נכון לפעול כאילו לוולקי יש סבירות גבוהה מהממוצע להיות ראש ארגון פשע, אנו מבינים כי הצהרתו של המפכ"ל מהווה ראיה בייסיאנית חזקה.

מערכת המשפט שלנו לא תאסור את וולקי על סמך הצהרת המפכ"ל בלבד. הצהרה מסוג זה אינה קבילה כראיה משפטית. אם תאסור כל אדם שהמפכ"ל מאמין שהוא ראש ארגון פשע, אולי תתפוס הרבה ראשי ארגוני פשע כשלב הראשון, יחד עם עוד כמה אנשים שהמפכ"ל פשוט לא מחבב. כוח נוטה להיות משחית. אך ככל שיעבור הזמן, תתפוס פחות ופחות ראשים אמיתיים של ארגוני פשע (שיתאמצו יותר להישאר אנונימיים) ויותר ויותר קורבנות חפים מפשע (כוח בלתי מרוסן מושך שחיתות כמו שדבש מושך זבובים). כל זה לא אומר שהצהרת המפכ"ל לא מהווה ראיה רציונלית. "יחס הסבירות" שלה עדיין נוטה לצד אחד, וזה עדיין יהיה טיפשי מצדך להתגרות בוולקי. אך ברמה החברתית, במסגרת הניסיון להשיג

מטרה חברתית כלשהי, אנו מגדירים "ראיה משפטית" באופן מכוון כך שתכלול רק סוגים מסוימים של ראיות, כמו תצפיותיו של המפכ"ל בעצמו בליל ה-4 באפריל. באופן אידיאלי, כל ראיה משפטית צריכה להיות גם ראיה רציונלית, אך לא להפך. אנו מחילים סטנדרטים מיוחדים, נוספים וחזקים יותר לפני שאנו מכתירים ראיה רציונלית כ"ראיה משפטית".

בשעה שאני כותב משפט זה, ב-33:20 שעון החוף המערבי, ב-18 באוגוסט 2007, אני גורב גרביים לבנים. כרציונליסט, האם אתה רשאי להאמין להצהרה הזו? כן. האם אוכל להעיד על כך בבית משפט? כן. האם זו הצהרה מדעית? לא, מפני שאין ניסוי שתוכל לבצע בעצמך על מנת לאמת אותה. מדע מורכב מהכללות שחלות על מקרים פרטיים רבים, בצורה כזו שניתן לערוך ניסויים בעולם האמיתי שיבדקו את ההכללה, ובכך לאמת אותה בעצמנו, מבלי להסתמך על סמכותו של אף אחד. מדע הוא הידע הניתן לשחזור ציבורי של האנושות.

כמו מערכת המשפט, המדע כתהליך חברתי מורכב מבני אדם שיכולים לטעות. אנו רוצים מאגר מוגן של אמונות אמינות במיוחד. ואנו רוצים מנגנון של כללים חברתיים שיעודדו יצירת ידע שכזה. זו הסיבה שאנו מטילים סטנדרטים מיוחדים, נוספים וחזקים יותר לפני שאנו מכתירים ידע רציונלי כ"ידע מדעי", ובכך מוסיפים אותו למאגר האמונות המוגן. האם רציונליסט רשאי להאמין בקיומו ההיסטורי של אלכסנדר הגדול? כן. אנו מחזיקים בתמונה כללית של יוון העתיקה, שאינה מהימנה לחלוטין אבל היא טובה יותר מאנטרופיה מקסימלית. אבל אנו תלויים בסמכויות כגון פלוטארכוס*. לא נוכל לנוח את פלוטארכוס ולאמת הכל בעצמנו. ידע היסטורי אינו ידע מדעי.

האם רציונליסט רשאי להאמין שהשמש תזרח ב-18 בספטמבר 2007? כן – לא בוודאות מוחלטת, אבל זה ההימור הנכון (לפרדנטיס: פרשו זאת כשמירה על מהירות הסיבוב של כדור הארץ ויציבות המסלול שלו ביחס לשמש). האם הצהרה זו, כשאני כותב מאמר זה

* פילוסוף והיסטוריון יווני, שחי במאות הראשונה והשנייה לספירה.

ב-18 באוגוסט 2007, היא אמונה מדעית?

שלילת התואר "מדעי" מהצהרות כמו "השמש תזרח ב-18 בספטמבר 2007" עשויה להישמע כמו מעשה חתרני. אם המדע לא היה יכול לחזות אירועים עתידיים – אירועים שטרם התרחשו – הוא היה חסר תועלת. הוא לא יכול היה לחזות דבר לפני עריכת ניסוי. התחזית שהשמש תזרח היא, ללא ספק, הסקה הנובעת מהכללות מדעיות. היא מבוססת על מודלים של מערכת השמש שתוכל לבדוק בעצמך באמצעות ניסוי.

אך דמיין שאתה מתכנן ניסוי כדי לאמת את תחזית מספר 27, בהקשר חדש, של תיאוריה מקובלת Q. ייתכן שאין לך שום סיבה קונקרטית לחשוד שהאמונה שגויה, אתה פשוט רוצה לבדוק אותה באותו הקשר חדש. נראה מסוכן לומר שקיימת "אמונה מדעית" לגבי התוצאה, לפני ביצוע הניסוי. ישנה "התחזית הקונבנציונלית" או "תחזית לפי תיאוריה Q". אבל אם אתה כבר יודע מראש את ה"אמונה המדעית" לגבי התוצאה, למה שתטרח לערוך את הניסוי? אני מקווה שאתה מתחיל להבין מדוע אני מזהה את המדע עם הכללות, ולא עם ההיסטוריה של ניסוי ספציפי כלשהו. אירוע היסטורי מתרחש פעם אחת; הכללות חלות על אירועים רבים. ההיסטוריה אינה ניתנת לשחזור; הכללות מדעיות כן.

האם ההגדרה שלי ל"ידע מדעי" היא נכונה? זו שאלה שאינה מוגדרת היטב. הסטנדרטים המחמירים שאנו מחילים על המדע נובעים מבחירות פרגמטיות. בשום מקום בין הכוכבים או על צלעות ההרים לא כתוב ש- $p < 0.05$ הוא הסטנדרט לפרסום מדעי. רבים טוענים כיום ש-0.05 הוא סטנדרט חלש מדי, וכי יהיה זה מועיל להוריד אותו ל-0.01 או 0.001.

אולי אנשי הדורות הבאים, שיפעלו על פי התפיסה שהמדע הוא הידע האנושי הניתן לשחזור ציבורי, יתייגו כ"מדעי" רק מאמרים שפורסמו בכתבי עת הפתוחים לכל הציבור. אם חלק מהידע מוגן בחומת תשלום, האם הוא באמת חלק מידע האנושות? האם נוכל לסמוך על תוצאה של ניסוי, אם אנשים צריכים לשלם כדי לבקר

אותה? האם זה באמת מדע?

השאלה "האם זה באמת מדע?" אינה מוגדרת היטב. האם כתב עת עם מחיר מנוי של 20,000 דולר לשנה באמת מהווה ראיה בייסיאנית? כמו בהבטחתו הפרטית של המפכ"ל לגבי זהותו של וולקי כראש הארגון, אני סבור שעלינו לענות "כן". אבל האם עלינו להכתיר את כתב העת הסגור כ"מדע"? האם עלינו לאפשר לו להיכנס למאגר האמונות המיוחד והמוגן? לכשעצמי, אני סבור שהמדע יצא נשכר אם נקבל את העקרון לפיו רק ידע פתוח נחשב כחלק ממאגר הידע האנושי הניתן לשחזור ציבורי.

23

כמה ראיות נדרשות?

מוקדם יותר הגדרתי ראייה כ"אירוע השזור, בקשרים של סיבה ותוצאה, בכל דבר שברצונך לדעת עליו", ואת המונח שזור כ"מתרחש באופן שונה עבור מצבים שונים של המטרה". אז איזו כמות של שזירה – איזה נפח של ראיות – נדרש כדי לתמוך באמונה?

נפתח בשאלה פשוטה מספיק כדי להיות מתמטית: כמה חזק תצטרך לשזור את עצמך עם הגרלת לוטו על מנת לזכות? נניח שישנו מיכל עם 70 כדורים, הנשלפים בזה אחר זה ללא החזרה, ועליך לנחש שישה מספרים רצופים כדי לזכות. אז ישנם 131,115,985 שילובים אפשריים, ולכן כרטיס שנבחר באקראי יהיה בעל סיכוי של $\frac{1}{131,115,985}$ לזכות (0.0000007%). על מנת לזכות בהגרלה, תזדקק לראיות סלקטיביות דיין כדי להעדיף שילוב אחד על פני 131,115,984 חלופות.

הבה נניח שישנן כמה בדיקות שאתה יכול לבצע, המבחינות, באופן הסתברותי, בין כרטיסי הגרלה מנצחים לכרטיסי הגרלה מפסידים. לדוגמה, אתה יכול להזין צירוף מספרים לקופסה שחורה קטנה שתמיד מצפצפת אם השילוב הוא המנצח, ויש לה סיכוי של $\frac{1}{4}$ (25%) בלבד לצפצף אם הרצף שגוי. במונחים בייסיאניים, נאמר שיחס ההסתברות הוא 4 ל-1. כלומר, הסיכוי של הקופסה לצפצף כאשר מזינים צירוף נכון הוא פי 4 גבוה יותר, בהשוואה לסיכוי שתצפצף עבור צירוף שגוי.

עדיין ישנם המון צירופים אפשריים. אם תזין 20 צירופים שגויים, הקופסה תצפץ עבור 5 מהם בממוצע באופן אקראי. אם תזין את כל 131,115,985 הצירופים האפשריים, הקופסה תצפץ בוודאות עבור הצירוף המנצח, אבל היא תצפץ גם עבור 32,778,996 צירופים מפסידים (בממוצע).

אז נכון שהקופסה הזו לא תאפשר לך לזכות בהגרלה, אבל זה עדיין עדיף מכלום. אם תשתמש בקופסה, סיכויך לזכות יעלו מ-1 ל-131,115,985 ל-1 ל-32,778,997. עשית התקדמות מסוימת לעבר המטרה שלך, מציאת האמת, בתוך מרחב האפשרויות העצום.

נניח שאתה יכול להשתמש בקופסה שחורה נוספת כדי לבדוק צירופים פעמיים, באופן בלתי תלוי. שתי הקופסאות יצפצפו בוודאות עבור הכרטיס המנצח, אבל הסיכוי שקופסה תצפץ עבור צירוף מפסיד הוא $\frac{1}{4}$ באופן בלתי תלוי עבור כל אחת מהקופסאות, ומכאן שהסיכוי ששתי הקופסאות יצפצפו עבור צירוף מפסיד הוא $\frac{1}{16}$. אפשר לומר שהראיות המצטברות, משתי בדיקות בלתי תלויות, הן בעלות יחס הסתברות של 16:1. מספר הכרטיסים המפסידים שיעברו את שתי הבדיקות יהיה (בממוצע) 8,194,749.

מאחר שיש 131,115,985 כרטיסי הגרלה אפשריים, היית חושב שתזדקק לראיות בעוצמה של כ-131,115,985 ל-1 – אירוע, או סדרת אירועים, שסבירותם להתרחש עבור צירוף מנצח גבוהה פי 131,115,985 מאשר עבור צירוף מפסיד. למעשה, כמות כזו של ראיות תספיק רק כדי לתת לך סיכוי של 50 אחוז לזכות בהגרלה. מדוע? כי אם תפעיל מסנן בעוצמה כזו על 131 מיליון כרטיסים מפסידים, אז בממוצע יהיה כרטיס מפסיד אחד שיעבור את המסנן. הכרטיס המנצח יעבור גם הוא את המסנן. או אז תישאר עם שני כרטיסים שעברו את המסנן, ורק אחד מהם מנצח. סיכויי זכייה של חמישים אחוז, בהנחה שניתן לקנות כרטיס אחד בלבד.

דרך טובה יותר להסתכל על הבעיה: בשלב הראשון יש כרטיס מנצח אחד ו-131,115,984 כרטיסים מפסידים, אז סיכויך לזכות הם 1:131,115,984. אם אתה משתמש בקופסה אחת, הסיכוי שהיא

תצפץף הוא 1 עבור כרטיס מנצח ו-0.25 עבור כרטיס מפסיד. אז אנחנו מכפילים 1:131,115,984 ב-0.25:1 ומקבלים 1:32,778,996. הוספת קופסה נוספת של ראיות מכפילה שוב את הסיכויים ב-0.25:1, כך שכעת הסיכויים הם כרטיס מנצח 1 מול 8,194,749 כרטיסים מפסידים.

נוח למדוד ראיות ב"ביטים" – לא ביטים כמו בהארד דיסק, אלא ביטים של מתמטיקאים, שהם שונים מהותית מבחינה קונספטואלית. ביטים של מתמטיקאים הם לוגריתמים בבסיס $\frac{1}{2}$ של הסתברויות. לדוגמה, אם יש ארבע תוצאות אפשריות א ב ג ו-ד, עם הסתברויות של 50%, 25%, 12.5% ו-12.5%, ואני אומר לך שהתוצאה היא ד, אז העברתי לך שלושה ביטים של מידע, כי הודעתי לך על תוצאה שהסתברות שלה היא $\frac{1}{8}$.*

מסתבר ש-131,115,984 הוא קצת פחות מ-2 בחזקת 27. לכן, 14 קופסאות שחורות או 28 ביטים של ראיות – אירוע שהסתברות שיקרה אם ההשערה של הכרטיס נכונה גבוהה פי 268,435,456 מאשר אם היא שגויה – ישנו את הסיכויים מ-1:131,115,984 ל-1:131,115,984:268,435,456 שמצמצם ל 1:2. סיכויים של 2 ל-1 פירושם שני סיכויים לזכות עבור כל סיכוי להפסיד, כך שהסתברות לזכות עם 28 ביטים של ראיות היא $\frac{2}{3}$. הוספת קופסה נוספת, עוד 2 ביטים של ראיות, תעלה את הסיכויים ל-1:8. הוספת עוד שתי קופסאות תעלה את הסיכוי לזכות ל-1:128.

אז אם אתה רוצה להצדיק אמונה חזקה לגבי האפשרות שתזכה בהגרלה – המוגדרת באופן שרירותי כבעלת פחות מ-1% הסתברות לטעות – 34 ביטים של ראיות לגבי הצירוף המנצח אמורים להספיק. באופן כללי, הכללים לשקילת "כמה ראיות נדרשות" פועלים על פי דפוס דומה: ככל שמרחב האפשרויות בו נמצאת ההשערה גדול יותר, או ככל שההשערה נראית מראש פחות סבירה בהשוואה לשכנותיה, או ככל שאתה רוצה להיות בטוח יותר, כך תזדקק ליותר

$$\left(\frac{1}{2}\right)^3 = \frac{1}{8} \text{ שהרי } \log_{\frac{1}{2}}\left(\frac{1}{8}\right) = 3 \quad *$$

ראיות.

אינך יכול להפר את הכללים: אינך יכול ליצור אמונות מדויקות על סמך ראיות שאינן מספקות. נניח שיש לך 10 קופסאות מסודרות בשורה, ואתה מתחיל להזין לתוכן צירופים. אינך יכול לעצור בצירוף הראשון שמקבל צפצופים מכל 10 הקופסאות, ולומר, "אבל הסיכויים שזה יקרה עבור צירוף מפסיד הם אחד למיליון! אני פשוט אתעלם מהחוקים הבייסיאניים המתנשאים האלו ואעצור כאן." בממוצע, 131 כרטיסים מפסידים יעברו מבחן כזה עבור כל כרטיס מנצח. בהתחשב במרחב האפשרויות ובהסתברות הנמוכה מלכתחילה, קפצת למסקנה חזקה מדי על סמך ראיות לא מספיקות. זו לא תקנה בירוקרטית חסרת טעם. זו מתמטיקה.

אתה עדיין יכול להאמין, כמובן, על סמך ראיות לא מספיקות, אם זה רצונך. אבל לא תוכל להאמין באופן מדויק. יהיה זה כמו לנסות לנהוג במכונית ללא דלק, כי אתה לא מאמין בקשקוש המקושקש הזה שצריך דלק כדי לנסוע. יהיה זה הרבה יותר כסף, והרבה יותר זול, אם פשוט נחליט לבטל את החוק לפיו מכוניות זקוקות לדלק. האין זה פשוט טוב יותר עבור כולם? ובכן, אתה יכול לנסות, אם זה רצונך. אתה יכול אפילו לעצום את עיניך ולדמיין שהמכונית זזה. אבל כדי להגיע באמת לאמונות מדויקות דרוש דלקראיות, וככל שתרצה להגיע רחוק יותר, כך תזדקק ליותר דלק.

24

יהירותו של איינשטיין

בשנת 1919, יצא סר ארתור אֵדינגְטון בראשן של משלחות מחקר לברזיל ולאי פֶרינסיפָה, במטרה לבחון באופן ניסויי את תורת היחסות הכללית החדשה של איינשטיין בעזרת צפייה בליקוי חמה.* עיתונאי כלשהו שאל את איינשטיין מה הוא יעשה אם תצפיותיו של אדינגטון לא יאששו את התאוריה שלו. איינשטיין ענה בתשובה האייקונית: "אז אצטער עבור האל הטוב. התאוריה נכונה."

נשמע כמו הצהרה נמהרת למדי, שמתריסה נגד עקרון היסוד של הרציונליות המסורתית, המעניק לניסוי את זכות המילה האחרונה. נראה כאילו איינשטיין נתפס ליהירות כה גדולה, שהוא סירב לכופף את ראשו ולהיכנע לדבר הטבע, כפי שמכתיבה חובתו כמדען. מי יכול לדעת שהתאוריה נכונה, לפני שהועמדה לניסוי?

הסתבר, כמובן, שאיינשטיין צדק. אני משתדל להימנע מביקורת על אנשים כשהם צודקים. אם הם באמת ראויים לביקורת, לרוב לא אצטרך לחכות זמן רב להזדמנות שבה יטעו.

ואולי איינשטיין לא היה כה נמהר כפי שנשמע...

כדי להקצות הסתברות של יותר מ-50% למועמד הנכון מתוך 100,000,000 השערות אפשריות, נדרשים לפחות 27 ביטים של ראיות (פחות או יותר). אי אפשר לצפות למצוא את המועמד הנכון

* המטרה היתה לבדוק האם האור אכן מושפע מכבידה, וספציפית, האם אורם של כוכבים רחוקים "מתעקם" בדרך אלינו כשהוא עובר בסמוך לשמש.

ללא מבחנים חזקים באותה מידה, משום שמבחנים חלשים יותר ינפקו יותר ממועמד אחד שעבר אותם. אם תנסה להחיל מבחן עם סיכוי של אחד למיליון לתוצאה שגויה (כ-20 ביטים), תישאר בסוף עם מאה מועמדים. עצם מציאת התשובה הנכונה, בתוך מרחב עצום של אפשרויות, דורשת כמות גדולה של ראיות.

הרציונליות המסורתית שמה דגש על הצדקה: "אם אתה רוצה לשכנע אותי ב-X, עליך להציג לי כמות של Y ראיות." אני עצמי מועד בלשוני לא אחת ומתבטא כך, כשאני אומר משהו בסגנון, "כדי להצדיק אמונה בהצעה זו בהסתברות של מעל 99%, נדרשים 34 ביטים של ראיות." או, "כדי להקצות מעל 50% הסתברות להשערתך, תזדקק ל-27 ביטים של ראיות." הניסוח המסורתי מרמז על כך שאתה מתחיל עם תחושה, או קו חשיבה אישי שמוביל להשערה המוצעת, ואז עליך לאסוף "ראיות" כדי לאשר אותה – לשכנע את הקהילה המדעית, או להצדיק את האמונה בתחושה שלך.

אך מנקודת מבט בייסיאנית, אתה זקוק לכמות ראיות השקולה בערך למורכבות ההשערה רק כדי לאתר את ההשערה במרחב התיאוריות. זה לא עניין של להצדיק משהו בפני מישור. אם יש מאה מיליון חלופות, תזדקק לפחות ל-27 ביטים של ראיות רק כדי למקד את תשומת הלב שלך באופן בלעדי בתשובה הנכונה.

התהליך הזה נכון גם אם תקרא לניחוש שלך "תחושה" או "אינטואיציה". תחושות ואינטואיציות הן תהליכים אמיתיים המתרחשים במוחות אמיתיים. אם המוח שלך לא מחזיק לפחות 10 ביטים של ראיות בייסיאניות תקפות ושזורות איתן יוכל לעבוד, אז הוא לא יוכל לבודד השערה נכונה בת 10 ביטים לתשומת לבך – לא באופן מודע ולא באופן לא-מודע. תהליכים של התת-מודע לא יכולים למצוא מטרה אחת מתוך מיליון אפשרויות עם 19 ביטים של שזירה, בדיוק כפי שתהליכים מודעים לא יכולים. תחושות-בטן עשויות להיות מסתוריות לבעל הבטן, אבל הן לא יכולות להפר את חוקי הפיזיקה. אתם מבינים לאן זה הולך: בזמן שניסח את ההשערה לראשונה – בפעם הראשונה שבה צצו המשוואות בראשו – איינשטיין כבר החזיק

בידיו מספיק ראיות תצפיתיות כדי לבודד את משוואות היחסות הכללית לתשומת לבו הבלעדית. אחרת, הוא לעולם לא היה מנחש אותן נכון.

עכשיו, מה הסיכוי שאיינשטיין החזיק בידיו בדיוק מספיק ראיות תצפיתיות כדי להביא את היחסות הכללית לתשומת הלב שלו, אבל להצדיק לה הקצאה של 55% הסתברות בלבד? נניח שהיחסות הכללית היא השערה בת 29.3 ביטים. כמה סביר שאיינשטיין נתקל בדיוק ב-29.5 ביטים של ראיות במהלך חקירותיו בתורת הפיזיקה? לא סביר! אם לאיינשטיין היו מספיק ראיות תצפיתיות כדי לבודד את משוואות היחסות הכללית מלכתחילה, היו לו כנראה מספיק ראיות כדי להיות בטוח למדי שתאוריית היחסות הכללית נכונה. למעשה, מאחר שהמוח האנושי לא מעבד מידע בצורה מושלמת, לאיינשטיין היו כנראה הרבה יותר ראיות ממה שנדרש, באופן עקרוני, לבייסיאני מושלם כדי להקצות ביטחון עצום ליחסות הכללית. "אז אצטער עבור האל הטוב. התאוריה נכונה." זה לא נשמע מזעזע כלל, כשמסתכלים על זה מנקודת מבט כזו. וזכרו, היחסות הכללית הייתה נכונה, מתוך כל מרחב האפשרויות העצום הזה.

25

תערו של אוקאם

ככל שהסבר כלשהו הוא מורכב יותר, כך נדרשות יותר ראיות רק כדי לאתרו בתוך מרחב האמונות האפשריות. (ברציונליות מסורתית, עקרון זה מנוסח לעיתים בצורה המטעה של: "ככל שהצהרה היא מורכבת יותר, כך נדרשות יותר ראיות כדי לתמוך בה.") אבל כיצד ניתן למדוד את מורכבותו של הסבר? כיצד ניתן לקבוע כמה ראיות נדרשות?

תערו של אוקאם מנוסח לעיתים קרובות בתור "ההסבר הפשוט ביותר שמתאים לעובדות". רוברט היינליין השיב לכך באמירה שההסבר הפשוט ביותר הוא "הגברת מהמשך הרחוב היא מכשפה. היא עשתה זאת". חדי העין יבחינו בכך שאורכו של משפט באנגלית לא מהווה מדד טוב ל"מורכבות". ו"התאמה" לעובדות רק על ידי הימנעות משלילתן אינה מספיקה.

אבל למה בעצם אורכו של משפט באנגלית הוא מדד גרוע למורכבות? משום שכאשר אתה הוגה משפט בקול רם, אתה משתמש בתוויות למושגים שהמאזין מכיר – המקבל כבר טען אותם במורכבות בעצמו. נניח שנקצר את משפטו של היינליין ל"המההמהע"ז!" כך שההסבר כולו מועבר במילה אחת בלבד. אולי אפילו ניתן לו תווית שרירותית קצרה יותר, כמו "פנורד!" האם קיצור זה מפחית את המורכבות? לא, משום שיש צורך לספר למאזין מראש ש"המההמהע"ז!" מייצג את "הגברת מהמשך הרחוב היא מכשפה.

היא עשתה זאת". המילה "מכשפה" לכשעצמה מהווה תווית לטענות יוצאות דופן – העובדה שכולנו מבינים למה היא מכוונת לא הופכת את המושג לפשוט.

הבזק עצום של חשמל נורה מהשמיים ופוגע במשהו, וילידים נורדיים אומרים: "אולי סוכן חזק כעס מאוד והטיל מכת ברק." המוח האנושי הוא המבנה המורכב ביותר ביקום הידוע. אם כעס נראה לנו פשוט, זה רק מפני שאיננו רואים את כל המעגלים העצביים מהם הרגש מורכב. (דמיינו לנסות להסביר לחיזורים ללא חוש הומור מדוע סאטרדיי נייט לייב מצחיק. אבל אל תעופו על עצמכם. לכם אין שום חוש לפנורד.) המורכבות של רגש הכעס, ובמובלע גם המורכבות של אינטליגנציה, נעלמה מעיני האנשים שהניחו את קיומו של ת'ור, סוכן הרעם.

לכן אנוש, משוואות מקסוול* דורשות הסבר ארוך בהרבה מאשר ת'ור. כבני אדם, אין לנו אוצר מילים מובנה לחשבון דיפרנציאלי כמו שיש לנו לכעס. עליך להסביר את השפה, ואת השפה שמאחורי השפה, ואת עצם המושג של מתמטיקה, לפני שתוכל אפילו להתחיל לדבר על חשמל.

ובכל זאת, נראה שצריך להיות מובן כלשהו שבו משוואות מקסוול הן פשוטות יותר ממוח אנושי, או מת'ור סוכן הרעם.

האמת שיש. מסתבר שכתובה של תוכנת מחשב המדמה את משוואות מקסוול היא דבר קל לאין שיעור, לעומת תוכנה המדמה מוח אינטליגנטי ורגשי כמו של ת'ור.

הפורמליזם של אינדוקציית סולומונוף מודד את "מורכבות התיאור" לפי אורכה של התוכנה הקצרה ביותר שמייצרת כפלט את אותו תיאור. כדי לדבר על "התוכנה הקצרה ביותר" שעושה משהו, יש להגדיר מרחב של תוכנות מחשב, שדורש שפת תוכנות ומפרש**. אינדוקציית סולומונוף משתמשת במכונות טיורינג, או ליתר דיוק,

* משוואות מקסוול, יחד עם משוואות ניוטון למקורות, מהוות תיאור מתמטי שלם של תורת החשמל והמגנטיות הקלאסית (ויקיפדיה).

** תוכנה המתרגמת ומבצעת את קוד המקור של שפת התיכנות (Interpreter).

אחרת של 20 הטלות – לא פחות ולא יותר – אך במקרה הזה אנו רואים השערה אחרת, שבעצמה לא נראית מסובכת מדי, ומתאימה לנתונים הרבה יותר טוב.

בכל פעם שמאפשרים לתוכנה לאחסן ביט אחד נוסף של מידע בינארי, היא תוכל לחתוך את מרחב האפשרויות בחצי, וכתוצאה מכך להקצות פי שניים הסתברות לכל הנקודות במרחב שנותר. אפקט זה מעלה את המחשבה על כך שביט אחד של מורכבות תוכנה צריך לעלות לפחות "פקטור של פי שניים" בהתאמה. אם אנו מנסים לתכנן תוכנה שמאחסנת במפורש תוצאה כמו עפפפעפ, ששת הביטים שאיבדנו במורכבות חייבים להרוס את כל הסבירות שהושגה משיפור של פי 64 בהתאמה. אחרת, במוקדם או במאוחר נאלץ להחליט שכל המטבעות ההוגנים הם קבועים.

אלא אם כן התוכנה חכמה ודוחסת את הנתונים, לא תצמח שום תועלת מהעברת ביט אחד מהנתונים לתיאור התוכנה.

אופן הפעולה של אינדוקציית סולומונוף לחיזוי רצפים הוא סכימה של כל התוכנות המותרות – אם כל התוכנות מותרות, האינדוקציה הופכת לבלתי ניתנת לחישוב – כאשר לכל תוכנה יש הסתברות מוקדמת של $\frac{1}{2}$ בחזקת אורך הקוד שלה בביטים, וכל תוכנה משוקללת לפי התאמתה לכל הנתונים שנצפו עד כה. שיטה זו נותנת לנו מעין תערובת משוקללת של מומחים שמסוגלת לחזות ביטים עתידיים.

הפורמליזם של אורך ההודעה המינימלי זהה לאינדוקציית סולומונוף כמעט לחלוטין. אתה שולח מחרוזת שמתארת קוד, ואז מחרוזת שמתארת את הנתונים באותו הקוד. ההסבר שמוביל להודעה הכוללת הקצרה ביותר מוכתר כטוב ביותר. אם חושבים על קבוצת הקודים המותרים כמרחב של תוכנות מחשב, ועל שפת תיאור הקוד כמכונה אוניברסלית, אז אורך ההודעה המינימלי הוא כמעט זהה לאינדוקציית סולומונוף. (כמעט, כי הוא בוחר את התוכנה הקצרה ביותר, במקום לסכום את כל התוכנות.)

כל זה מאפשר לנו לראות בבירור את הבעייתיות שבשימוש

ב"הגברת מהמשך הרחוב היא מכשפה. היא עשתה זאת" על מנת להסביר את התבנית ברצף 0101010101. אם אתה שולח הודעה לחבר ומנסה לתאר את הרצף שראית, תצטרך לומר: "הגברת מהמשך הרחוב היא מכשפה. היא גרמה לרצף לצאת 0101010101." אישום בכישוף לא יאפשר לך לקצר את שאר ההודעה. עדיין תצטרך לתאר במלוא הפירוט את הנתונים להם גרם הכישוף.

כישוף עשוי להתאים לתצפיות שלנו במובן של אי־שלילה שלהן. אבל זה רק מפני שכישוף מאפשר הכל, כמו להכריז "פְּלוֹגִיסְטוֹן!" באותו אופן, גם לאחר שאתה מכריז "מכשפה", עדיין תצטרך לתאר במלוא הפירוט את כל הנתונים שנצפו. לא דחסת את האורך הכולל של ההודעה שמתארת את התצפיות שלך על ידי שליחת ההודעה לגבי הכישוף. כל שעשית היה להוסיף הקדמה חסרת משמעות, שרק הגדילה את האורך הכולל של ההודעה.

הערמומיות האמיתית הוסתרה במילה "זאת" במשפט "מכשפה

עשתה זאת". מכשפה עשתה מה ?

כמובן, הודות להטיית החכמה בדיעבד, עיגון, הסברים מזויפים, סיבתיות מזויפת, הטיית האישוש וקוגניציה מוטיבציונית, זה עשוי להיראות שקוף לחלוטין שאם אישה היא מכשפה, אז ברור שהיא תגרום למטבע לצאת 0101010101. אבל אגיע לזה בקרוב...

26

כוח העל שלך כרציונליסט

המקרה שאתאר להלן קרה לי בחדר צ'אט של IRC*, לפני זמן רב כל כך, שעוד נהגתי להסתובב בחדרי צ'אט של IRC. הזמן טישטש את הזיכרון וייתכן שהדיווח שלי לא יהיה מדויק.

אז הסתובבתי שם, בחדר צ'אט של IRC, ומישהו כתב שחבר שלו צריך ייעוץ רפואי. החבר סיפר שהוא חש כאבים פתאומיים בחזה, אז הוא הזמין אמבולנס. האמבולנס הגיע, אבל הפרמדיקים אמרו לו שזה שום דבר ונסעו משם, ועכשיו הכאבים בחזה מחמירים. מה החבר שלו אמור לעשות?

הסיפור הזה בלבד אותי. נזכרתי שקראתי על חסרי בית בניו יורק שנוהגים להזמין אמבולנסים רק כדי להעביר את הלילה במקום חם, ושהפרמדיקים תמיד מחוייבים לקחת אותם לחדר המיון, גם בפעם ה-27. אם הם לא יקחו אותם, חברת האמבולנסים עלולה להיתבע על סכומי עתק. באופן דומה, חדרי מיון מחוייבים על פי חוק לטפל בכל אדם, ללא קשר ליכולת התשלום שלו. (ובתי החולים סופגים את העלויות העצומות, אז בתי חולים סוגרים את חדרי המיון שלהם... מה שמעלה את התהיה – מה הטעם להעסיק כלכלנים אם בכל מקרה אנחנו מתעלמים מהם.) אז לא ממש הבנתי איך שרשרת האירועים שתוארו יכולה להתרחש. כל מי שמדווח על כאבים פתאומיים בחזה

* גלגול מוקדם של חדרי צ'אט.

אמור להילקח באמבולנס ללא שאלות. וכאן כשלת כרציונליסט. נזכרתי במספר מקרים שבהם הרופא שלי לא התרגש כלל מדיווח על תסמינים שלי באופן אישי נראו מדאיגים מאוד, ובכל אותם מקרים הצדק היה עם נציג הממסד הרפואי. בכל פעם ופעם. גם לי היו פעם כאבים בחזה, והרופא הסביר לי בסבלנות שמדובר בכאב בשרירי-החזה, ולא בהתקף לב. אז כתבת בצ'אט: "טוב, אם הפרמדיקים אמרו לחבר שלך שזה שום דבר, זה כנראה כאמת שום דבר. אם היה ולו סיכוי קל שבקלים לבעיה רצינית – הם היו לוקחים אותו מיד."

ככה הצלחתי להסביר את הסיפור בתוך המודל הקיים שלי, למרות שההסבר עדיין הרגיש קצת מאולץ...

מאוחר יותר, הבחור חזר לצ'אט ואמר שהחבר שלו המציא את כל הסיפור. מסתבר שהוא לא היה החבר הכי אמין בחבורה.

הייתי צריך להבין, אולי, שמכר לא ידוע של מכר בחדר צ'אט של IRC עשוי להיות אמין פחות ממאמר שפורסם בכתב עת. אך אבוי, קל יותר להאמין מאשר לא להאמין. אנחנו מאמינים באופן אינסטינקטיבי, אבל איאמונה דורשת מאמץ מודע.¹

אז במקום זאת, תוך התפתלויות רבות, כפיתי על מודל-המציאות שלי להסביר אנומליה שמעולם לא קרתה. וידעתי כמה מביך זה היה. ידעתי שערכו של מודל לא נמדד במה שהוא יכול להסביר, אלא במה שהוא לא יכול. השערה שאינה אוסרת דבר, למעשה מאפשרת הכול, ולכן אינה מצליחה לצמצם את מרחב הציפיות.

כוח העל שלך כרציונליסט הוא היכולת שלך להיות מבולבל יותר מבדיה מאשר ממציאות. אם אתה מסוגל להסביר באותה מידה כל תוצאה, יש לך אפס ידע.

כולנו חלשים, מפעם לפעם. החלק העצוב הוא שיכולתי להיות חזק יותר. היה בידי כל המידע שהייתי צריך כדי להגיע לתשובה הנכונה, אפילו שמתי לב לבעיה, ואז התעלמתי ממנה. תחושת הבלבול שלי

הייתה רמוז, ואני התעלמתי מהרמז הזה.*
הייתי צריך לייחס חשיבות רבה יותר לתחושה הזו של עדיין
מרגיש קצת מאולץ. זו אחת התחושות החשובות ביותר עבור אדם
המחפש אחר האמת, חלק מכוח העל שלך כרציונליסט. פגם עמוק
בתכנון הקוגניטיבי האנושי הוא שהתחושה הזו מתבטאת רק כמתח
שקט באחורי הראש, במקום כצופר אזעקה מהדהד ושלט ניאון זוהר
שאומר:

או שהמודל שלך שגוי, או שהסיפור הזה לא נכון.

1. Daniel T. Gilbert, Romin W. Tatarodi, and Patrick S. Malone, "You Can't Not Believe Everything You Read," *Journal of Personality and Social Psychology* 65 (2 1993): 221-233, doi:10.1037/0022-3514.65.2.221.

* "...ובמקום כלשהו באחורי תודעתו נשמע צליל קטנטן של בלבול, תחושה שמשהו לא בסדר בסיפור הזה; וזה היה אמור להיות חלק מהאמנות של הארי, לשים לב לצליל הקלוש הזה, אבל דעתו הייתה מוסחת. שהרי זהו כלל עצוב שבדיוק ברגע שבו אתה זקוק ביותר לאומנותך כרציונליסט, חזקה עליך שתשכח אותה."

היעדר ראיות הוא ראיה להיעדר

מתוך ספרו של רוברט דוס, "בחירה רציונלית בעולם של חוסר ודאות":

למעשה, התאמה בדיעבד של ראיות לתיאוריה הייתה מעורבת באחד הפרקים העגומים בהיסטוריה של ארצות הברית: כליאתם של אמריקאים ממוצא יפני בתחילת מלחמת העולם השנייה. כאשר מושל קליפורניה דאז, אָרל וורן, שהעיד בשימוע בפני הקונגרס בסן פרנסיסקו ב-12 בפברואר 1942, נשאל על כך שלא תועדו שום מעשי חבלה או ריגול מצד אמריקאים ממוצא יפני עד אותו מועד. על כך השיב וורן: "אני סבור שהיעדר זה [של פעילות חתרנית] הוא הסימן המאיים ביותר בכל המצב הזה. הוא משכנע אותי יותר מכל גורם אחר שהחבלה שאנו עתידים לחוות, פעילויות הגֵיס החמישי הצפויות לנו, הן מתוזמנות ומתואמות בדיוק כפי שתוזמן ותואם פרל הארבור... אני מאמין שפשוט מרדמים אותנו בתחושת ביטחון כוזבת."

הבה נבחן את טיעונו של וורן מנקודת מבט בייסיאנית. כאשר אנו מבחינים בראיות, להשערות שייחסו סבירות גבוהה יותר לראיות אלו תוקצה סבירות רבה יותר על חשבון השערות שייחסו סבירות נמוכה יותר לאותן ראיות. זוהי התופעה של סבירויות יחסיות והסתברויות יחסיות. ניתן לייחס סבירות גבוהה לראיות ועדיין 'לאבד נקודות' להשערה אחרת, אם אותה השערה מייחסת סבירות גבוהה עוד יותר. וורן טוען, ככל הנראה, שבהינתן שלא נצפו מעשי חבלה, הדבר

מאשש את קיומו של גֵּיס חמישי. ניתן לטעון שגֵּיס חמישי עשוי לעכב את פעולות החבלה שלו, אך הסבירות שבהיעדר גֵּיס חמישי פשוט לא יתרחשו מעשי חבלה גבוהה יותר מאשר הסבירות שבהינתן קיומו של גֵּיס חמישי לא יתרחשו מעשי חבלה.

נניח ש- E מייצג תצפית של מעשי חבלה, ו- $\neg E$ מייצג תצפית של היעדר חבלה. הסמל H_1 מייצג את ההשערה של קיום גֵּיס חמישי יפני-אמריקאי, ו- H_2 מייצג את ההשערה שאין שום גֵּיס חמישי. ההסתברות המותנית $P(E|H)$, או " E בהינתן H ", היא מידת הביטחון שבה היינו מצפים לראות את הראיה E אם נניח את נכונותה של השערה H . לא משנה מהי הסבירות שגֵּיס חמישי לא יבצע חבלה, $P(\neg E|H_1)$, היא לעולם לא תהיה גבוהה כמו הסבירות שבהיעדר גֵּיס חמישי לא יתרחשו מעשי חבלה, $P(\neg E|H_2)$. לפיכך, תצפית של היעדר חבלה מגבירה את ההסתברות לכך שאין גֵּיס חמישי.

היעדר חבלה לא מוכיח שאין גֵּיס חמישי. היעדר הוכחה לא מהווה הוכחה להיעדר. בלוגיקה, $(A \Rightarrow B)$, שפירושו " A גורר B ", אינו שקול ל- $(\neg A \Rightarrow \neg B)$, שפירושו " A גורר לא- B ". אך בתורת ההסתברות, היעדר ראיות הוא תמיד ראיה להיעדר. אם E הוא אירוע בינארי ו- $P(H|E) > P(H)$, כלומר, התצפית על E מגבירה את ההסתברות של H , אזי $P(H|\neg E) < P(H)$, כלומר, היעדרה של תצפית על E מפחיתה את ההסתברות של H . ההסתברות $P(H)$ היא שילוב משוקלל של $P(H|E)$ ו- $P(H|\neg E)$, ונמצאת בהכרח בין שניהם. אם כל זה נשמע לכם מבלבל, עיינו באינטרלוד הסבר אינטואיטיבי למשפט בייס*.

ברוב המכריע של המקרים בחיים האמיתיים, סיבה עשויה שלא להפיק סימנים לקיומה באופן עקבי, אך סביר אפילו פחות המצב בו היעדר הסיבה יפיק את הסימנים הללו. היעדר תצפית עשוי להוות ראיה חזקה או ראיה חלשה מאוד להיעדר, תלוי עד כמה זה סביר שהסיבה תייצר את התצפית. היעדר תצפית שמותרת רק באופן חלש

* מקומו של אינטרלוד זה בכרך השלישי של רציונליות מ-א' ועד ת'.

גם אם ההשערה החלופית לא מאפשרת אותה כלל) הוא ראיה חלשה מאוד להיעדר (אם כי עדיין ראיה). זו הטעות של "חורים ברצף המאובנים"* – מאובנים נוצרים לעיתים נדירות בלבד. יהיה זה חסר תועלת להכריז על היעדר תצפית שמותרת באופן חלש, כאשר תצפיות חיוביות רבות כל כך כבר תועדו. אך אם אין תצפיות חיוביות בכלל, זה הזמן להתחיל לדאוג. מכאן נובע הפרדוקס של פרמי.** כוח העל שלך כרציונליסט הוא היכולת שלך להיות מבולבל יותר מבידיה מאשר ממציאות. אם אתה מסוגל להסביר באותה מידה כל תוצאה, יש לך אפס ידע. חוזקו של מודל לא נמדד במה שהוא יכול להסביר, אלא במה שהוא לא יכול, שכן רק איסורים מסוגלים לצמצם את מרחב הציפיות. אם אינך מבחין כאשר המודל שלך הופך את הראיות לבלתי סבירות, באותה מידה יכולת שלא להחזיק בשום מודל, ובאותה מידה יכולת שלא להחזיק בראיות. לא מוח ולא עיניים.

1. Daniel T. Gilbert, Romin W. Tafarodi, and Patrick S. Malone, "You Can't Not Believe Everything You Read," *Journal of Personality and Social Psychology* 65 (2 1993): 221-233, doi:10.1037/0022-3514.65.2.221.

* הביטוי *חורים ברצף המאובנים* ("gaps in the fossil record") מתאר פערים ברצף האבולוציוני התיאורטי, כפי שהוא משתקף בפריטי המאובנים שנמצאו. למשל – מין שאמור "לגשר" אבולוציונית בין אב קדמון לבין מין שחי כיום, אבל לא נמצא לו שום תיעוד בשכבות מאובנים. הביטוי משמש גם כמטבע לשון, בין השאר גם בקרב בריאתנים ומכחישי אבולוציה, כהוכחה לכך שתיאוריית האבולוציה לא מסוגלת להסביר את המציאות.

** איך זה שיש כל כך הרבה כוכבים, אבל אין שום סימן לחיזורים.

28

חוק שימור הראיות הצפויות

פרידריך שְפֶה פון לְאָנְגְנֶפֶלד, כומר ששמע את הווידויים של מכשפות שנידונו למוות, כתב ב-1631 את ה-*Cautio Criminalis* ("זהירות במקרים פליליים"), שבו תיאר בחריפות את עץ ההחלטות להרשעת נשים שמואשמות בכישוף: אם המכשפה ניהלה חיים מרושעים ולא ראויים, היא אשמה; אם היא ניהלה חיים טובים וראויים, גם זו הוכחה לאשמתה, כי מכשפות נוהגות להתחזות ומנסות להיראות צדיקות במיוחד. לאחר שהאישה נכלאה: אם פחדה, סימן שהיא אשמה; אם לא פחדה, סימן שהיא אשמה, כי מכשפות ינהגו כחפות מפשע וישמרו על חזות איתנה. או בזמן שהיא שומעת על האשמתה בכישוף, היא עשויה לברוח או להישאר: אם ברתה, סימן שהיא אשמה; אם נשארה, השטן עיכב אותה כדי שלא תוכל לברוח.

שְפֶה ערך וידויים למכשפות רבות. אשר על כן, הוא ישב בעמדה ממנה יכל לצפות בכל ענף של עץ ההאשמות, כך שיהיה מה שיהיה מה שהנאשמת אמרה או עשתה, יחשב הדבר כהוכחה נגדה. בכל מקרה בודד, תוכל לשמוע רק ענף אחד של הדילמה. זו הסיבה שבעטיה מדענים רושמים את התחזיות הניסוייות שלהם מבעוד מועד.

אבל אי אפשר לאחוז במקל משני קצותיו – כמאפיין של תורת ההסתברות, גם אם לא מצד ההגינות. הכלל של "היעדר ראיות הוא ראיה להיעדר" הוא מקרה פרטי של חוק כללי יותר, אותו אכנה חוק שימור הראיות הצפויות: הציפייה של ההסתברות בדיעבד, לאחר

שצפית בראיות, חייבת להיות שווה להסתברות שחושבה טרם עריכת הניסוי.

$$P(H) = P(H, E) + P(H, -E)$$

$$P(H) = P(H|E) \times P(E) + P(H|-E) \times P(-E)$$

לכן, עבור כל ציפייה לראיות, ישנה ציפייה שווה ומנוגדת לה לראיות נגדיות.

אם אתה מצפה להסתברות גבוהה למציאת ראיות חלשות בכיוון אחד, הציפייה הזו חייבת להתאזן בציפייה חלשה למציאת ראיות חזקות בכיוון ההפוך. אם יש לך בטחון גבוה בתיאוריה שלך, ובהתאם לכך אתה מצפה לראות תוצאה התואמת את ההשערה שלך, זה יכול לספק רק תוספת קטנה מאוד לאמונה שלך (משום שהיא גם ככה קרובה ל-1), אבל כישלון בלתי צפוי של התחזית שלך יצטרך (בהכרח) לפגוע קשות בביטחונך. בממוצע, עליך לצפות להיות בדיוק באותה רמת ביטחון שהיתה לך כשהתחלת. במילים אחרות, עצם הציפייה לפגוש ראיות – לפני שראית אותן בפועל – לא אמורה לשנות את האמונות המקוריות שלך. (שוב, אם זה לא אינטואיטיבי, עיינו באינטרלוד "הסבר אינטואיטיבי למשפט בייס").

אז אם אתה טוען ש"היעדר חבלה" מהווה ראיה לקיומו של ג'יס חמישי, עליך להחזיק מנגד בעמדה לפיה כל גילוי של מעשה חבלה יפעל נגד הטענה בעד קיומו של ג'יס חמישי. אם אתה טוען ש"חיים טובים וראויים" מהווים ראיה לכך שאישה היא מכשפה, אזי חיים מרושעים ולא ראויים מוכרחים להוות ראיה לכך שהיא אינה מכשפה. אם אתה טוען שאלוהים, במטרה לבחון את אמונתם של בני האדם, מסרב לחשוף את קיומו, אזי הניסים המתוארים בתנ"ך חייבים להוות טיעון נגד קיומו של אלוהים.

זה לא נשמע לגמרי נכון, הלא כן? שים לב לתחושה הזו של "זה נשמע קצת מאולץ", המתח השקט הזה בחלק האחורי של הראש. הוא חשוב.

עבור רציונליסט בייסיאני אמיתי, חיפוש אחר ראיות המאששות תיאוריה הוא מעשה בלתי אפשרי. אין כל תוכנית שתוכל לתכנן, שום

אסטרטגיה מתוחכמת או מנגנון ערמומי, שבאמצעותם תוכל לצפות באופן לגיטימי שרמת הביטחון שלך בתאוריה כלשהי יהיה גבוה יותר (בממוצע) מאשר לפני כן. באפשרותך רק לחפש ראיות לבדיקה של תיאוריה, לא לאישורה.

זו תובנה שבכוחה להוריד הרבה עומס מהתודעה שלך. היא משחררת אותך מהצורך לפרש כל תוצאה ניסויית אפשרית כך שתאשש את התיאוריה שלך. אין צורך לתכנן כיצד לגרום לכל בדל ראיה לאשש את התיאוריה שלך, משום שאתה יודע שאל מול כל ציפייה לראיות, קיימת ציפייה שווה ומנוגדת לראיות נגדיות. אם תנסה להחליש את הראיות הנגדיות של תצפית "לא נורמלית", תימצא מחליש בעל כורחך גם את התמיכה בתצפית "נורמלית", בדיוק באותה המידה. זה משחק סכום אפס. לא משנה כיצד תתכנן, לא משנה כיצד תטען, לא משנה באיזו אסטרטגיה תבחר, אינך יכול לצפות שהתוכנית שיצרת תשנה את אמונותיך (בממוצע) בכיוון מסוים.

יהיה אפקטיבי באותה מידה פשוט להישען לאחור בשלווה בזמן שאתה מחכה להופעתן של ראיות.
...הפסיכולוגיה האנושית כל כך דפוקה.

29

החוכמה בדיעבד מפחיתה מערכו של המדע

מאמר זה מבוסס על קטע מתוך *Exploring Social Psychology*¹ של מאיירס; הקטע כולו שווה קריאה לכשעצמו. קאלן מרפי, עורך *The Atlantic*, ציין שהמדעים החברתיים לא מגלים "שום רעיונות או מסקנות שלא ניתן למצוא באנציקלופדיה של ציטוטים... יום אחר יום, חוקרי מדעי החברה יוצאים לעולם. יום אחר יום הם מגלים שהתנהגותם של אנשים היא פחות או יותר מה שהיית מצפה שהיא תהיה".

כמובן, ה"ציפייה" הזו מורכבת כולה מחוכמה בדיעבד. (הטיית החוכמה בדיעבד: נבדקים שיודעים את התשובה האמיתית לשאלה כלשהי מייחסים הסתברות גבוהה בהרבה לאפשרות שהם "היו מצליחים" לנחש את התשובה, בהשוואה לנבדקים שצריכים לנחש את התשובה מבלי לדעת אותה מראש.)

ההיסטוריון ארתור שלזינגר ג'וניור תיג מחקרים מדעיים על חוויותיהם של חיילים במלחמת העולם השנייה כ"הדגמות מסורבלות" של היגיון בריא. לדוגמה:

1. חיילים משכילים יותר סבלו יותר מבעיות הסתגלות בהשוואה לחיילים משכילים פחות. (אינטלקטואלים היו פחות מוכנים ללחצים של שדה הקרב מאשר אנשים עם חוכמת רחוב.)

2. חיילים דרומיים התמודדו טוב יותר עם האקלים החם של איי דרום האוקיינוס השקט מאשר חיילים צפוניים. (דרומיים רגילים יותר למזג אוויר חם.)
3. טוראים לבנים היו להוטים יותר לקבל קידום לדרגות פיקוד מאשר טוראים שחורים. (שנים של דיכוי משפיעות על דחף ההישגיות.)
4. שחורים דרומיים העדיפו קצינים לבנים דרומיים על פני קצינים לבנים צפוניים. (קצינים דרומיים היו מנוסים יותר באינטראקציה עם שחורים.)
5. כל עוד נמשכו הקרבות, חיילים היו להוטים יותר לחזור הביתה מאשר לאחר סיום המלחמה. (בזמן הקרבות, החיילים ידעו שהם בסכנת מוות.)

כמה מהממצאים האלו אתה חושב שהיית מנחש בעצמך? שלושה מתוך חמישה? ארבעה מתוך חמישה? האם יש מקרים שבהם היית מנחש את ההפך – מקרים שבהם המודל שלך נפגע? חשוב רגע לפני שתמשיך...

...

בהדגמה זו (של פול לאַזרספלד, אותה מצטט מאירס), כל הממצאים לעיל מתארים את ההפך ממה שנמצא בפועל.² כמה פעמים חשבת שהמודל שלך נפגע? כמה פעמים הודית בפני עצמך שהיית טועה? זהו המדד האמיתי לאיכותו של המודל שלך. כוח העל שלך כרציונליסט הוא היכולת להיות מבולבל יותר מבדיה מאשר ממצאות.

אלא אם כן, כמוכן, הפכתי שוב את התוצאות. מה אתה חושב? האם תהליכי החשיבה שלך כעת, כשאתה באמת לא יודע את התשובה, הם שונים מתהליכי החשיבה בהם השתמשת כדי להצדיק

כל צד של התשובה "הידועה"?

דפנה בֶּרֶץ חשפה סטודנטים לזוגות של ממצאים לכאורה, אחד אמיתי ("בתקופות של שגשוג כלכלי אנשים מוציאים חלק גדול יותר מהכנסתם מאשר בתקופות מיתון") ואחד הופכי לאמת.³ עבור שני ההיגדים, הסטודנטים דירגו את הממצא לכאורה בתור "מה שהם היו מנחשים". מקרה קלאסי של הטיית החוכמה בדיעבד.

מה שגורם לאנשים לחשוב שהם לא זקוקים למדע, כי הם "היו יכולים לחזות" את זה.

(בריוק מה שהיית מצפה, לא ?)

חוכמה בדיעבד גורמת לנו שוב ושוב להעריך בחסר את ה"מפתיעות" של ממצאים מדעיים, במיוחד את התגליות שאנו מבינים – אלה שנראות לנו אמיתיות, אלה שאנו יכולים להתאים בקלות למודלים שלנו לגבי העולם. אם אתה מבין נירולוגיה או פיזיקה וקורא חדשות בתחום, אתה כנראה מזלזל במפתיעותם של ממצאים גם בתחומים אלה. החוכמה בדיעבד גורמת להמעטה בערכם של חוקרים באופן לא הוגן. גרוע מכך, היא תמנע ממך לשים לב כשאתה פוגש ראיות שאינן מתאימות למה שבאמת היית מצפה. אנו צריכים לעשות מאמץ מודע על מנת להיות מופתעים מספיק.

1. David G. Meyers, *Exploring Social Psychology* (New York: McGraw-Hill, 1994), 15-19.

2. Paul F. Lazarsfeld, "The American Solidier—An Expository Review," *Public Opinion Quarterly* 13, no. 3 (1949): 377-404.

3. Daphna Baratz, *How Justified Is the "Obvious" Reaction?* (Stanford University, 1983).

תשובות מסתוריות

תשובות מסתוריות שואל האם המדע פותר את הבעיות הללו עבורנו. מדענים מבססים את המודלים שלהם על ניסויים ניתנים לשחזור, ולא על ספקולציות או שמועות. ולמדע יש רקורד מצויין בהשוואה לאנקדוטות, דת, ו... כל דבר אחר, בגדול. האם עדיין עלינו לחשוש מפני אמונות "מזויפות", הטיית האישור, הטיית החוכמה בדיעבד וכיוצא באלה, בזמן שאנו עובדים עם קהילה של אנשים המנסים להסביר תופעות, ולא רק לספר סיפורים שנעים לשמוע?

— רוב בנסינגר,

הטיות: מבוא

30

למקד את אי-הוודאות

היה היה פעם מרצה שלימד סטודנטים לפיזיקה. יום אחד קרא להם המרצה לכיתה והראה להם לוח מתכת רחב ומרובע, שהונח בסמוך לרדיאטור חם. כל אחד מהסטודנטים הוזמן להניח את ידו על הלוח וגילה שהצד הקרוב לרדיאטור קריר למגע, ואילו הצד המרוחק – חם. והמרצה שאל: "מדוע לדעתכם זה קורה?" כמה סטודנטים שיערו שמדובר בהסעה של זרמי אוויר, ואחרים שיערו שהלוח מכיל סוג כלשהו של מתכות משונות. כולם הוגיעו את מוחם בהסברים יצירתיים לרוב, אך איש מהם לא השפיל עצמו לומר "אני לא יודע" או "זה נראה בלתי אפשרי".

התשובה הייתה שלפני שהסטודנטים נכנסו לחדר, המרצה הפך את הלוח.¹

חישבו על הסטודנט שמגמגם בבהלה: "אה, אולי בגלל הולכת חום או משהו כזה?" אני שואל: האם תשובה זו מהווה אמונה תקינה? די קל להכריז את המילים הללו – לומר אותן בקול רם וסמכותי. אך האם המילים הללו באמת שולטות בציפייה?
חישבו על מילת הקישור התמימה הזו, "בגלל", שמופיעה לפני "הולכת החום". הרהרו בכמה מהדברים האחרים שיכולנו לשים אחריה. יכולנו לומר, למשל, "בגלל פְּלוֹגִיסְטוֹן", או "בגלל קסם".

"קסם!" אתם קופצים. "זה לא הסבר מדעי!"* בהחלט, קל לזהות שהביטויים "בגלל הולכת חום" ו"בגלל קסם" שייכים לסוגות ספרותיות שונות. "הולכת חום" הוא ביטוי שספוק עשוי לומר במסע בין כוכבים, ואילו "קסם" הוא משהו שג'יילס היה אומר בבאפי ציינת הערפדים.

עם זאת, כבייסיאנים, איננו מייחסים חשיבות יתרה לסוגות ספרותיות. עבורנו, מהותו של מודל היא השליטה שהוא מפעיל על הציפייה. אם תאמרו "הולכת חום", לאיזו חוויה יגרום לכם הדבר לצפות? בנסיבות רגילות, הדבר יגרום לכם לצפות לכך שאם תניחו את ידכם על צד הלוח הקרוב לרדיאטור, הצד הזה יהיה חם למגע יותר מאשר הצד השני. אם "בגלל הולכת חום" יכול להסביר גם מצב שבו הצד הצמוד לרדיאטור קר יותר למגע, אז הוא יכול להסביר פחות או יותר כל דבר.

וכפי שכולנו כבר יודעים בשלב זה (כך אני מקווה), אם אתם מסוגלים להסביר כל תצפית באותה מידה, יש לכם אפס ידע. "בגלל הולכת חום", בשימוש כזה, הוא למעשה היפותזה מוסווית של אנטרופיה מרבית. הוא איזומורפי מבחינת הציפייה לאמירה "קסם". זה מרגיש כמו הסבר, אבל זה לא.

נניח שבמקום לנחש, היינו מודדים את החום על לוח המתכת בנקודות שונות ובזמנים שונים. כאשר אנו רואים לוח מתכת ליד רדיאטור, היינו מצפים שטמפרטורות הנקודות יקיימו שיווי משקל של משוואת הדיפוזיה בהתאם לתנאי הגבול שהוגדרו על ידי הסביבה. אולי לא תדעו את הטמפרטורה המדויקת של הנקודה הראשונה, אבל לאחר מדידת הנקודות הראשונות – אינני פיזיקאי מספיק כדי לדעת כמה נקודות נדרשות – תוכלו לתת ניחושים מצויינים לגבי כל השאר. אמן אמיתי באמנות השימוש במספרים כדי לתחום את הציפייה

* הארי התקשה לנשום, "אבל מה בדיוק קורה פה?" "קסם", אמרה פרופסור מקגונגל.

"זאת סתם מילה! אפילו אחרי שאת אומרת לי את זה, אני לא יכול לבצע תחזיות חדשות!"

לגבי תופעות חומריות – "פיזיקאי" – היה מבצע כמה מדידות ואומר, "הלוח הזה היה בשיווי משקל עם הסביבה לפני שתיים וחצי דקות, התהפך לצד השני, וכעת הוא נע בחזרה לשיווי משקל".

הטעות העמוקה יותר של הסטודנטים אינה בכך שהם נכשלו בתיחום הציפייה. שגיאתם העמוקה יותר טמונה בכך שהם חשבו על עצמם כ"עושים פיזיקה". הם אמרו את המילה "בגלל", ולאחריה את סוג המילים שספוק עשוי לומר *במסע בין כוכבים*, וחשבו שבכך הם נכנסים לתחום הידע של המדע.

ולא היא. הם פשוט העבירו את הקסם שלהם מסוגה ספרותית אחת לאחרת.

1. Search for "heat conduction." Taken from Joachim Verhagen, [http:// web.archive.org/web/20060424082937/http://www.nvon.nl/scheik/best/diversen/seijokes/seijokes.txt](http://web.archive.org/web/20060424082937/http://www.nvon.nl/scheik/best/diversen/seijokes/seijokes.txt), archived version, October 27, 2001.

31

לנחש את מילת הקוד של המורה

כשהייתי צעיר, קראתי ספרי מדע פופולרי כמו *QED*: התיאוריה המוזרה של אור וחומר מאת ריצ'רד פיינמן. ידעתי שאור הוא גל, קול הוא גל, חומר הוא גל. בהיותי בן תשע, התגאיתי מאוד בידיעותיי המדעיות.

עברו כמה שנים, וכשהתחלתי לקרוא את הרצאות פיינמן על פיזיקה, נתקלתי בפנינה שנקראה "משוואת הגלים". הצלחתי לעקוב אחר הנגזרת של המשוואה, אך במבט לאחור, לא הצלחתי לקלוט את משמעותה האמיתית בקריאה מהירה. אז חשבתי על משוואת הגלים במשך שלושה ימים, לסירוגין, עד שקלטתי שהיא מובנת מאליה באופן מביך. וכשנפל לי האסימון סוף סוף, קלטתי שבמשך כל הזמן הזה פשוט סמכתי על המילה של פיזיקאים לגבי כך שאור הוא גל, קול הוא גל, חומר הוא גל, ולא היה לי שמץ של מושג למה פיזיקאי מתכוון כשהוא אומר "גל".

ישנה נטייה אינסטינקטיבית לחשוב שאם פיזיקאי אומר "אור עשוי מגלים", והמורה שואל "ממה עשוי אור?", והתלמיד עונה "גלים!", אזי תשובתו של התלמיד נכונה. נשמע פייר, לא? אנו מקבלים את "גלים" כתשובה נכונה עבור הפיזיקאי. האם לא יהיה זה בלתי הוגן לדחות אותה עבור התלמיד? התשובה "גלים!" היא או נכונה או שגויה, אין אפשרות אחרת. נכון? וזהו עוד הרגל רע מבית הספר שממנו עלינו להיגמל. למילים אין

משמעות בפני עצמן. אם אני שומע את ההברות "בו-נה" וחושב על מכרסם גדול, זוהי עובדה לגבי מצב התודעה שלי, לא עובדה הקשורה להברות "בו-נה". רצף ההברות "עשוי מגלים" (או "בגלל הולכת חום") אינו היפותזה, הוא תבנית של תנודות המתקדמות באוויר, או של דיו על נייר. הוא יכול להיות מקושר להיפותזה בתודעה של מישהו, אבל כשלעצמו, הוא לא נכון או שגוי. אבל בכית הספר, אם תאמר "עשוי מגלים" תקבל מדבקה מהמורה, וזו ודאי התשובה הנכונה מכיוון שהמורה שמע פיזיקאי מפיק מהפה את אותן תנודות אוויר. מכיוון שהתנהגות מילולית (מדוברת או כתובה) היא זו שמזכה במדבקה, תלמידים מתחילים לחשוב שלהתנהגות מילולית יש ערך־אמת. אחרי הכל, או שאור עשוי מגלים, או שלא, נכון?

וזה מוביל להרגל גרוע עוד יותר. נניח שהמורה מציג בפניכם בעיה מבבלת הכוללת לוח מתכת ליד רדיאטור, והצד הרחוק חם יותר למגע מהצד הסמוך לרדיאטור. המורה שואל "מדוע?" אם תאמרו "אני לא יודע", אין לכם שום סיכוי לקבל מדבקה – זה אפילו לא ייחשב כהשתתפות בשיעור. אבל במהלך הסמסטר, אותו מורה השתמש בביטויים כמו "בגלל הסעת חום", "בגלל הולכת חום", ו"בגלל קרינת חום". אחד מאלה הוא כנראה מה שהמורה רוצה לשמוע. אתם אומרים, "אה, אולי בגלל הולכת חום?"

זו אינה היפותזה לגבי לוח המתכת. זו אפילו לא אמונה תקינה. זהו ניסיון לנחש את מילת הקוד של המורה.

אפילו לכתוב בראשכם את הסמלים של משוואת הדיפוזיה (המתמטיקה המתארת הולכת חום) אין פירושו יצירת היפותזה אנדוט לוח המתכת. זה לא בית ספר: אנחנו לא בודקים את הזיכרון שלכם כדי לראות אם אתם מסוגלים לכתוב את משוואת הדיפוזיה. זוהי "אומנות בייסאינית": אנו מעניקים נקודות לפי הציפיות שלכם מהחוויה. אם אתם משתמשים במשוואת הדיפוזיה, על ידי כך שאתם מודדים מספר נקודות בעזרת מדחום ואז מנסים לחזות מה המדחום יאמר במדידה הבאה, אז היא בהחלט מחוברת לחוויה. גם אם התלמיד רק מדמיין משהו זורם, וכדי לבדוק זאת הוא מחזיק גפרור ליד הצד הקריר יותר

של הלוח כדי לנסות ולמדודר לאן החום הולך, אז הדימוי המנטלי הזה של "זרימה" מתחבר לחווייה. הוא שולט בציפייה.

אם אינכם משתמשים במשוואת הדיפוזיה, כלומר מזינים מספרים ומקבלים תוצאות השולטות בציפייה שלכם לחוויות מסוימות, אזי הקשר בין המפה לשטח נותק כמו בסכין. מה שנוותר אינו אמונה, אלא התנהגות מילולית גרידא.

בעולם של בית הספר, התנהגות מילולית היא כל העניין, בין אם היא נכתבת על נייר או נאמרת בקול רם. התנהגות מילולית מזכה אתכם במדבקה או בציון נכשל. חלק מהגמילה מההרגל הרע הזה כולל פיתוח מודעות אקטיבית להבדל שבין הסבר אמיתי לבין מילות קוד.

זה נשמע לכם מחמיר מדי? כשאתם ניצבים מול לוח מתכת מבלבל, האם לא יתכן ש"הולכת חום"? יכול להיות צעד ראשון בדרך למציאת התשובה? אולי, אבל רק אם אינכם נופלים למלכודת שגורמת לכם לחפש מילת קוד. מה אם אין שום מורה שיאמר לכם שנשלחתם? במקרה כזה אתם עלולים לחשוב ש"אור הוא וַאקאליקסים" הוא הסבר טוב, ש"וַאקאליקסים" היא מילת הקוד הנכונה. זה בדיוק מה שקרה לי כשהייתי בן תשע – לא בגלל שהייתי טיפש, אלא בגלל שזה מה שקורה לבני אדם ככרירת מחדל. כך בני אדם חושבים, אלא אם כן מאמנים אותם שלא ליפול למלכודת. האנושות נשארה תקועה בכורות האלה במשך אלפי שנים.

אולי, אם נאמן את התלמידים שמילים אינן נחשבות, רק שולטי ציפיות נחשבים, התלמיד לא ייתקע על "הולכת חום"? לא? אולי הסעת חום? גם זה לא? "אולי אז, המחשבה על הביטוי "הולכת חום" תוביל לנתיב מועיל באמת, כמו:

- "הולכת חום?"
- אבל זה רק ביטוי – מה המשמעות שלו?
- משוואת הדיפוזיה?

* וַאקאליקס (Wakalix) הוא מונח מומצא שטבע פיינמן כדי להמחיש את ההבדל בין שינון שמות לבין הבנה אמיתית של המציאות.

- אבל אלה רק סמלים – איך אני מיישם אותם?
 - למה יישום משוואת הדיפוזיה גורם לי לצפות?
 - הוא בטח לא גורם לי לצפות שהצד של לוח מתכת שרחוק יותר מהרדיאטור יהיה חם יותר למגע.
 - אני שם לב שאני מבולבל. אולי הצד הקרוב רק מרגיש קר יותר, בגלל שהוא עשוי מחומר מבודד יותר ומעביר ליד שלי פחות חום? אנסה למדוד את הטמפרטורה...
 - אוקיי, זה לא העניין. האם אני יכול לנסות לוודא אם משוואת הדיפוזיה בכלל נכונה לגבי הלוח הזה? האם החום זורם כפי שהוא זורם בדרך כלל, או שמהו אחר קורה כאן?
 - אני יכול להחזיק גפרור ליד הלוח ולנסות למדוד איך החום מתפשט לאורך זמן...
- אם לא נקפיד על כך ש"אה, אולי בגלל הולכת חום?" הוא הסבר מזויף, סביר להניח שהתלמיד ייתקע על איזו מילת-קוד-מסוג-ואקאליקסים. זה קורה כברירת מחדל: זה קרה למיין האנושי כולו במשך אלפי שנים.

32

מדע כסממן קבוצתי

בקדימון לסרט "אקס-מן" נשמע קריין שאומר: "בכל יצור אנושי... קיים הקוד הגנטי... למוטציה". מסתבר שבאמצעות מוטציה ניתן לפתח כל מיני יכולות מגניבות. המוטנטית סופה, למשל, מסוגלת להטיל ברקים.

אני מפציר בפניך, קורא יקר, להרהר במנגנון הביולוגי הנדרש לייצור חשמל, את ההתאמות הביולוגיות שיידרשו כדי להימנע מפגיעת חשמל, ואת המעגלים הקוגניטיביים הנדרשים לשליטה מדויקת בברקים. לו באמת היינו חוזים באורגניזם כלשהו המפתח יכולות כאלו בתוך דור אחד, כתוצאה ממוטציה, הדבר היה מפרך לחלוטין את כל המודל הניאו-דרוויניסטי של הברירה הטבעית. יהיה זה גרוע יותר ממציאת מאובני ארנבים בתקופת הפרה-קמבריון. אם היה ביכולתה של תורת האבולוציה להימתח עד כדי תיאור המקרה של סופה, אז היא, דה פקטו, היתה מסוגלת להסביר כל דבר, וכולנו יודעים מה זה אומר.

הקומיקס של "אקס-מן" משתמש במונחים כמו "אבולוציה", "מוטציה" ו"קוד גנטי", אך ורק כדי למקם את עצמו במה שכותביו וקוראיו תופסים כסוגה הספרותית של "מדע". החלק שמפחיד אותי הוא המחשבה על כמות האנשים, במיוחד בתקשורת, שמבינים מדע דק בתור סוגה ספרותית.

אני פוגש אנשים שבהחלט מאמינים באבולוציה, המלגלים על

בורותם של בריאתנים,* בעוד להם עצמם אין שום מושג לגבי מה תורת הביולוגיה האבולוציונית מתירה ומה היא אוסרת. הם ידברו על "השלב הבא באבולוציה של האנושות", כאילו שהברירה הטבעית הגיעה עד לכאן בכך שעקבה אחרי הוראות של איזו תוכנית. או גרוע מכך, הם ידברו על משהו שנמצא לחלוטין מחוץ לתחום של ביולוגיה אבולוציונית, כמו עיצוב משופר לשבבי מחשב, או פיצול של תאגידים, או בני אדם המגבים את עצמם לתוך מחשבים, והם יקראו לזה "אבולוציה". אם ביולוגיה אבולוציונית היתה יכולה לתאר את זה, היא היתה יכולה לתאר כל דבר.

סביר להניח שרובם המכריע של האנשים שמאמינים באבולוציה משתמשים בביטוי "בגלל האבולוציה" משום שהם רוצים להרגיש חלק מהחברה המדעיים – אמונה כסממן קבוצתי, כמו ללבוש חלוק מעבדה. אם אותם חברה מדעיים היו משתמשים במקום זאת בביטוי "בגלל תכנון תבוני", אותם אנשים היו מאמצים אותו באותה התלהבות – זה לא היה משנה דבר לגבי מנגנוני הציפייה שלהם. מבחינתם, האמירה "בגלל האבולוציה" במקום "בגלל תכנון תבוני" לא אוסרת את קיומה של סופה. עבורם, מטרתה היחידה של אותה אמירה היא להזדהות עם שבט.

אני פוגש אנשים שבהחלט מוכנים לשקול את הרעיון של בינה מלאכותית טיפשה יותר מבני אדם, או אפילו בינה מלאכותית קצת יותר חכמה מבני אדם. הצג בפניהם את הרעיון של בינה מלאכותית על-אנושית חזקה, ומיד הם יתייגו זאת כ"פסאודו מדע". זה לא שהם חושבים שיש להם איזו תיאוריית אינטליגנציה שבעזרתה הם מסוגלים לחשב גבול עליון תיאורטי לכוחו של תהליך אופטימיזציה. במקום זאת, הם משייכים בינה מלאכותית על-אנושית חזקה לסוגה הספרותית של ספרות אפוקליפטית, בעוד שבינה מלאכותית המנהלת

* בריאתנים (Creationists) – אנשים המאמינים שהעולם, בעלי החיים והאדם נבראו באופן על-טבעי ע"י אל כלשהו, ולא התפתחו מתוך המפץ הגדול והאבולוציה.

תאגיד קטן משויכת לז'אנר הספרותי של מגזין "Wired"*. הם לא מדברים מתוך מודל של קוגניציה. הם לא מבינים שהם זקוקים למודל. הם לא מבינים שמדע עוסק במודלים. הביקורות הקטלניות שלהם מורכבות אך ורק מהשוואות לספרות אפוקליפטית, במקום, נניח, חוקים ידועים האוסרים תוצאה מהסוג הזה. הם מבינים מדע רק כסוגה ספרותית, או כקבוצה אליה ניתן להשתייך. התלבושת הזו לא נראית להם כמו חלוק מעבדה, זו פשוט לא הקבוצה שהם מעודדים. האם יש רעיון מדעי כלשהו שאתם גאים להאמין בו, למרות שאינכם משתמשים באמונה זו באופן מקצועי? מוטב שתשאלו את עצמכם מה לא ייתכן שיקרה, על פי אותה אמונה. זהו סך כל מה שהטמעתם והפכתם לחלק אמיתי מעצמכם. כל דבר אחר הוא כנראה מילות קוד או חלוקי מעבדה.

* מגזין טכנולוגיה וחדשנות אמריקאי.

33

סיבתיות מדומה

במאה ה-18, "פְּלוֹגִיסְטוֹן" היה התשובה העדכנית ביותר ל'יסוד האש' של האלכימאים היוונים. הציטו עץ, תנו לו לבעור. מהו אותו בוחק כתום המכונה "אש"? מדוע העץ הופך לאפר? לשתי השאלות הללו, הכימאים של המאה השמונה-עשרה ענו: "פּלוֹגִיסְטוֹן".

...וזה הכל, כן? זאת הייתה התשובה שלהם: "פּלוֹגִיסְטוֹן".

פּלוֹגִיסְטוֹן השתחרר מחומרים בוערים בתור אש נראית לעין. ככל שהשתחרר עוד פּלוֹגִיסְטוֹן, החומרים הבוערים איבדו פּלוֹגִיסְטוֹן ובכך הפכו לאפר, שהוא ה"חומר האמיתי". להבות שבערו במיכלים סגורים כבו משום שהאוויר הפך רווי בפּלוֹגִיסְטוֹן, ועל כן לא יכל להכיל עוד. פחם-עץ הותיר אחריו מעט מאוד רמזים לאחר בעירה מכיוון שהיה למעשה פּלוֹגִיסְטוֹן כמעט טהור.

כמובן שתיאוריית הפּלוֹגִיסְטוֹן לא שימשה איש כדי לחזות את התוצאה של תגובה כימית. קודם הסתכלת על התוצאה, ואז השתמשת בתיאוריית הפּלוֹגִיסְטוֹן כדי להסביר אותה בדיעבד. זה לא שתיאוריטיקנים של הפּלוֹגִיסְטוֹן חזו שלֵהבה תכבה במיכל סגור. הם הדליקו להבה במיכל, צפו בה בזמן שכבתה ואז אמרו: "האוויר בטח רווי עכשיו בפּלוֹגִיסְטוֹן". לא יכולת אפילו להשתמש בתיאוריית הפּלוֹגִיסְטוֹן כדי לומר מה לא אמור לקרות; היה בכוחה להסביר כל דבר.

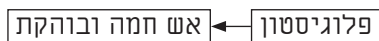
היה זה עידן מוקדם יותר של המדע. במשך זמן רב, איש לא

הבחין בכך שיש בעיה. הסברים מדומים לא מרגישים מדומים. זה מה שעושה אותם למסוכנים.

מחקר מודרני מציע שבני אדם חושבים על סיבה ותוצאה באמצעות משהו הדומה לגרפים מכוונים לא־מעגליים (DAGs) של רשתות בייסיאניות. מכיוון שירד גשם, המדרכה רטובה. מכיוון שהמדרכה רטובה, היא חלקלקה:



מכאן אנו יכולים להסיק – או, ברשת בייסיאנית, לחשב בקפדנות בעזרת הסתברויות – שכאשר המדרכה חלקלקה, כנראה ירד גשם. אבל אם אנחנו כבר יודעים שהמדרכה רטובה, הגילוי שהמדרכה חלקלקה לא מלמד אותנו שום דבר נוסף לגבי השאלה אם ירד גשם. מדוע בעירת האש היא חמה ובוהקת?



הנימוק הזה נשמע כמו הסבר. הוא מיוצג באותו פורמט נתונים קוגניטיבי. אלא שהמוח האנושי לא מזהה באופן אוטומטי מתי סיבה מסוימת שולחת חץ שאינו מגביל לכיוון התוצאה שלה. גרוע מכך, הודות להטיית החוכמה בדיעבד, אנו עשויים לחוש כאילו הסיבה כן מגבילה את התוצאה, בעוד שלמען האמת היא רק הותאמה לה בדיעבד.

מעניין לציין, שההבנה המודרנית שלנו לגבי דרכי החשיבה ההסתברותית על סיבתיות יכולה לתאר במדויק איפה טעו התיאורטיקנים של הפלוגיסטון. אחד ממקורות ההשראה העיקריים לרשתות בייסיאניות היה בעיית "ההצבעות הכפולות" של ראיות במקרים בהם ההסקה מיטלטלת בין תוצאה לסיבה. לדוגמה, נניח שאני מקבל מידע לא אמין האומר שהמדרכה רטובה. מידע זה אמור לגרום לי לחשוב שיש סבירות גבוהה יותר שיורד גשם. אבל, אם סביר יותר שיורד גשם, האם זה לא מעלה את הסבירות למדרכה רטובה? והאם זה לא מעלה את הסבירות למדרכה חלקלקה? אבל אם המדרכה

חלקלקה, היא כנראה רטובה, ואז עלי להעלות שוב את ההסתברות לכך שיורד גשם...

יהודה פָּרַל משתמש במטאפורה של אלגוריתם לספירת חיילים בטור.¹ נניח שאתה עומד בטור, ואתה רואה שני חיילים סמוכים אליך – אחד מלפנים ואחד מאחור. סך הכל שלושה חיילים, כולל אותך. אז אתה שואל את החייל מאחוריך, "כמה חיילים אתה רואה?" הוא מסתכל סביב ואומר, "שלושה." אז בסך הכל שישה חיילים. זו, כמובן, לא הדרך הנכונה לעשות זאת.

דרך חכמה יותר תהיה לשאול את החייל שמלפניך, "כמה חיילים לפניך?" ואת החייל מאחוריך, "כמה חיילים מאחוריך?" את השאלה "כמה חיילים לפניך?" אפשר להעביר הלאה כמסר ללא בלבול או "הצבעה כפולה". אם אני בראש הטור, אני מעביר את המסר "חייל אחד לפניך", עבור עצמי. האדם הצמוד אלי מאחור מקבל את המסר "חייל אחד לפניך", ומעביר הלאה את המסר "2 חיילים לפניך" לחייל שמאחוריו. באותו הזמן, כל חייל מקבל גם את המסר "N חיילים מאחור" מהחייל שמאחוריו, ומעביר אותו כ- $N + 1$ חיילים מאחור" לחייל שלפניו. כמה חיילים בסך הכל? חבר את שני המספרים שקיבלת, בתוספת אחד עבור עצמך: זהו המספר הכולל של החיילים בטור.

הרעיון המרכזי הוא שהחייל חייב לעקוב אחר שני המסרים בנפרד, המסר־קדימה והמסר־אחורה, ולחבר אותם רק בסוף התהליך. לעולם אינך מוסיף חיילים מהמסר־אחורה שקיבלת למסר־קדימה שאתה מעביר חזרה. למעשה, המספר הכולל של החיילים לעולם אינו מועבר כמסר – הוא לעולם לא נאמר בקול רם.

עיקרון דומה פועל בחשיבה הסתברותית קפדנית על סיבתיות. אם אתה לומד משהו על ירידת הגשם, ממקור כלשהו שאינו צפייה במדרכה הרטובה, הדבר ישלח מסר־קדימה מ־[גשם] ל־[מדרכה רטובה] ויעלה את הציפייה שלנו שהמדרכה תהיה רטובה. אם אתה צופה במדרכה הרטובה, הדבר שולח מסר־אחורה לאמונה שלנו שיורד גשם, והמסר הזה מתפשט מ־[גשם] לכל הצמתים השכנים למעט צומת

מדרכה רטובה. מדרכה רטובה. אנו סופרים כל פיסת ראייה בדיוק פעם אחת. שום מסר לעולם לא "מקפץ" הלוך ושוב. את האלגוריתם המדויק ניתן למצוא בספרו הקלאסי של יהודה פרל: *הסק הסתברותי במערכות תבוניות: רשתות של היקש סביר*¹.

אז מה השתבש בתיאוריית הפלוגיסטון? כשאנו מבחינים שהאש היא חמה, צומת אש יכול לשלוח מסר־ראייה־לאחור לצומת פלוגיסטון, מה שגורם לנו לעדכן את אמונותינו לגבי פלוגיסטון. אך אם זה המצב, איננו יכולים לספור זאת כחיזוי מוצלח (קדימה) של תיאוריית הפלוגיסטון. המסר אמור ללכת בכיוון אחד בלבד, מבלי לקפוץ בחזרה.

למרבה הצער, בני אדם אינם משתמשים באלגוריתם קפדני לעדכון רשתות אמונה. אנו לומדים על צְמֵת־יהורה מצפייה בילדים, ומנבאים צְמֵת־ילד מאמונות לגבי הורים. אבל אנחנו לא מנהלים בקפדנות פנקסים נפרדים עבור המסר־אחורה והמסר־קדימה. אנחנו פשוט זוכרים שפלוגיסטון הוא חם, מה שגורם לאש להיות חמה. אז נראה שתיאוריית הפלוגיסטון חוזה את חומה של האש. או גרוע מכך, אנו פשוט מרגישים שפלוגיסטון הופך את האש לחמה.

עד היום שבו תשים לב ששום חיזוי מוקדם אינו מתבצע, צומת הסיבתיות הלא־מגביל אינו מסומן כ"מזויף". הוא מיוצג באותו אופן כמו כל צומת אחר ברשת האמונות שלך. זה מרגיש כמו עובדה, כמו כל שאר העובדות שאתה מכיר: פלוגיסטון הופך את האש לחמה.

AI שתוכנן כראוי היה מבחין בבעיה מיד. זה אפילו לא ידרוש קוד ייעודי, רק ניהול נכון של פנקסי רשת האמונות. (למרבה הצער, אנו בני האדם לא מסוגלים לשכתב את הקוד של עצמנו, כפי ש-AI שתוכנן כראוי יכול היה לעשות.)

הדיבור על "הטיית החוכמה בדיעבד" הוא פשוט הדרך הלא־טכנית לומר שבני אדם אינם מפרידים בקפדנות בין מסרים קדימה ואחורה, ובכך מאפשרים למסרים־קדימה להזדהם על ידי מסרים־אחורה.

אלו שצעדו בנתיב הפלוגיסטון לפני שנים רבות כל כך לא ניסו

להיות טיפשים. אף מדען לא מנסה להיתקע ברחוב ללא מוצא בכוונה תחילה. האם בראש שלך קיימים הסברים מדומים כלשהם? אם כן, אני מבטיח לך שהם לא מסומנים בתווית "הסבר מזויף", כך שסריקת המחשבות שלך אחר מילת המפתח "מזויף" לא תניב תוצאות. הודות להטיית החוכמה בדיעבד, אין די בבדיקת יכולתה של התיאוריה שלך "לחזות" עובדות שאתה כבר מכיר. עליך לחזות עבור מחר, לא עבור אתמול. זו הדרך היחידה שבה מוח אנושי מبولגן יכול להבטיח שהוא שולח אך ורק מסר-קדימה.

Judea Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks .1 of Plausible Inference* (San Mateo, CA: Morgan Kaufmann, 1988).

34

תמרורי עצור סמנטיים

והילד שאל:

ש: מאיפה הגיעה האבן הזו?

ת: אני סיתתי אותה מהסלע הגדול, במרכז הכפר.

ש: מאיפה הגיע הסלע?

ת: הוא כנראה ניתק והתגלגל מההר העצום שמתנשא מעל הכפר שלנו.

ש: מאיפה הגיע ההר?

ת: מאותו מקום כמו כל האבנים: אלו עצמותיו של יימיר, הענק הקדמוני.

ש: מאיפה הגיע יימיר, הענק הקדמוני?

ת: מהתהום הגדולה, גינונגגאפ.

ש: מאיפה הגיעה התהום הגדולה גינונגגאפ?

ת: לעולם אל תשאל את השאלה הזו.

חשבו על הכאילרפרדוקס של הסיבה הראשונית. המדע שחזר את סדר האירועים לאחור עד למפץ הגדול, אבל למה המפץ הגדול בכלל קרה? זה טוב ויפה לומר שהמפץ הגדול היווה את נקודת האפס של הזמן – שאין שום דבר לפניו מבחינת השטף הרגיל של שעות ודקות. אבל אמירה כזו למעשה מניחה מראש את מערכת החוקים הפיזיקליים שלנו, שבעצמה נראית מורכבת להפליא. היא ממש זועקת להסבר.

מאיפה הגיעו חוקי הפיזיקה? אפשר לטעון שכולנו חיים בסימולציה מחשב, אבל אז סימולציה המחשב רצה על חוקי פיזיקה של עולם אחר – מאיפה הגיעו חוקי הפיזיקה האלה?

בשלב זה, יש אנשים שיאמרו "אלוהים!"

לעולם לא אבין מה יכול לגרום למישהו, אפילו אדם דתי במיוחד, לחשוב שזה עוזר לפתור את פרדוקס הסיבה הראשונה ולו כזית? למה שלא תשאל מיד, "מאיפה הגיע אלוהים?" האמירה "אלוהים הוא חסר סיבה" או "אלוהים ברא את עצמו" משאירה אותנו בדיוק באותו מצב כמו "הזמן התחיל עם המפץ הגדול". אנחנו פשוט שואלים מדוע כל מערכת העל קיימת מלכתחילה, או מדוע לאירועים מסוימים מותר להיות חסרי סיבה, אך לאחרים אסור.

מטיתי כאן אינה לדון בפרדוקס לכאורה של הסיבה הראשונה, אלא לשאול איך יכול מישהו לחשוב ש"אלוהים!" מסוגל לפתור את הפרדוקס. אמירת "אלוהים!" היא למעשה דרך להשתייך לשבט (ראו בפרק הצהרות וקריאות עידוד), מה שמעודד אנשים לומר זאת לעתים קרובות ככל האפשר – יש אנשים שאפילו משתמשים באמירה זו כתשובה לשאלות כמו "מדוע ההוריקן היכה בניו אורלינס?" למרות זאת, הייתי מצפה שאנשים יבחינו בכך שלגבי החידה הספציפית של הסיבה הראשונה, אמירת "אלוהים!" לא עוזרת. היא לא גורמת לפרדוקס להיראות פחות פרדוקסלי גם אם היא נכונה. איך ייתכן שמישהו לא יבחין בכך?

ג'ונתן וואלאס הציע ש"אלוהים!" משמש בתור תמרור עצור סמנטי – כלומר, הוא לא מהווה טענה עובדתית של ממש, אלא מעין תמרור קוגניטיבי: אל תחשוב מעבר לנקודה זו. אמירת "אלוהים!" לא באמת פותרת את הפרדוקס, כמו שהיא מציבה תמרור קוגניטיבי שנועד לעצור את ההמשך המתבקש של פינג-פונג השאלות והתשובות. כמוכן שאתה, אתיאסט משכיל שכמוך, לעולם לא תעשה דבר שכזה, נכון? אבל "אלוהים!" הוא לא תמרור העצור הסמנטי היחיד. הוא רק הדוגמה הראשונה והברורה ביותר.

הטכנולוגיות הטרנס-הומניסטיות – ננוטכנולוגיה מולקולרית,

ביוטכנולוגיה מתקדמת, הנדסה גנטית, בינה מלאכותית וכו' – מציבות שאלות מדיניות קשות. איזה תפקיד, אם בכלל, צריך להיות לממשלה בסוגיות כמו פיקוח על בחירת הגנים של עוברים על ידי הוריהם? האם הורים יוכלו לבחור בכוונה גנים הגורמים לסכיזופרניה? אם שדרוג האינטליגנציה של ילד הוא תהליך יקר, האם על הממשלות לסייע בהבטחת הנגישות אליו, כדי למנוע את עלייתן של אליטות קוגניטיביות? אפשר להציע את הקמתם של מוסדות שונים כדי לענות על שאלות מדיניות כאלו – למשל, ארגוני צדקה פרטיים שיספקו סיוע כספי לשדרוג אינטליגנציה – אבל השאלה המתבקשת היא, "האם מוסד כזה יהיה יעיל?" אם נסתמך על תביעות בגין אחריות מוצר כדי למנוע מתאגידים לבנות ננוטכנולוגיה מזיקה, האם זה באמת יעבוד? אני מכיר מישהו שהתשובה שלו לכל אחת מהשאלות הללו היא "דמוקרטיה ליברלית!" זהו. זאת התשובה שלו. אם תשאלו את השאלה המתבקשת "עד כמה דמוקרטיה ליברלית טיפלו היטב בבעיות מסובכות כאלה בעבר?" או "מה אם הדמוקרטיה הליברלית תעשה משהו טיפשי?" אז אתם אוטוקרטים, או ליברטופים, או סתם אנשים מאוד רעים. אסור לאף אחד לפקפק בדמוקרטיה.

פעם קראתי לסוג כזה של חשיבה "הזכות הקדושה של הדמוקרטיה". אבל מדויק יותר לומר ש"דמוקרטיה!" תפקדה עבור אותו מכר כתמרור עצור סמנטי. אם מישהו היה אומר לו "בוא ניתן לתאגיד קוקה־קולה את האחריות על זה!", הוא היה שואל את השאלות המתבקשות: "למה? מה תאגיד קוקה־קולה יעשה בנידון? למה שנסמוך עליהם? האם הם הצליחו בעבר בבעיות מסובכות כל כך?"

נניח שמישהו אומר "אמריקאים־מקסיקנים זוממים לסלק את כל החמצן מהאטמוספירה של כדור הארץ". סביר להניח שתשאלו, "למה שהם יעשו את זה? האם אמריקאים־מקסיקנים לא צריכים לנשום? האם אמריקאים־מקסיקנים בכלל מתפקדים בתור קנוניה?" אם אינכם שואלים את השאלות המתבקשות האלה כשמישהו אומר, "תאגידים זוממים לסלק את כל החמצן של כדור הארץ", אז "תאגידים!" מתפקד עבורכם כתמרור עצור סמנטי.

היזהרו שלא ליצור כאן טיעון נגד גנרי נגד דברים שאתם לא אוהבים – "אה, זה סתם תמרור עצור!" אף מילה לא מהווה תמרור עצור בפני עצמה. השאלה היא האם למילה יש השפעה כזו על אדם מסוים. קיומם של רגשות עזים כלפי משהו לא הופך אותו לתמרור עצור. קשה להאשים אותי באהבת טרוריסטים או בעוינות לקונספט של רכוש פרטי, אבל זה לא אומר ש"טרוריסטים!" או "קפיטליזם!" הם תמרורים קוגניטיביים עבורי. (למילה "אינטליגנציה" הייתה בעבר השפעה כזו עליי, אך לא עוד.) ייחודו של תמרור עצור סמנטי הוא בכך שהוא מונע מאיתנו לחשוב על השאלה המתבקשת הבאה.

35

תשובות מסתוריות לשאלות מסתוריות

דמיינו שאתם מביטים בכף ידכם, ואינכם יודעים שום דבר על תאים, שום דבר על ביוכימיה, שום דבר על דנ"א. למדתם קצת אנטומיה מניתוח גופות, אז אתם מודעים לכך שהיד מכילה שרירים, אבל אין לכם מושג מדוע השרירים נעים במקום פשוט לשכב שם כמו חימר. היד שלכם היא פשוט... חומר... ומסיבה כלשהי היא נעה לפי פקודתכם. האם אין זה מעשה קסמים?

הגוף החי אינו פועל כמנוע תרמודינמי... התודעה מלמדת כל אדם כי הגוף, במידה מסוימת, כפוף לרצונו. נראה, אפוא, כי ליצורים חיים ישנה היכולת להפעיל באופן מיידי, על חלקיקי חומר נעים מסוימים בתוך גופם, כוחות שבאמצעותם תנועותיהם של חלקיקים אלו מכוונות ליצירת תולדות מכניות נגזרות... השפעתם של חיות או צמחים על החומר מונחת רחוק לאין שיעור מהשיג ידה של כל חקירה מדעית שנעשתה עד כה. יכולתם המופלאה לכוון את תנועותיהם של חלקיקים נעים, בנס היומיומי המוכח של הרצון החופשי האנושי, ובצמיחת דור אחר דור של צמחים מתוך זרע בודד, שונים לאין שיעור* מכל תוצאה אפשרית של גיבוב מקרי של

* "לורד קלווין בהחלט התלהב מאוד מאי ידיעה." (הארי ג'יימס פוטר-אוואנס-ורס)

אטומים... ביולוגים מודרניים הגיעו פעם נוספת לכדי הסכמה על דבר מה, ודבר זה הוא עיקרון החיוניות.

לורד קלווין¹

זו הייתה תיאוריית הניטליזם, בה ההבדל המסתורי בין חומר חי לחומר דומם הוסבר על ידי אלן ויטל (élan vital) או ויס ויטליס (vis vitalis). אלן ויטל הוטמע בחומר החי וגרם לו לנוע בהתאם להכוונה של המודעות. אלן ויטל השתתף בטרנספורמציות כימיות ששום חלקיקים דוממים לא יכלו לעבור – הסינתזה המאוחרת יותר של וֶלְרֶ לְשֶׁתָּנָן, מרכיב בשתן, היוותה מכה קשה לתיאוריה הויטליסטית מכיוון שהיא הראתה שבכוחה של כימיה טהורה לשכפל תוצר ביולוגי.

לתאר את ה"אלן ויטל" כהסבר, אפילו הסבר מדומה כמו פלוגיסטון, ייחשב כנראה כמחמאה גדולה מדי עבורו. הוא תפקד בעיקר כבלם סקרננות. שאלתם "למה?" והתשובה הייתה "אלן ויטל!" כשאתם אומרים "אלן ויטל!", אתם מרגישים כאילו ידעתם למה היד שלכם זזה. יש לכם תרשים סיבתי קטן בראש שמוֹרָה:

אלן ויטל ← היד זזה

אך למען האמת אתם לא יודעים שום דבר שלא ידעתם קודם. אתם לא יודעים, למשל, אם היד שלכם תייצר חום או תקלוט חום, אלא אם כן צפיתם בכך בעבר. אם לא צפיתם בכך בעבר, לעולם לא תוכלו לחזות זאת מראש. אתם מרגישים שסקרנותכם באה על סיפוקה, אך למעשה היא לא הוזנה בדבר. ברגע שאפשר לומר "למה? אלן ויטל!" עבור כל תצפית אפשרית, אז זו תשובה שמסבירה באותה מידה כל תוצאה, השערה מוסווית של אנטרופיה מקסימלית, וכולי.

אך הלקח הגדול יותר טמון בהערצתם של הויטליסטים לאלן ויטל, בלהיטותם להכריז עליו כמסתורין שמעבר לכל מדע. כשפגשו את הדרקון 'הלא-נודע' הגדול, הויטליסטים לא שלפו את חרבותיהם אלי קרב, אלא פשטו את צווארם בהכנעה. הם התגאו בבורותם, הפכו

את הביולוגיה למסתורין מקודש, ובכך סדרו מלוותר על בורותם כשהראיות דפקו בדלת.

סוד החיים היה רחוק לאין שיעור מהישג ידו של המדע! לא רק קצת רחוק, רבותי, אלא רחוק לאין שיעור! לורד קלווין בהחלט הפיק סיפוק אדיר מאידיעת משהו.

אבל הבורות נמצאת במפה, לא בשטח. אם אני בור לגבי תופעה כלשהי בעולם, זוהי עובדה לגבי מצב התודעה שלי, לא לגבי התופעה עצמה. תופעה יכולה להיראות מסתורית עבור אדם מסוים. אין תופעות שהן מסתוריות כשלעצמן. להעריך תופעה משום שהיא נראית מסתורית באופן נפלא, פירושו להעריך את בורותך שלך.

הויטליזם חָלַק עם הפלוגיסטון את הטעות של גילום המסתורין בתור ישות. האש הייתה מסתורית, ותיאוריית הפלוגיסטון גילמה את המסתורין בדמות חומר מסתורי שנקרא "פלוגיסטון". החיים היו מסתורין מקודש, והויטליזם גילם את המסתורין המקודש בתור ישות מסתורית שנקראת "אלן ויטל". אף אחת מהתשובות לא עזרה לרכז את צפיפות ההסתברות של המודל, כלומר להפוך תוצאות מסוימות לקלות יותר להסבר ביחס לאחרות. ה"הסבר" פשוט עטף את השאלה לכדי כדור שחור קטן, קשה ואטום.

בקומדיה שכתב מולייר, רופא מסביר את כוחה של תרופת שינה באומרו שהיא מכילה "סגולה מרדמת". אותו עיקרון. זהו כשל בפסיכולוגיה האנושית שכאשר אנו ניצבים מול תופעה מסתורית, אנו נוטים להניח את קיומן של ישויות פנימיות מסתוריות במקום לשקול תהליכים מורכבים העומדים בבסיסה.

אך הכשל העמוק יותר הוא ההנחה שתשובה יכולה להיות מסתורית. אם אנו מרגישים שתופעה היא מסתורית, זוהי עובדה לגבי מצב הידע שלנו, לא עובדה לגבי התופעה עצמה. הויטליסטים זיהו פער מסתורי בידע שלהם, והניחו את קיומה של ישות מסתורית שסוגרת את הפער. בעשותם כן, הם בלבדו בין המפה לבין השטח. כל הבלבול והמבוכה קיימים בתודעתנו בלבד, לא בישויות מגולמות. זהו ההסבר האולטימטיבי והכללי מדוע שוב ושוב לאורך

ההיסטוריה האנושית, אנשים נדהמים לגלות שלשאלה מסתורית להפליא יש תשובה לא־מסתורית. מסתורין הוא תכונה של שאלות, לא של תשובות.

זו הסיבה שאני מכנה תיאוריות כגון ויטליזם תשובות מסתוריות לשאלות מסתוריות.

אלו הם הסימנים של תשובות מסתוריות לשאלות מסתוריות:

- ראשית, ההסבר מתפקד כבלם סקרנות ולא כשולט בציפיות.
- שנית, אין בהשערה חלקים נעים – המודל אינו מנגנון ספציפי מורכב, אלא חומר או כוח מוצק ואטום. ניתן לומר שהחומר המסתורי או הכוח המסתורי נמצא פה או שם, גורם לזה או לזה, אך הסיבה מדוע אותו כוח מסתורי מתנהג כך עטופה באחדות אטומה.
- שלישית, אלה המציעים את ההסבר מוקירים את בורותם. הם מדברים בגאווה על האופן שבו התופעה מביסה את המדע הרגיל או שאינה דומה לתופעות ארציות גרידא.
- רביעית, גם לאחר שניתנה התשובה, התופעה עודנה בגדר מסתורין ובעלת אותה איכות של בלתי־מוסברות מופלאה שהייתה לה מלכתחילה.

1. Silvanus Phillips Thompson, *The Life of Lord Kelvin* (American Mathematical Society, 2005).

36

חוסר התוחלת של מערכות מתהוות

כישלונותיהם של הפְּלוּגִיסְטוֹן והוֹיִטְלִיזִם מתגלים לנו בפרספקטיבה היסטורית בחוכמה שבדיעבד. האם ארהיב עוז ואנקוב בשמה של תיאוריה שרווחת כיום, וסובלת לדעתי מאותו כשל בדיוק? אני נוקב בשם התהוות (Emergence) או "תופעות מתהוות" – המוגדרות לרוב כחקר מערכות שדפוסייהן ברמה גבוהה "נובעים" או "מתהווים" מתוך אינטראקציה של אלמנטים רבים ברמה נמוכה (ויקיפדיה: "האופן שבו מערכות ודפוסים מורכבים נוצרים מתוך ריבוי של אינטראקציות פשוטות יחסית"). אם נפרש את התיאור הזה באופן מילולי, נגלה שהוא מתאים לכל תופעה ביקום שלנו שמתרחשת מעל לרמת הקווארקים הבודדים. זה חלק מהבעיה. דמיינו שמישהו מצביע על קריסה בשוק המניות ואומר "זה לא קווארק!" האם זה נשמע כמו הסבר? לא? אז גם אמירה כמו "זו תופעה מתהווה!" לא אמורה להישמע כך.

שם העצם "התהוות" הוא הדבר לו אני מתנגד, ולא הפועל "מתהווה מתוך". אין שום דבר רע באמירה "א מתהווה מ-ב", כאשר ב הוא מודל ספציפי ומפורט שיש לו חלקים נעים. גם "נובע מ-" הוא ביטוי לגיטימי שמשמעותו זהה בדיוק: כוח הכבידה נובע מעקמומיות הזמן־מרחב, לפי המודל המתמטי הספציפי של תורת היחסות הכללית. כימיה נובעת מאינטראקציות בין אטומים, לפי המודל הספציפי של אלקטרודינמיקה קוונטית.

כעת נניח שאומר שכוח הכבידה מוסבר על ידי "התהוות" או שכמייה היא "תופעה מתהווה", ואטען שזהו הסבר מספק. הביטוי "מתהווה מ-" הוא קביל, בדיוק כפי ש"נובע מ-" או "נגרם על ידי" הם קבילים, אם מיד אחריהם יתואר מודל ספציפי כלשהו אותו ניתן לשפוט בפני עצמו.

אולם, זו אינה הדרך בה משתמשים ב"התהוות" בדרך כלל. "התהוות" נהייתה נפוצה כהסבר בפני עצמו.

הפסקתי כבר לספור את כמות הפעמים בהן שמעתי אנשים אומרים: "אינטליגנציה היא תופעה מתהווה!" כאילו שזה מסביר את האינטליגנציה. שימוש כזה בביטוי עונה על כל הסעיפים עבור תשובה מסתורית לשאלה מסתורית. מה אתם יודעים, לאחר שאמרתם שאינטליגנציה היא "תופעה מתהווה"? אינכם יכולים לבצע תחזיות חדשות. אינכם יודעים דבר שלא ידעתם קודם על התנהגות של מוחות בעולם האמיתי. אתם מרגישים כאילו אתם מאמינים בעובדה חדשה, אך אינכם מצפים לשום תוצאות שונות. אתם מרגישים שסקרנותכם באה על סיפוקה, אך למעשה היא לא הוזנה בדבר. אין בהשערה חלקים נעים – אין מודל פנימי מפורט שניתן לתמוך. אלה המציעים את ההשערה של "תופעה מתהווה" מודים למעשה בכורותם לגבי המנגנונים הפנימיים, ומתגאים בכך. הם מציבים את מדע ה"התהוות" כמנוגד למדעים אחרים, שהם בעיניהם ארציים בלבד.

וגם לאחר שהתשובה "למה? התהוות!" ניתנה, התופעה עדיין מסתורית ובעלת אותה חסינות מקודשת לפתרון שהייתה לה מלכתחילה.

תרגיל כף שיכול לעזור לנו לבדוק את העניין הוא להעלים את שם התואר "מתהווה" מכל משפט שבו הוא מופיע, ולראות אם המשפט מקבל משמעות שונה:

- לפני: אינטליגנציה אנושית היא תוצר מתהווה של ירי נוירונים.
- אחרי: אינטליגנציה אנושית היא תוצר של ירי נוירונים.
- לפני: התנהגות של מושבת נמלים היא התוצאה המתהווה של האינטראקציות בין נמלים בודדות רבות.

- אחרי: התנהגות של מושבת נמלים היא התוצאה של האינטראקציות בין נמלים בודדות רבות.
 - עוד יותר טוב: מושבה עשויה מנמלים. אנו מסוגלים לחזות בהצלחה כמה היבטים של התנהגות המושבה באמצעות מודלים הכוללים נמלים בודדות בלבד, ללא משתני מושבה גלובליים, מה שמראה שאנו מבינים כיצד התנהגויות מושבה אלו נובעות מהתנהגויות הנמלים.
- תרגיל כיף נוסף הוא להחליף את המילה "מתהווה" במילה הישנה, בהסבר בו אנשים נאלצו להשתמש לפני שהומצאה ההתהוות:
- לפני: החיים הם תופעה מתהווה.
 - אחרי: החיים הם תופעה קסומה.
 - לפני: אינטליגנציה אנושית היא תוצר מתהווה של ירי נוריונים.
 - אחרי: אינטליגנציה אנושית היא תוצר קסום של ירי נוריונים.
- האין כל אחת מהאמירות האלו מעבירה בדיוק את אותה כמות ידע על התנהגות התופעה? האין כל אחת מההשערות האלו מתאימה בדיוק לאותה קבוצה של תוצאות?
- "התהוות" הפכה להיות מאוד פופולרית, בדיוק כפי שלומר "קסם" היה מאוד פופולרי בעבר. מושג ה"התהוות" קוסם לפסיכולוגיה האנושית באותו אופן, ומאותה הסיבה בדיוק. "התהוות" היא הסבר פשוט ונוח, ונעים לומר אותו. הוא מספק לנו מסתורין מקודש שאפשר להעריך. התהוות פופולרית בגלל שהיא הג'אנק-פוד של הסקרנות. אפשר להסביר הכל בעזרת התהוות, ואנשים אכן עושים זאת. כי להסביר דברים מעניק לנו תחושה נפלאה כל כך. בני אדם הם עדיין בני אדם, גם אם לקחו כמה קורסים מדעיים באוניברסיטה. ברגע שהם מוצאים דרך להשתחרר מכבליו של המדע המבוסס, הם חוזרים לאותם תעלולים כמו אבותיהם הקדמונים – לבושים בסוגה הספרותית של "מדע", אבל בני אדם הם עדיין בני אדם, והפסיכולוגיה האנושית היא עדיין פסיכולוגיה אנושית.

37

אל תגיד "סיבוכיות"

היה היה פעם...

זהו סיפור מהפגישה הראשונה שלי עם מרסלו, איתו עבדתי אחר כך במשך שנה על תאוריית הבינה המלאכותית, אך בשלב זה עדיין לא קיבלתי אותו כחניך. ידעתי שהוא התחרה באולימפיאדות למתמטיקה ומחשבים ברמה הלאומית, מה שהספיק כדי למשוך את תשומת ליבי לבחינה קרובה יותר, אך עדיין לא ידעתי אם הוא מסוגל ללמוד איך לחשוב על בינה מלאכותית.

ביקשתי ממרסלו לומר כיצד הוא חושב שבינה מלאכותית עשויה לגלות איך לפתור קובייה הונגרית. לא בעזרת תכנות מראש, זה טריוויאלי, אלא איך הבינה המלאכותית עצמה עשויה לפענח את החוקים של יקום הקובייה ולהסיק איך ניתן לנצל אותם. כיצד בינה מלאכותית תמציא לעצמה את המושג "אופרטור", או "מאקרו", שהוא המפתח לפתרון הקובייה ההונגרית.

בשלב מסוים בדיון, מרסלו אמר: "טוב, אני חושב שה-AI זקוק לסיבוכיות כדי לעשות את X, וסיבוכיות כדי לעשות את Y –"

ואני אמרתי, "אל תגיד 'סיבוכיות'."

מרסלו שאל, "למה לא?"

אמרתי, "סיבוכיות לעולם לא צריכה להיות מטרה בפני עצמה. ייתכן שתצטרך להשתמש באלגוריתם מסוים שמוסיף מידה של סיבוכיות, אבל סיבוכיות לשם סיבוכיות רק מקשה על הדברים."

(חשבתי על כל הפעמים ששמעתי אנשים טוענים שהאינטרנט "יתעורר" ויהפוך לבינה מלאכותית כאשר הוא יהיה "מורכב מספיק"). ומרסלו אמר, "אבל חייבת להיות כמות מסוימת של סיבוכיות שעושה את זה."

עצמתי את עיני לרגע, וניסיתי לחשוב איך להסביר את כל העניין במילים. כשלעצמי, פשוט הרגשתי שהאמירה "סיבוכיות" היא הצעד השגוי בריקוד הבינה המלאכותית. איש אינו מסוגל לחשוב מהר מספיק כדי לשקול בכובד ראש, במילים, כל משפט בזרם התודעה של עצמו, שכן הדבר ידרוש רקורסיה אינסופית. אנחנו חושבים במילים, אבל זרם התודעה שלנו מנווט מתחת לרובד המילולי, מונחה על ידי שרידים של תובנות עבר וניסיון מר...

אמרתי, "קראת את 'הסבר טכני של הסבר טכני'?" * "כן," אמר מרסלו. "אוקיי," אמרתי. "להגיד 'סיבוכיות' לא מעלה את הצפיפות של מסת ההסתברות שלך."

"אה," אמר מרסלו, "כמו 'התהוות'. הא. אז... עכשיו אני צריך לחשוב איך X באמת עשוי לקרות..."

זה היה הרגע בו חשבתי לעצמי, "אולי הוא מסוגל ללמוד." סיבוכיות אינה מושג חסר תועלת. מוצמדות אליה הגדרות מתמטיות, כמו סיבוכיות קולמוגורוב וסיבוכיות פֶּנְיִקֶצ'רְבוֹנְגְּקִיס. אפילו ברמה האינטואיטיבית, לעתים קרובות כדאי לחשוב על סיבוכיות – עליך לשפוט את הסיבוכיות של השערה כלשהי ולהחליט אם היא "מסובכת מדי" בהינתן הראיות התומכות בה, או להסתכל על תכנון כלשהו ולנסות לפשט אותו.

אך מושגים אינם שימושיים או חסרי תועלת כשלעצמם. רק שימושים הם נכונים או שגויים. בצעד שמרסלו ניסה לבצע בריקוד, הוא ניסה להסביר משהו בחינם, לקבל משהו תמורת לאיכלום. זוהי מעידה נפוצה ביותר, לפחות בתחום שלי. אתה יכול להצטרף לדיון על בינה מלאכותית כללית (AGI) ולראות אנשים עושים את אותו

* מאמר של יודקובסקי.

הדבר על ימין ועל שמאל, שוב ושוב – מדלגים ללא הרף על דברים שאינם מבינים, מבלי להבין שזה מה שהם עושים. זה קורה כהרף עין: הצבת צומת סיבתי שאינו שולט בציפיות מאחורי משהו מסתורי, צומת סיבתי שנותן תחושה של הסבר, אך אינו כזה. הטעות מתרחשת מתחת לרובד המילים. היא אינה דורשת פגם אופי מיוחד – כך בני אדם חושבים כברירת מחדל, וכך הם חשבו מאז ימי קדם.

הדבר ממנו יש להימנע הוא הדילוג מעל החלק המסתורי. עליך להתעכב על המסתורין כדי להתעמת איתו ישירות. ישנן מילים רבות המסוגלות לדלג מעל תעלומות, וחלקן יהיו לגיטימיות בהקשרים אחרים – "סיבוכיות", למשל. אבל הטעות המהותית היא אותו דילוג, בלי קשר לסוג הצומת הסיבתי שמוצב מאחוריו. הדילוג אינו מחשבה, אלא מיקרומחשבה. עליך להקדיש לכך תשומת לב מרבית כדי לתפוס את עצמך בשעת מעשה. וכאשר תאמן את עצמך להימנע מדילוגים, יהפוך עבורך הדבר לעניין של אינסטיקט, ולא של חשיבה מילולית. שומה עליך להרגיש אילו חלקים במפה שלך עדיין ריקים, וחשוב מכך, לשים לב לאותה הרגשה.

אני חושד שבאקדמיה קיים לחץ עצום לטאטא בעיות מתחת לשטיח כדי שאפשר יהיה להציג מאמר שנראה שלם. תקבל יותר שבחים על מודל שלם לכאורה הכולל כמה "תופעות מתהוות", לעומת מפה חסרה-במפורש, שלידה כתוב: "אין לי מושג איך החלק הזה עובד" או "ואז מתרחש נס". כתב עת עשוי אפילו לדחות את המאמר האחרון, שכן מי יודע אם הצעדים הלא-ידועים הם לא המקום שבו מתרחשים כל הדברים המעניינים? וכן, קורה לפעמים שכל החלקים הלא-קסומים במפה שלך מתגלים גם כלא-חשובים. זה המחיר שמשלמים לפעמים על כניסה לארץ הלא נודעת ועל ניסיון לפתור בעיות שלב אחרי שלב. אבל זה הופך את הידיעה שעדיין לא סיימת לחשובה אפילו יותר. רוב האנשים כלל לא מעזים להיכנס לארץ הלא נודעת, עקב הפחד המצמית מבזבוז זמנם. ואם אתה עובד בסטארטאפ מהפכני בתחום ה-AI, הלחץ

לטאטא בעיות מתחת לשטיח עצום אף יותר, שהרי אם לא תעשה כן, תיאלץ להודות בפני עצמך שאתה עדיין לא יודע איך לבנות בינה מלאכותית, ותוכניות החיים שבנית יקרסו כליל אל מול עיניך. אך אולי אני מסביר בפירוט יתר, שכן אותו דילוג קורה אצל בני אדם כברירת מחדל. אם אתה מחפש דוגמאות, פשוט הקשב לאנשים דנים בנושא דת, פילוסופיה, רוחניות או כל תחום מדעי שבו אין להם הכשרה מקצועית.

מרסלו ואני פיתחנו מוסכמה בעבודת ה-AI שלנו: כאשר נתקלנו במשהו שלא הבנו, דבר שקרה לעתים קרובות, היינו אומרים "קסם". לדוגמה, "X עושה את Y באופן קסום" – כדי להזכיר לעצמנו שיש לנו כאן בעיה לא פתורה, פער בהבנה שלנו. הרבה יותר טוב להגיד "קסם" מאשר "סיבוכיות" או "התהוות". האחרונות יוצרות אשליה של הבנה. חכם יותר להגיד "קסם", ולהשאיר לעצמך סימניה, תזכורת לעבודה שתצטרך לעשות מאוחר יותר.

38

הטיה חיובית: להביט אל תוך האפלה*

אני מעביר שיעור בכיתה, ואני כותב על הלוח שלושה מספרים: 2-4-6. "אני חושב על כלל", אני אומר, "שחל על רצפים בני שלושה מספרים. רצה הגורל, והרצף 2-4-6 מציית לכלל הזה. על השולחן לפניכם תוכלו למצוא ערימה של כרטיסיות. רשמו על הכרטיסיה הראשונה רצף של שלושה מספרים, ואני אסמן אותה ב'כן' אם הרצף מתאים לכלל, או 'לא' אם הוא לא מתאים לכלל. לאחר מכן תוכלו לרשום קבוצה נוספת של שלושה מספרים ולשאול שוב אם היא מתאימה, וכך הלאה. כשתהיו בטוחים שאתם יודעים מהו הכלל, רשמו אותו על כרטיסייה. באפשרותכם לבדוק כמה שלשות שתרצו."

הנה התיעוד של ניחושי אחד הסטודנטים:

לא 4-6-2

כן 4-6-8

כן 10-12-14

בנקודה זו הסטודנט רשם את הניחוש שלו לגבי הכלל. איך אתם חושבים שהכלל נראה? האם הייתם רוצים לבדוק שלשה נוספת, ואם כן, איזו? חשבו על זה לרגע לפני שתמשיכו.

* "אבל אלא אם קוראים את הספרים הממש ממש טובים, לא לומדים איך לעסוק במדע כמו שצריך" (הארי ג'יימס וכולי)

האתגר לעיל מבוסס על ניסוי קלאסי של פיטר גייסון, שנקרא משימת ה-2-4-6. למרות שהנבדקים שקיבלו את המשימה הזו הביעו בדרך כלל ביטחון עצמי גבוה לגבי הניחושים שלהם, רק 21% מהנבדקים ניחשו את הכלל האמיתי בהצלחה, ושחזורים שנעשו מאז המשיכו להראות שיעורי הצלחה של כ-20%¹.

שמו של המחקר הוא "על הקושי בפסילת היפותזות במשימה תפיסתית". נבדקים הניגשים למשימת ה-2-4-6 מנסים בדרך כלל לייצר דוגמאות חיוביות, ולא דוגמאות שליליות – הם מיישמים בראשם את הכלל המשוער על מנת לייצר מקרה מייצג, ובודקים אם הוא מסומן כ"כן".

לפיכך, מי שמגבש את ההשערה "מספרים הגדלים בשתיים" יבדוק את השלשה 8-10-12, ישמע שהיא מתאימה, ויכריז בביטחון על הכלל. מי שמגבש את ההשערה "X-2X-3X" יבדוק את השלשה 3-6-9, יגלה שהיא מתאימה, ואז יכריז על הכלל הזה.

בכל המקרים הכלל האמיתי הוא זהה: שלושת המספרים חייבים להיות בסדר עולה.

אבל כדי לגלות זאת, היה עליכם לייצר שלשות שלא אמורות להתאים לכלל שלכם, כגון 20-23-26, ולראות אם הן מסומנות ב"לא". זהו דבר שאנשים נוטים שלא לעשות בניסוי זה, משום מה. במקרים מסוימים, נבדקים חושבים, "בודקים" ומכריזים על כללים מסובכים הרבה יותר מהתשובה האמיתית.

תופעה קוגניטיבית זו נכרכת בדרך כלל יחד עם "הטיית האישוש". עם זאת, נראה לי שיש להבחין בין התופעה בה אנו מנסים לבדוק דוגמאות חיוביות במקום שליליות, לבין התופעה בה אנו מנסים לשמר את האמונה שאיתה התחלנו. "הטיה חיובית" משמשת כביטוי נרדף ל"הטיית האישוש", ומתאימה בהרבה לתיאור הפגם הספציפי הזה.

פעם היה נדמה שתיאוריית הפלוגיסטון יכולה להסביר מדוע להבת נר נכבית במיכל סגור (האוויר הופך רווי בפלוגיסטון ולא מותיר מקום לפלוגיסטון נוסף להשתחרר), אבל תיאוריית הפלוגיסטון

הייתה יכולה להסביר באותה מידה מצב שבו הלהבה לא כבתה. כדי להבחין בכך, עליכם לחפש דוגמאות שליליות במקום דוגמאות חיוביות, להסתכל לכיוון האפס במקום לכיוון האחד, דבר שנוגד את מה שהוכח כאינסטינקט אנושי.

שכן על פי אינסטינקט, אנו בני האדם חיים רק בחצי מהעולם. אדם יכול לשמוע שיעורים על הטיה חיובית במשך ימים שלמים, ועדיין לפספס אותה ברגע האמת. הטיה חיובית אינה משהו שאנו עושים מתוך היגיון, או אפילו מתוך היקשרות רגשית. משימת ה-4-6-2 היא משימה "קרה", לוגית, נעדרת כל נופך רגשי. ואף על פי כן הטעות מתרחשת ברובד עמוק יותר מהמילולי, רובד של דימויים, של תגובות אינסטינקטיביות. מכיוון שהבעיה לא נובעת ממעקב אחר כלל מכוון שאומר "חשוב רק על דוגמאות חיוביות", לא ניתן לפתור אותה רק בעזרת ידיעה מילולית של "עלי לחשוב גם על דוגמאות חיוביות וגם על שליליות". איזו דוגמה קופצת לכם לראש? עליכם ללמוד, ללא מילים, לשבור ימינה כשהראש מושך שמאלה. עליכם ללמוד להימשך לכיוון האין, במקום להירתע ממנו.

אני כותב כבר זמן מה על כך שכוחה של השערה לא טמון במה שהיא יכולה להסביר, אלא במה שהיא לא יכולה – אם אתם יכולים להסביר כל תוצאה במידה שווה, יש לכם אפס ידע. לכן, כדי לזהות הסבר שאינו מועיל, לא מספיק לחשוב על מה שהוא מסוגל להסביר היטב – שומה עליכם לחפש גם תוצאות שהוא לא יכול להסביר, ושם טמון כוחה האמיתי של התיאוריה.

אז אמרתי את כל זה, ואז אתגרתי את התועלת שבשימוש במושג "התהוות". מגיב אחד ציין מוליכות על ופרומגנטיות כדוגמאות להתהוות. עניתי לו שגם אימוליכות על ואיפרומגנטיות הן דוגמאות להתהוות, וזו בדיוק הבעיה. אך חלילה לי מלבקר את המגיב! למרות שקראתי רבות על "הטיית האישוש", גם אני נפלתי בפח שטמנה לי משימת ה-4-6-2 כשנתקלתי בה לראשונה. זהו רפלקס תת־הכרתי שיש לאמן מחדש. אני עדיין עובד על זה בעצמי.

חלק גדול כל כך מהמיומנות של הרציונליסט נמצא מתחת לרובד

המילולי. זה הופך את הניסיון להעביר את האמנות הזו באמצעות מילים לאתגר לא פשוט. אנשים יסכימו איתך, אבל אז, ממש במשפט הבא, יעשו משהו תת־הכרתי בכיוון ההפוך. לא שאני מתלונן! סיבה עיקרית לכך שאני כותב זאת היא כדי להתבונן במה שהמילים שלי לא העבירו.

האם אתם מחפשים דוגמאות חיוביות להטיה חיובית ברגע זה, או שאתם מקדישים חלק מהחיפוש שלכם למה שהטיה חיובית אמורה לגרום לכם לא לראות? האם אתם מביטים אל האור או אל האפלה?

1. Peter Cathcart Wason, "On the Failure to Eliminate Hypotheses in a Conceptual Task," *Quarterly Journal of Experimental Psychology* 12, no. 3 (1960): 129–140, doi:10.1080/17470216008416717.

39

אי-ודאות שיטתית

בספרו "בחירה רציונלית בעולם של חוסר ודאות", מתאר רובין דוס ניסוי שערך עמוס טְבֶרְסְקִי:^{2,1}

ניסויים פסיכולוגיים רבים נערכו בסוף שנות ה-50 ותחילת שנות ה-60, בהם התבקשו נבדקים לנבא את תוצאותיו של אירוע שהיה בעל רכיב אקראי אך עם זאת בעל שיעור בסיס כלשהו הניתן לחישוב. לדוגמה, נבדקים התבקשו לנבא האם הקלף הבא שהנסיין ישלוף מהחפיסה יהיה אדום או כחול, בהינתן ש-70% מהקלפים הם כחולים, אך רצף הקלפים האדומים והכחולים הוא אקראי לחלוטין.

במצב כזה, האסטרטגיה שתניב את שיעור הצלחה הגבוה ביותר היא לנבא את האירוע הנפוץ יותר. למשל, אם 70% מהקלפים הם כחולים, אזי ניבוי של 'כחול' בכל ניסיון מניב שיעור הצלחה של 70%.

אולם, מה שהנבדקים נטו לעשות במקום זאת, היה לבצע מעין "התאמת הסתברויות" – כלומר, לנבא את האירוע הסביר יותר בתדירות היחסית שבה הוא אמור להתרחש. למשל, נבדקים נטו לנבא ב-70% מהמקרים שיופיע קלף כחול, וב-30% מהמקרים שיופיע קלף אדום. אסטרטגיה כזו מניבה שיעור הצלחה של 58%, מכיוון שהנבדקים צודקים ב-70% מהזמן לגבי קלף כחול (שקורה בהסתברות של 0.70) וב-30% מהזמן לגבי קלף אדום (שקורה בהסתברות של 0.30):

$$. (0.70 \times 0.70) + (0.30 \times 0.30) = 0.58$$

למען האמת, נבדקים נוטים לנבא את האירוע התדיר יותר בהסתברות מעט גבוהה יותר מזו שבה הוא מתרחש בפועל, אך לעולם אינם מתקרבים לנבא את התרחשותו ב-100% מהמקרים, אפילו כשהם מקבלים תשלום עבור רמת הדיוק של הניבויים שלהם... למשל, נבדקים שקיבלו 5 סנט עבור כל ניבוי נכון לאורך אלף ניסיונות... ניבאו [את האירוע הנפוץ יותר] ב-76% מהמקרים.

אל לכם לחשוב שהניסוי הזה עוסק בפגם קטן באסטרטגיות של הימורים. הוא ממחיש בתמציתיות את הרעיון החשוב ביותר בתורת הרציונליות כולה.

הנבדקים פשוט ממשיכים לנחש מדי פעם 'אדום', כאילו הם חושבים שבאפשרותם לנחש בדרך כלשהי את הרצף האקראי. על ניסוי זה דוס ממשיך ואומר: "חרף המשוב החוזר לאורך אלף ניסיונות, נבדקים אינם מסוגלים להביא את עצמם להאמין שהם במצב בו הם לא מסוגלים לנבא."

אבל הטעות בוודאי עמוקה יותר מכך. גם אם נבדקים חושבים שיש להם השערה כלשהי, הם לא חייבים להמר בפועל על בסיס אותה השערה על מנת לבדוק אותה. הם יכולים פשוט לומר, "אם ההשערה הזו נכונה, הקלף הבא יהיה אדום" – ובכל זאת להמר על כחול. הם יכולים לבחור כל פעם בכחול, ולצבור סנטים רבים ככל הניתן, תוך שהם רושמים לעצמם בראש את הניחושים הפרטיים שלהם לגבי כל דפוס שנדמה להם שזיהו. אם הניבויים שלהם יתבררו כנכונים, אז הם יוכלו לעבור לרצף החדש שגילו.

לא הייתי מאשים נבדק על כך שהוא ממשיך להמציא השערות – איך הוא יכול לדעת שהרצף באמת נמצא מעבר ליכולת הניבוי שלו? אך כן הייתי מאשים נבדק על כך שהוא מהמר על הניחושים, כשזה לא הכרחי לצורך איסוף מידע, ולאחר שמאות ניחושים קודמים הופרכו, פשוטו כמשמעו.

האם ייתכן שבטחוננו העצמי של אדם יהיה כל כך מופרז?
הייתי חושד שמהשווה פשוט יותר מתרחש כאן – שאסטרטגיית

”תמיד כחול” פשוט לא עלתה בדעתם של הנבדקים.
 אנשים רואים תערוכת של הרבה קלפים כחולים עם מעט קלפים
 אדומים, ומניחים שאסטרטגיית ההימורים האופטימלית היא כנראה
 תערוכת של הרבה קלפים כחולים עם מעט קלפים אדומים.
 זהו רעיון לא־אינטואיטיבי, שבהינתן מידע חלקי, אסטרטגיית
 ההימורים האופטימלית אינה דומה לרצף אופייני של קלפים.
 זהו רעיון לא־אינטואיטיבי, שהאסטרטגיה האופטימלית היא
 להתנהג בצורה שיטתית, אפילו בסביבה המורכבת גם מאלמנטים
 אקראיים.

נראה כאילו ההתנהגות שלך אמורה להיות בלתי צפויה, בדיוק
 כמו הסביבה – אבל לא! מפתח אקראי אינו פותח מנעול אקראי רק
 מפני ששניהם ”אקראיים”.

לא נלחמים באש באמצעות אש. נלחמים באש באמצעות מים.
 אבל מחשבה כזו כרוכה בצעד נוסף, מושג חדש שאינו מופעל ישירות
 על ידי הצגת הבעיה, ולכן זה לא הרעיון הראשון שעולה לראש.
 בדילמה של הקלפים הכחולים והאדומים, הידע החלקי שלנו
 אומר לנו – בכל סיבוב וסיבוב – שההימור הטוב ביותר הוא כחול.
 העצה הזו של הידע החלקי שלנו היא זהה בכל סיבוב וסיבוב. אם
 ב-30% מהזמן אנו הולכים נגד הידע החלקי שלנו ומהמרים במקום
 זאת על אדום, אנו נשיג עקב כך תוצאה גרועה יותר – מכיוון שכעת
 אנו נוהגים בטיפשות גמורה, ומהמרים על מה שאנו יודעים שהוא
 התוצאה הפחות סבירה.

אם הייתם מהמרים על אדום בכל סיבוב, הייתם משיגים את
 התוצאה הגרועה ביותר האפשרית. הייתם 100% טיפשים. אם אתם
 מהמרים על אדום ב-30% מהזמן, מול 30% קלפים אדומים, אז אתם
 הופכים את עצמכם ל-30% טיפשים.

כשהידע שלכם חלקי – כלומר, כשנראה שהעולם מכיל רכיב
 של אקראיות – הפיכת הפעולות שלכם לאקראיות אינה פותרת
 את הבעיה. הפיכת הפעולות לאקראיות מרחיקה אתכם מהמטרה,
 לא מקרבת אתכם אליה. בעולם שהוא מעורפל מלכתחילה, זריקת

האינטליגנציה שלכם לפח רק מחמירה את המצב. זהו רעיון לא־אינטואיטיבי, שהאסטרטגיה האופטימלית יכולה להיות לחשוב בצורה שיטתית, אפילו בתנאים של אי־ודאות. ועל כן הרציונליות אינה חזון נפרץ, שכן רוב האנשים, כשיפגשו את אקראיותו של העולם, ינסו להילחם בכאוס בעזרת כאוס. עליכם לעשות צעד נוסף, ולחשוב על משהו שלא קופץ לראשכם כמחשבה ראשונה, כדי לדמיין את עצמכם נלחמים באש בעזרת משהו שאינו אש. ודאי שמעתם את הלא־מוארים אומרים, "רציונליות עובדת נהדר עבור התמודדות עם אנשים רציונליים, אבל העולם אינו רציונלי". אך כשאתם ניצבים מול יריב לא רציונלי, השלכת ההיגיון שלכם מהחלון לא תועיל כהוא זה. ישנן צורות חשיבה חוקיות שעדיין מייצרות את התגובה הטובה ביותר, גם כאשר ניצבים מול יריב שמפר את אותם חוקים. תורת ההחלטות לא עולה בלהבות ומתאינת ברגע שהיא ניצבת מול יריב שאינו מציינת לתורת ההחלטות. זה לא יותר מובן מאליו מהרעיון של להמר "תמיד כחול", כשניצבים מול רצף של קלפים כחולים ואדומים גם יחד. אבל כל הימור שאתם מבצעים על אדום הוא הפסד צפוי, וכך גם כל סטייה מן "הדרך" בחשיבה שלכם עצמכם. לאור כל זאת, כמה פרקים של מסע בין כוכבים מופרכים כרגע? כמה תאוריות של AI?

1. Dawes, Rational Choice in An Uncertain World; Yaacov Schul and Ruth Mayo, "Searching for Certainty in an Uncertain World: The Difficulty of Giving Up the Experiential for the Rational Mode of Thinking," *Journal of Behavioral Decision Making* 16, no. 2 (2003): 93-106, doi:10.1002/bdm.434.

2. Amos Tversky and Ward Edwards, "Information versus Reward in Binary Choices," *Journal of Experimental Psychology* 71, no. 5 (1966): 680-683, doi:10.1037/h0023123.

40

נעוריי הפרועים והפוחזים

אומרים שהורים עושים את כל הדברים שהם אומרים לילדיהם לא לעשות, שהרי כך הם יודעים שלא כדאי לעשותם.

לפני זמן רב, בעבר הרחוק באופן בלתי נתפס, הייתי רציונליסט מסורתי¹ נאמן. החשבתי את עצמי למיומן בתורה זו, אך חסרת ידיעה אמיתית לגבי דרכו של בייס. כאשר עמד אליעזר הצעיר מול שאלה שנראתה מסתורית, עקרונות הרציונליות המסורתית לא מנעו ממנו להמציא תשובה מסתורית. זו הטעות המביכה ביותר שעשיתי בחיי, בפער עצום, ואני עדיין מתכווץ כשאני חושב עליה.

מה הייתה אותה תשובה מסתורית לשאלה מסתורית? זאת לא אתאר, שכן מדובר בסיפור ארוך ומסובך. הייתי רציונליסט מסורתי צעיר שלא הכיר את תורתם של טֶבֶרְסְקִי וְכֶנְמַן לעומקה. הכרתי את תערו של אוקאם, אך לא את כשל החיתוך. חשבתי שאני יכול להתמודד עם מחשבות מסובכות בכוחות עצמי, בסגנון המחשבות המסובכות שקראתי בספרי מדע, מבלי להבין שמורכבות הנושאת ערך של אמת אפשרית רק כאשר כל צעד במהלך הוא מהודק ומעוגן באופן מוחלט. כיום, אחת העצות העיקריות שאני נותן לרציונליסטים צעירים היא "הימנעו מליצור תכניות מסובכות או שרשראות ארוכות של הסקת מסקנות."²

* "כל מזימה שדורשת שיקרו יותר משלושה דברים שונים, לעולם לא תעבוד בחיים האמיתיים" (לוציוס מאלפוי)

אומר רק זאת: גם לאחר שהמצאתי את ה"תשובה" שלי, התופעה נותרה בעיניי בחזקת תעלומה, והחזיקה באותה איכות של בלתי חדירות מופלאה שהייתה לה מלכתחילה.

שלא תטעו, אליעזר הצעיר לא היה טיפש. כל השגיאות שאליעזר הצעיר אשם בהן עדיין נעשות כיום על ידי מדענים מכובדים בכתבי עת רציניים. נדרשה מיומנות עדינה יותר בכדי להגן עליו, יותר מכל מה שלימדו אותו אי פעם להיות כרציונליסט מסורתי.

ואכן, אליעזר הצעיר תעה בדרכו תוך שהוא עוקב בדקדקנות אחר ההוראות של הרציונליות המסורתית.

בתור רציונליסט מסורתי, אליעזר הצעיר הקפיד לוודא שתשובתו המסתורית תייצר ניבוי נועז לגבי חוויה עתידית. ספציפית, ציפיתי שנוירולוגים בעתיד יגלו שנוירונים מנצלים כבידה קוונטית, משהו בסגנון סר רוג'ר פֶּנְרוֹז. תיאור כזה דרש מנוירונים לשמור על רמה מסוימת של קוהרנטיות קוונטית, שזה משהו שאפשר לחפש, ובהתאם למצוא או לא למצוא. או שהתצפית תאשר את התחזית או שלא, נכון? אך ההשערה שיצרתי לא סיפקה שום ניבויים רטרואפקטיביים. על פי המדע המסורתי, ניבויים רטרואפקטיביים לא נחשבים – אז למה שאטרח לייצר כאלה? לביסיאני, לעומת זאת, אם להשערה אין סבירות יחסית חיובית לעומת "אני לא יודע" כבר היום, אז עולה השאלה מדוע היום אתה מאמין במשהו מורכב יותר מ"אני לא יודע". אבל ידוע לא ידעתי את דרכו של בייס, ולכן לא חשבתי על סבירות יחסית או על מיקוד צפיפות הסתברות. אני הרי "יצרתי ניבוי הניתן להפרכה". אלה החוקים, לא?

כרציונליסט מסורתי, אליעזר הצעיר הקפיד שלא להאמין בקסמים, מיסטיקה, שוביניזם פחמני, ועוזריהם משאר העמים. הכרזתי בגאווה על תשובתי המסתורית: "זה בסך הכל פיזיקה, בדיוק כמו שאר הפיזיקה!" כאילו שאוכל למנוע מקסם מלהיות מקבילה קוגניטיבית של קסם, על ידי כך שאקרא לו "כבידה קוונטית". אבל ידוע לא ידעתי את דרכו של בייס, והרובד בו הרעיון שלי היה איזומורפי לקסם נסתר מעיניי לחלוטין. נשבעתי אמונים לפיזיקה, אך שבועה זו לא הצילה

אותי. מה יודעת תורת ההסתברות על נאמנות? הדרתי את רגליי מכל המקומות אליהם אסרה עלי הרציונליות המסורתית ללכת, אך מה שנשאר עדיין היה קסם.

אין עוררין על כך שנאמנותי לרציונליות המסורתית עזרה לי לצאת מהבור אותו חפרתי לעצמי. אם לא הייתי רציונליסט מסורתי, הייתי אבוד לחלוטין. אבל רציונליות מסורתית עדיין לא הספיקה כדי להבין באמת את כל העסק. היא בסך הכל הובילה אותי לטעויות שונות מאלה שהיא אסרה במפורש.

כשאני חושב על האופן בו פעלתי בזהירות רבה כל כך לפי כללי הרציונליות המסורתית בצעירותי כל הדרך עד להשגת התשובה הלא נכונה, אני מבין טוב יותר מדוע אנשים המכנים עצמם "רציונליסטים" אינם שולטים בעולם. אתה צריך לצבור חתיכת קילומטראז' של רציונליות לפני שהיא תעשה משהו חוץ מלהוביל אותך לטעויות חדות ומעניינות.

רציונליות מסורתית נלמדת כאומנות, לא כמדע. אתה קורא ביוגרפיות של פיזיקאים מפורסמים המתארים את השיעורים שהחיים לימדו אותם, ואתה מנסה לעשות מה שהם אומרים לך לעשות. אבל לא חיית את חייהם, וחצי ממה שהם מנסים לתאר הוא למעשה אינסטינקט שאומן והוטמע בהם בעמל רב.

לפי האופן שבו בנויה הרציונליות המסורתית, לא מין הנמנע שהייתי יכול לבזבז שלושים שנה על הרעיון הטיפשי שהיה לי, כל עוד הייתי מצליח להפריך אותו בסופו של דבר, והייתי כן מול עצמי לגבי מה שהתיאוריה שלי ניבאה, ומקבל את ההפרכה כשהייתה מגיעה, וכו'. זה מספיק כדי לאפשר למכונת המדע להתקדם צעד נוסף קדימה, אבל זה קצת נוקשה עבור האנשים שמבזבזים שלושים שנה מחייהם. רציונליות מסורתית היא הליכה, לא ריקוד. היא נועדה להביא אותך לאמת בסופו של דבר, ונותנת לך יותר מדי זמן להנות מהנוף שבדרך.

רציונליסטים מסורתיים יכולים להסכים שלא להסכים. הרציונליות המסורתית לא מחזיקה באידיאל לפיו חשיבה היא אומנות מדויקת

שבה יש רק הערכת הסתברות נכונה אחת בהינתן סט הראיות. ברציונליות מסורתית מותר לך לנחש, ואז לבדוק את הניחוש שלך. אבל הניסיון לימד אותי שאם אתה מנחש כשאתה לא יודע, סופך לטעות.

דרכו של בייס גם היא אומנות שאינה מדויקת, לפחות באופן בו אני מדבר עליה. מאמרים אלו הם בסך הכל ניסיונות מגושמים לנסח במילים כמה שיעורים שטוב יותר ללמוד מתוך ניסיון. אך לפחות יש מתמטיקה בבסיס הדברים, כמו גם ראיות אמפיריות מתחום הפסיכולוגיה הקוגניטיבית על האופן בו בני אדם באמת חושבים. אולי זה יספיק כדי לעמוד ברף הגבוהה-ברמה-סטרטוספירית הנדרש עבור דיסציפלינה שבאמת מאפשרת לך להבין את כל העסק, במקום רק להוביל אותך לטעויות חדשות ומעניינות.

41

חוסר היכולת ללמוד מההסטוריה

לפני שנים רבות, בנעוריי הפרועים והפוחזים, בטרם ידעתי את דרכו של בייס, סיפקתי תשובה מסתורית לשאלה מסתורית לכאורה. מבין כשלונות רבים שצברתי בזה אחר זה, טעות אחת מתבלטת כקריטית ביותר: אליעזר הצעיר לא הבין שפתרון תעלומה אמור להפוך אותה למבלבלת פחות. ניסיתי להסביר תופעה מסתורית – תהליך שעבורי, פירושו היה לספק לה סיבה ולשבץ אותה בתוך מודל משולב של המציאות. מדוע שזה יהפוך את התופעה למסתורית פחות, כשזהו טבעה? ניסיתי להסביר את התופעה המסתורית, לא להפוך אותה (על ידי איזו אלכימיה בלתי אפשרית) לתופעה ארצית, תופעה שאפילו לא הייתה דורשת הסבר חריג מלכתחילה.

כרציונליסט מסורתי, הכרתי את המעשיות ההיסטוריות על אסטרולוגים ואסטרונומיה, על אלכימאים וכימיה, על ויטליסטים וביולוגיה. אבל התופעה המסתורית שלי לא הייתה כזו. היא הייתה משהו חדש, משהו מוזר יותר, משהו קשה יותר, משהו שהמדע הרגיל לא הצליח להסביר במשך מאות שנים –

– כאילו שכוכבים וחומר וחיים לא היו תעלומות במשך מאות ואלפי שנים, משחר המחשבה האנושית ועד לרגע בו המדע פתר אותן לבסוף –

אנו לומדים בבית הספר על אסטרונומיה וכימיה וביולוגיה, ונדמה לנו שהנושאים הללו היו מאז ומתמיד נחלתו הברורה של המדע,

שהם מעולם לא היו מסתוריים.* בכל פעם שהמדע מעז לאתגר חידה משמעותית חדשה, ילדי אותו דור מגיבים לכך בספקנות, שכן הם מעולם לא ראו את המדע מסביר משהו שמדגיש מסתורי. המדע מועיל רק להסברת נושאים מדעיים, כמו כוכבים וחומר וחיים.

אניחשבתישהלקח מאותן מעשיות היסטוריות הוא שלאסטרוולוגים ולארכימאים ולויטליסטים היה איזה פגם אופי פנימי, מעין נטייה למסתורין, שהוביל אותם להמציא הסברים מסתוריים לנושאים לא-מסתוריים. אבל ברור שאם תופעה היא באמת מאוד מוזרה, הסבר מוזר עשוי להיות במקום?

רק מאוחר יותר, כשהתחלתי להבין את המבנה הארצי והנורמלי שבתוך המסתורין, הבנתי איזה מסלול שחזרו צעדיי. רק אז הבנתי עד כמה סביר נראה הויטליזם בזמנו, עד כמה מפתיעה ומביכה הייתה תשובתו של היקום: "החיים הם ארציים, ואינם זקוקים להסברים מוזרים."

אנו קוראים היסטוריה אבל לא חיים אותה, אנו לא חווים אותה. לו רק הייתי מאמין באופן אישי בקיומן של תעלומות אסטרוולוגיות ואז מגלה את המכניקה הניוטונית, מאמין בתעלומות אלכימיות ואז מגלה את הכימיה, מאמין בתעלומות ויטליסטיות ואז מגלה את הביולוגיה. הייתי חושב על התשובה המסתורית שלי ואומר לעצמי: שום סיכוי שאני נופל בזה שוב.

* "הכוכבים היו פעם תעלומות. כל תעלומה פתורה הייתה חידה משחר המין האנושי ועד לרגע שבו מישהו פתר אותה." (הארי ג'יימס וזה)

42

להנגיש את ההיסטוריה

ישנו הרגל חשיבתי לו אני קורא הכשל הלוגי של הכללה מראיות בדיוניות. לדוגמא, עיתונאים שמזכירים את הסרט שליחות קטלנית בכתבות על AI, בדרך כלל לא מתייחסים אליו כאל נבואה או איזו אמת אבסולוטית, אבל הכתב נזכר בסרט – הוא זמין לו – כאילו היה זה מקרה היסטורי מייצג. כאילו הכתב ראה את התרחיש הזה מתממש בכוכב אחר, כך שהוא עשוי להתממש כאן באותה מידה. הרחבתי על נושא זה בחלק 7 של "הטיות קוגניטיביות המשפיעות פוטנציאלית על שיפוט של סיכונים גלובליים"¹.

ישנה טעות הופכית להכללה מראיות בדיוניות: חוסר היכולת להיות מושפע במידה מספקת מראיות היסטוריות. הבעיה עם הכללה מראיות בדיוניות היא שמדובר בבדיה – הן לא התרחשו מעולם. היא לא נובעת מאותה התפלגות כמו היקום האמיתי בו אנו חיים. בדיה שונה מהמציאות באופן יסודי. אבל ההיסטוריה כן קרתה, והיא צריכה להיות זמינה.

בסביבה הקדמונית שלנו, לא היו שום סרטים. מה שראיתו במו עיניך היה אמת. האם עלינו להתפלא מכך שהבדיות שאנו רואים בתמונות נעות אמיתיות משפיעות עלינו בצורה חזקה כל כך? לעומת זאת, את הדברים שבאמת קרו, אנו פוגשים כדיו על נייר. הם אמנם קרו, אבל מעולם לא ראינו אותם קורים. אנחנו לא זוכרים שהם קרו לנו.

מהותה של הטעות ההופכית היא להתייחס להיסטוריה כאל סיפור גרידא. לעבד אותה בעזרת אותו חלק במוח שמטפל בספרים שאנו קוראים. שפתיך אולי יאמרו שזו "אמת", ולא "בדיה", אך אין פירושו של דבר שאתה מושפע במידה בה אתה אמור להיות. הטיות רבות כרוכות בחוסר היכולת שלנו להיות מושפעים במידה מספקת ממידע יבש ומופשט.

לפני שנים רבות, סיפקתי תשובה מסתורית לשאלה מסתורית, מבלי להבין שעשיתי בדיוק את אותה הטעות שעשו אסטרונומים שהמציאו הסברים מיסטיים לכוכבים, או אלכימאים שהמציאו תכונות קסומות של חומר, או ויטליסטים ששיערו את קיומו של אֶלֶן ויטל בלתי-חדיר כדי להסביר את כל הביולוגיה.

כשהבנתי לבסוף איזה מסלול שחזרו צעדיי, חווייתי הלם פתאומי של חיבור בלתי צפוי לעבר. הבנתי שהמצאתו של הויטליזם, כמו גם קריסתו – עליהן קראתי רק בספרים – באמת קרו לאנשים אמיתיים, שחוו זאת באותו האופן בו חווייתי אני את המצאתה וקריסתה של התשובה המסתורית שלי. באותו רגע גם הבנתי שאילו הייתי חווה את העבר בפועל – אילו הייתי חי בעצמי בזמן מהפכות מדעיות מהעבר, במקום לקרוא עליהן בספרי היסטוריה – סביר להניח שלא הייתי חוזר על אותה הטעות. לא הייתי ממציא תשובה מסתורית נוספת. אלף הנפילות הראשונות היו מעבירות את המסר.

אם כך (חשבתי לעצמי), על מנת להרגיש את כוחה של ההיסטוריה במידה מספקת, עלי לנסות לדמיין את המחשבות של אליעזר שכן חי לאורך כל ההיסטוריה – עלי לנסות ולחשוב כאילו כל מה שאי פעם קראתי בספרי ההיסטוריה קרה לי עצמי (כמובן כולל שקלול-מחדש בהתאם להטיית הזמינות של ספרי היסטוריה – עלי לזכור שהייתי אלף איכרים על כל פעם שהייתי שליט). עלי לשקע את עצמי בהיסטוריה, לדמיין שאני ממש חי במשך עידנים שראיתי רק כדיו על נייר.

למה שאזכור את הטיסה הראשונה של האחים רייט? לא הייתי שם. אך כרצינוליסט, איך אוכל שלא לזכור, כשהאירוע אכן קרה?

האם יש הבדל גדול כל כך בין לצפות באירוע בעיניך שלך – שזו למעשה שרשרת סיבתית המערבת החזר של פוטונים, ולא חיבור ישיר – לבין לצפות באירוע דרך ספר היסטוריה? פוטונים וספרי היסטוריה נובעים שניהם מהאירוע עצמו בשרשראות של סיבתיות. היה עלי להתגבר על השכחה הכפויה של מי שנולד בזמן מסוים. היה עלי לזמן ולהנגיש את כל הזיכרונות, לא רק את אלו אשר עקב צירוף מקרים טהור היו שייכים לי עצמי ולעידן שלי.

כדור הארץ נעשה עתיק יותר, לפתע.

בזיכרוני הקודם, ארצות הברית התקיימה מאז ומתמיד – מעולם לא היה זמן שבו לא הייתה ארצות הברית. לא זכרתי, עד אותו רגע, כיצד קמה האימפריה הרומית, כיצד הביאה שלום ושגשוג, כיצד התקיימה לאורך מאות שנים, עד אשר שכחתי כי לא תמיד היו כך פני הדברים. אך למרות הכל האימפריה נפלה, וברברים שטפו את עירי, והידע שהיה ברשותי אבד. העולם המודרני הפך לשביר יותר בעיניי. הוא לא היה העולם המודרני הראשון.

טעויות רבות כל כך, שנעשו שוב ושוב, משום שלא זכרתי שעשיתי אותן, בכל עידן שבו מעולם לא חייתי...

ולחשוב שאנשים תוהים לפעמים אם התגברות על הטיות היא מיומנות חשובה.

האם אינכם זוכרים כמה פעמים נהרגתם באשמת ההטיות שלכם? לא? שמתי לב שאובדן זכרון פתאומי הוא תופעת לוואי נפוצה של טעות קטלנית. אבל תאמינו לי, זה קרה. אני זוכר, לא הייתי שם.

אז בפעם הבאה שאתם מטילים ספק במוזרותו של העתיד, זכרו איך נולדתם בשבט של ציידים-לקטים לפני עשרת אלפים שנה, כשאף אחד לא ידע מה זה מדע בכלל. זכרו איך הזדעזעתם עד עמקי נשמתכם כשהמדע הסביר את התעלומות הגדולות, הקדושות והנוראות שפעם הערצתם כל כך. זכרו איך פעם האמנתם שתוכלו לעוף אם תאכלו את הפטריות הנכונות, ואיך קיבלתם באכזבה את העובדה שלעולם לא תעופו, ואז עפתם. זכרו איך תמיד חשבתם שעבדות היא דבר נכון וראוי, ואז שיניתם את דעתכם. אל תדמינו איך יכולתם לחזות את

השינוי, כי זהו אובדן זכרון. זכרו שלמעשה, לא ניחשתם. זכרו איך שוב ושוב, במשך מאות שנים, העולם השתנה בדרכים שלא ניחשתם. אולי אז תהיו מזועזעים פחות ממה שיקרה בהמשך.

1. Eliezer Yudkowsky, "Cognitive Biases Potentially Affecting Judgment of Global Risks," in *Global Catastrophic Risks*, ed. Nick Bostrom and Milan M. Ćirković (New York: Oxford University Press, 2008), 91-119.

43

להסביר/לסגוד/להתעלם?

בעוד השבט שלנו נודד בערבות העשב בחיפוש אחר טרף ועצי פרי, קורה מדי פעם שמים נשפכים מטה מהשמיים. "למה לפעמים נופלים מים מהשמיים?" אני שואל את האיש החכם והמזוקן של השבט.

הוא חושב לרגע, שכן השאלה הזו לא עלתה בדעתו מעולם, ואז אומר: "מפעם לפעם, רוחות השמיים נלחמות, וכשהן נלחמות, דמן מטפף מהשמיים."

"מאיפה מגיעות רוחות השמיים?" אני שואל.

קולו יורד לכדי לחישה. "מהזמן הקדום. מהעבר הרחוק־רחוק." כשיורד גשם, ואתה לא יודע למה זה קורה, יש לך כמה אפשרויות. ראשית, אתה יכול פשוט לא לתהות למה – לא לשאול שאלות המשך, או פשוט לא לחשוב על השאלה מלכתחילה. זוהי פקודת התעלם, שבה בחר במקור האיש החכם והמזוקן. שנית, אתה יכול לנסות להמציא סוג כלשהו של הסבר, פקודת הסבר, כפי שעשה האיש המזוקן בתגובה לשאלתך הראשונה. שלישית, אתה יכול ליהנות מתחושת המסתורין – פקודת סגוד.

כפי שודאי שמתם לב מהסיפור, בכל פעם שאתם בוחרים הסבר, התרחיש הטוב ביותר הוא שאתם מקבלים הסבר כלשהו, כמו "רוחות השמיים". אבל אז ההסבר עצמו כפוף לאותה דילמה – הסבר, סגוד, או התעלם? בכל פעם שאתם לוחצים על הסבר, גלגלי המדע

'טוחנים' למשך זמן מה, מחזירים הסבר, ואז הודעה נוספת קופצת. כרציונליסטים טובים, אנו חשים חובה להמשיך ללחוץ על הסבר, אך נראה שזהו נתיב שאין לו סוף.

אתה לוחץ על הסבר עבור חיים, ומקבל כימיה. אתה לוחץ על הסבר עבור כימיה, ומקבל אטומים. אתה לוחץ על הסבר עבור אטומים, ומקבל אלקטרונים וגרעינים. אתה לוחץ על הסבר עבור גרעינים, ומקבל כרומודינמיקה קוונטית וקווארקים. אתה לוחץ על הסבר לגבי איך הקווארקים הגיעו לשם, ומקבל בחזרה את המפץ הגדול...

אנו יכולים ללחוץ על הסבר עבור המפץ הגדול, ולהמתין בעוד המדע מריץ את התהליך שלו, ואולי יום אחד הוא יחזיר הסבר מושלם. אבל אז זה רק יקפיץ הודעה נוספת. אז, אם נתמיד בכך מספיק זמן, ודאי נגיע להודעה קופצת מיוחדת, אפשרות חדשה, הסבר שאינו זקוק להסבר, מקום שבו השרשרת מסתיימת – וזה יהיה, אולי, ההסבר היחיד שראוי לדעת. הנה – הרגע לחצתי על סגוד.

זכרו תמיד שיש דרכים רבות יותר לסגוד למשהו חוץ מאשר להדליק נרות סביב מזבח.

אילו הייתי אומר, "הא, זה באמת נראה פרדוקסלי. אני תוהה איך הפרדוקסלכאורה ייפתר לבסוף?" זו היתה לחיצה על הסבר, שאכן לוקח לו לעיתים זמן מה לספק תשובה.

ואם כל העניין נראה לכם לא חשוב, או לא רלוונטי, או אם אתם מעדיפים לדחות את המחשבה על כך למחר, הרי שלחצתם על התעלם.

בחרו את התשובה שלכם בתבונה.

44

"מדע" כעוצר סקרנות

דמיינו שאני, אל מול מצלמות טלוויזיה בשידור חי, מושיט את ידי ומכריז אברקדברה וגורם לאור בוהק להיווצר, להתלקח בחלל הריק מעבר לידי המושטות. דמיינו שאני מבצע את אקט הכישוף הבוטה והבלתי־ניתן־לערעור הזה תחת פיקוחם המלא של ג'יימס רנדי וכל צבאות הספקנים.* רוב האנשים, אני חושב, היו סקרנים למדי לגבי מה שקורה שם.

אבל כעת נניח שאני לא מופיע בטלוויזיה. אני לא מעוניין לחלוק את הכוח שלי, וגם לא את האמת שמאחוריי. אני רוצה לשמור את הכישוף שלי בסוד. אבל אני עדיין רוצה להטיל את הכשפים שלי מתי ואיפה שמתחשק לי. אני רוצה להטיל את הבזק האור הבוהק שלי כדי שאוכל לקרוא ספר ברכבת – מבלי שאנשים יהיו סקרנים. האם יש כישוף שעוצר סקרנות?

אכן כן! בכל פעם שמישהו ישאל "איך עשית את זה?", פשוט אומר "מדע!"

זה לא באמת הסבר, כמו שזה פשוט עוצר סקרנות. זה לא אומר לכם אם האור יתבהר או ידעך, ישנה צבע מבחינת גוון או עוצמה, וזה ודאי לא אומר לכם איך לייצר אור דומה בעצמכם. אתם למעשה לא

* ג'יימס רנדי (James Randi) היה קוסם שהתפרסם כחושף תרמיות פסאודו־מדעיות, נודע במיוחד בזכות "אתגר המליון" שהציע לשלם לכל אדם שיוכיח יכולת על־טבעית תחת תנאי בקרה מדעיים.

יודעים שום דבר מעבר למה שידעתם לפני שאמרתי את מילת הקסם. אבל אתם פונים וממשיכים הלאה, מרוצים מכך ששום דבר יוצא דופן לא מתרחש.

מה שיפה זה שאותו טריק עובד עם מתג רגיל של אור.

אתה לוחץ על מתג ונורת חשמל נדלקת. למה?

בבית הספר, מלמדים אותנו שמילת הקוד לנורה היא ”חשמל!” בשלב זה, אני מקווה, אתם נזהרים מלסמן את הנורה כ”מובנת” על בסיס הסבר כזה. האם האמירה ”חשמל!” מאפשרת לכם לבצע חישובים שישלטו בציפייה שלכם לחוויה? יש, לכל הפחות, עוד הרבה מאוד מה ללמוד. (קוראים עם תואר בפיזיקה מוזמנים להתעלם מהפסקה הזו ולהחליף אותה בבעיה בתורת האבולוציה, שבה תוכן התיאוריה נמצא בחישובים שמעט אנשים יודעים לבצע.)

אם הייתם חושבים שהנורה כלתי ניתנת להסבר מדעי, היא הייתה תופסת את מלוא תשומת הלב שלכם. הייתם עוזבים כל דבר אחר שאתם עושים, ומתמקדים באותה נורה.

אבל מה משמעות הביטוי ”ניתן להסבר מדעי”? זה אומר שמישהו אחר יודע איך הנורה עובדת. כשאומרים לכם שהנורה ”ניתנת להסבר מדעי”, אתם לא יודעים יותר משידעתם קודם לכן. אתם לא יודעים אם אור הנורה יתחזק או ידעך. אבל מכיוון שמישהו אחר יודע, זה מפחית את ערך הידע בעיניכם. אתם נהיים פחות סקרנים.

מישהו בטח יאמר, ”אם המדע לא היה מכיר את הנורה, היית יכול לזכות בתהילה ועושר על ידי חקירתה.” אבל אני לא מדבר על תאוות בצע. אני לא מדבר על שאיפות קריירה. אני מדבר על הרגש הגולמי של סקרנות – התחושה של להיות מסוקרן. למה שהסקרנות שלך תפחת בגלל שמישהו אחר, לא אתה, יודע איך הנורה עובדת? האם אין זו צרות־עין? לא מספיק שאתה תדע. אנשים אחרים מוכרחים להיות בורים גם הם, או שלא תוכל להיות מאושר?

מלבד סקרנות, ישנן מטרות נעלות אחרות שהידע עשוי לספק, כמו למשל התועלת החברתית של טכנולוגיה. עבור מטרות מעשיות מעין אלו, יש משמעות לכך שישות אחרת במרחב המקומי כבר

יודעת. אבל עבור הסקרנות שלי עצמי, למה שזה ישנה?*

חוץ מזה, חשבו על ההשלכות אם תאפשרו ל"מישהו אחר יודע את התשובה" לתפקד כעוצר-סקרנות. יום אחד אתם נכנסים לסלון שלכם ורואים פיל ירוק ענק, נראה כמרחף באוויר, מוקף בהילה של אור כסוף.

"מה לעזאזל?" אתם אומרים.

ואז בוקע קול מעל לפיל, שאומר:

"מישהו כבר יודע למה הפיל הזה כאן."

"אה", אתם אומרים, "אם ככה, אז לא חשוב" וממשיכים למטבח. אני לא מכיר את התיאוריה המאוחדת הגדולה של חוקי הפיזיקה השולטת ביקום הזה. אני גם לא יודע הרבה על אנטומיה אנושית, למעט המוח. אני לא יודע להצביע על מקומן של הכליות בגופי, ואני לא יכול להיזכר מה עושה הכבד בלי לבדוק. (אני לא גאה בזה. אך למרבה הצער, עם כל המתמטיקה שעלי ללמוד, לא סביר שאגיע לאנטומיה בזמן הקרוב.)

ככל שזה נוגע לסקרנות, האם בורותי לגבי חוקי הפיזיקה האולטימטיביים צריכה לסקרן אותי יותר מכך שאיני יודע כמעט כלום על מה שקורה בתוך גופי שלי?

אם הייתי מושיט את ידי ומטיל כישוף של אור, כנראה שהייתם מסתקרנים. האם אתם צריכים להיות פחות סקרנים לגבי עצם העובדה שהושטתי את ידי? כשאתם מרימים את הזרוע ומנופפים ביד, פעולה זו של רצון מתואמת על ידי המוח הקטן שלכם (בין שאר אזורי המוח). אני מוכן להתערב שאתם לא יודעים איך המוח הקטן עובד. אני יודע קצת – אמנם רק את צורתם הכללית של הדברים, לא מספיק כדי לבצע חישובים כלשהם... אז מה? מה זה משנה, אם אתם לא יודעים? למה צריך להיות מוסר כפול של סקרנות עבור כישוף ועבור תנועות ידיים?

הביטו בעצמכם במראה. האם אתם יודעים על מה אתם מביטים?

* "אמת עצובה היא, שאנחנו חושבים רק על חלק זעיר מכל התופעות שאנו נתקלים בהן..." (הגפא"ו)

האם אתם יודעים מה הוא שמביט החוצה מאחורי העיניים שלכם?
האם אתם יודעים מה אתם? עבור חלק מהשאלה הזו, למדע יש
תשובות, ועבור חלק אין לו תשובות. אבל למה שההבחנה הזו תשנה
משהו לגבי הסקרנות שלכם, אם אתם לא יודעים?
האם אתם יודעים איך עובדות הברכיים שלכם? האם אתם יודעים
איך יוצרו הנעליים שלכם? האם אתם יודעים למה צג המחשב שלכם
זוהר? האם אתם יודעים למה מים הם רטובים?
העולם סביבכם מלא בחידות.* קבעו סדרי עדיפויות, אם אין
ברירה. אך אל תתלוננו על כך שהמדע האכזר רוקן את העולם
ממסתורין. עם לוגיקה כזו, בסוף תתעלמו גם מפיל בסלון.

* ”העולם לעולם לא ירעב ממחסור בפלאות, אלא רק ממחסור בפליאה.” (ג.ק. צ’סטרטון)

45

חלק ממך באמת

מאמר קלאסי מאת דרו מקדרמוט, "אינטליגנציה מלאכותית פוגשת טיפשות טבעית", התייחס בבקורת לתוכנות AI שניסו לייצג רעיונות כמו אושר הוא מצב תודעתי באמצעות רשת סמנטית:

מצבתודעתי

^

הוא |

|

אושר

וכמובן שאין שום דבר בתוך הצומת של "אושר". זה בסך הכל טוקן *LISP** ריק עם שם שמרמז על משמעות.

אז מקדרמוט אומר, "מבחן טוב עבור המתכנת הממושמע הוא לנסות להשתמש בסמלים גנריים בנקודות מפתח ולראות אם הוא עדיין מתפעל מהמערכת שבנה. למשל, אם נשנה את שמו של "מצב תודעתי" ל-*G1073*... אז נקבל: הוא(אושר, *G1073*)** "... שנראה הרבה יותר מפוקפק."

* שפת תכנות ששימשה במשך שנים במחקר של בינה מלאכותית.

** לקוראים שאינם בקיאים בתחביר של שפות תכנות, ניתן לראות ברישום פונקציה(פריט א, פריט ב) ייצוג טקסטואלי לרשת הסמנטית שתוארה בתחילת הפרק. המילה שמחוץ לסוגריים מציינת את סוג הקשר הסמנטי שבין הפריטים. שבתוך הסוגריים.

או כפי שאני הייתי מנסח את הרעיון: אם הייתם מחליפים את כל השמות בסמלים אקראיים, ודאי לא הייתם מצליחים להבין את המשמעות של $G1071(G1072, G1073)$.

האם תוכנת ה-AI התכוונה לייצג המבורגרים? תפוחים? אושר? מי יודע? אם אתם מוחקים את השמות המרמזים, הם לא צומחים חזרה.

נניח שפיזיקאי אומר לכם ש"אור הוא גלים", ואתם מאמינים לפיזיקאי. כעת יש לכם רשת קטנה בראש שאומרת: הוא (אור, גלים).

אם מישהו שואל אתכם "ממה עשוי אור?" תוכלו לומר "גלים!". כפי שאומר מקדרמוט, "הבעיה היא לגרום לשומע להבחין במה שנאמר לו. לא 'להבין', אלא 'להבחין'". נניח שבמקום זאת היה הפיזיקאי אומר לכם, "אור עשוי מדברים קטנים ומסולסלים." (הוא לא, דרך אגב.) האם הייתם מבחינים בהבדל כלשהו בחוויה הצפויה?

איך תוכלו לדעת שלא כדאי לכם לסמוך על הידע לכאורה שלכם לגבי "אור הוא גלים"? מבחן אחד שאפשר ליישם הוא לשאול: "האם הייתי יכול ליצור את הידע הזה מחדש לו היה נמחק איכשהו ממוחי?"

רוח מעין זו מושלת גם בשיטה בה אנו משנים את השמות האינדיקטיביים בשפת LISP בתוכנת ה-AI שלנו לשמות חסרי משמעות, ואז בודקים אם מישהו אחר יכול להבין למה הם "מתייחסים" לכאורה. אותה רוח שורה גם באבחנה כי מודל AI אריתמטי שבלע ואז פלט:

ועוד (שבע, שש) = שלוש-עשרה

לא יוכל ליצור מחדש את הידע אם יימחק מהזיכרון, עד שאותו ידע יוזן מחדש למסד הנתונים ע"י אדם אחר. בדיוק כפי שאם הייתם שוכחים ש"אור הוא גלים", לא הייתם יכולים לקבל חזרה את הידע אלא באותה הדרך בה קיבלתם אותו מלכתחילה – על ידי כך ששאלתם פיזיקאי. לא הייתם יכולים ליצור את הידע הזה בעצמכם, באותה הדרך בה פיזיקאים יצרו אותו במקור.

אותן חוויות שמובילות אותנו לגבש אמונה, מחברות את האמונה הזו לידע אחר, לקלט חושי ולפלט מוטורי. אם אתם רואים בונה מכרסם בול עץ, אז אתם יודעים איך נראה הדבר הזה שמכרסם כולו עץ, ותוכלו לזהות אותו בעתיד, בין אם הוא נקרא "בונה" ובין אם לאו. אבל אם רכשתם את האמונות שלכם לגבי בונים על ידי כך שמישהו אחר סיפר לכם עובדות לגבי "בונים", ייתכן שלא תזהו בונה כשתראו אותו.

זוהי הסכנה הנוראה הטמונה בניסיון לספר לבינה מלאכותית עובדות שהיא לא יכולה ללמוד בעצמה. זוהי גם הסכנה הנוראה הטמונה בניסיון לספר למישהו אמיתות פיזיקליות שהוא לא יכול לאמת בעצמו. שהרי כשפיזיקאים אומרים "גל" הם לא מתכוונים "דבר קטן ומתפתל" אלא למושג מתמטי טהור.

כפי שדוידסון מציין, אם אתם מאמינים ש"בונים" חיים במדבריות, בעלי צבע לבן ושוקלים בכגרותם 150 קילו, אזי אין לכם שום אמונות לגבי בונים, נכונות או שגויות. האמונה שלכם לגבי "בונים" אינה נכונה מספיק כדי להיות שגויה. אם אין לכם מספיק ניסיון כדי ליצור מחדש אמונות כשהן נמחקות, אז האם יש לכם בכלל מספיק ניסיון כדי לחבר את האמונה הזו למשהו? ויטגנשטיין: "גלגל שניתן לסובבו אף ששום דבר אחר אינו נע עימו, אינו חלק מהמנגנון."^{*}

כמעט ברגע שהתחלתי לקרוא על AI – אפילו לפני שקראתי את מקדרמוט – הבנתי שזה יהיה רעיון ממש טוב לשאול את עצמי תמיד: "איך הייתי יוצר מחדש את הידע הזה אילו היה נמחק ממוחי?"

ככל שהמחיקה עמוקה יותר, כך המבחן מחמיר יותר. אם כל ההוכחות למשפט פיתגורס היו נמחקות ממוחי, האם הייתי מסוגל להוכיח אותו מחדש? אני חושב שכן. אם כל הידע על משפט פיתגורס היה נמחק ממוחי, האם הייתי מבחין במשפט פיתגורס על מנת להוכיחו מחדש? בזה קשה יותר להתרברב, מבלי להעמיד זאת למבחן, אך אם הייתם מראים לי משולש ישר-זווית עם ניצבים באורך

* מתוך חקירות פילוסופיות, מאת לודוויג ויטגנשטיין, בתרגומה של עדנה אולמן-מרגלית.

3 ו-4, ואומרים לי שאורך היתר ניתן לחישוב, אני חושב שהייתי מסוגל לחשב אותו, אם עדיין הייתי יודע את כל שאר המתמטיקה שלמדתי.

מה לגבי המושג של הוכחה מתמטית? אם מעולם לא הייתי שומע עליו, האם הייתי מסוגל להמציא את זה מחדש על בסיס אמונות אחרות שברשותי? הייתה תקופה בה האנושות לא הכירה את המושג הזה. מישהו ודאי המציא אותו. במה הוא הבחין? האם אני הייתי מבחין אילו הייתי רואה משהו חדשני באותה מידה וחשוב באותה מידה? האם הייתי מסוגל לחשוב רחוק כל כך מחוץ לקופסה? כמה מהידע שלכם הייתם יכולים ליצור מחדש? מאיזה עומק של מחיקה? זהו לא רק מבחן שנועד לסלק אמונות שאינן מחוברות דיין. זוהי דרך לייסד מעיין של ידע, ולא רק עובדה אחת.

רועה צאן בונה מערכת ספירה שמבוססת על השלכה של אבן אחת לתוך דלי בכל פעם שכבשה יוצאת מהדיר, והוצאה של אבן אחת בכל פעם שכבשה חוזרת. אם אתם, כחניכים, לא מבינים את המערכת הזו – אם עבורכם מדובר בקסם שעובד ללא שום סיבה נראית לעין – אז לעולם לא תדעו מה לעשות אם תפילו בטעות אבן נוספת לתוך הדלי. את מה שאינכם יכולים ליצור בעצמכם, אינכם יכולים ליצור מחדש כשהמצב דורש זאת. אינכם יכולים לחזור למקור, לשנות מעט את אחד הפרמטרים וליצור מחדש את הפלט, ללא המקור. אם "שתיים ועוד ארבע שווה שש" היא עבורכם עובדה גולמית, ואז אחד האלמנטים משתנה ל"חמש", איך תדעו ש"שתיים ועוד חמש שווה שבע" כשפשוט סיפרו לכם ש"שתיים ועוד ארבע שווה שש"?

אם אתם רואים צמח קטן שמפיל זרע בכל פעם שחולפת על פניו ציפור, לא יעלה בדעתכם שתוכלו להשתמש בצמח הזה כדי להפוך את סופרהכבשים לחציאוטומטי. למרות שלמדתם משהו שהיוצר המקורי היה ודאי משתמש בו כדי לשפר את המצאתו, לא תוכלו לחזור למקור וליצור אותה מחדש.

כשאתם מחזיקים בראשכם את המקור של מחשבה, אותה מחשבה תוכל להשתנות יחד איתכם ככל שאתם רוכשים ידע ומיומנויות

חדשים. כשאתם מחזיקים בראשכם את המקור של מחשבה, היא הופכת להיות חלק מכם באמת וגדלה יחד איתכם. שאפו להפוך את עצמכם למקור של כל מחשבה שראוי לחשוב. אם המחשבה הגיעה מבחוץ, וראו שהיא מגיעה גם מבפנים. שאלו את עצמכם ללא הרף: "איך הייתי יוצר את המחשבה הזו מחדש לו נמחקה ממוחי?" כשיש לכם תשובה, דמיינו שגם הידע הזה נמחק. וכשאתם מוצאים מעיין, ראו מה עוד יכול לנבוע.

1. Drew McDermott, "Artificial Intelligence Meets Natural Stupidity," *SIGART Newsletter*, no. 57 (1976): 4–9, doi:10.1145/1045339.1045340.

2. Richard Rorty, "Out of the Matrix: How the Late Philosopher Donald Davidson Showed That Reality Can't Be an Illusion," *The Boston Globe* (October 2003).

אינטרלוד: האמת הפשוטה

אני זוכרת עבודה שכתבתי על אקזיסטנציאליזם. המורה החזירה לי אותה עם ציון נכשל. היא הדגישה את המילים "אמת" ו"אמיתי" בכל מקום שהן הופיעו במאמר, בערך עשרים פעמים, עם סימן שאלה ליד כל אחת. היא רצתה לדעת למה התכוונתי כשכתבתי "אמת".

— דניאל איגן, עיתונאית

מאמר זה נועד להשיב לנו מבט נאיבי על האמת. מישהו אומר לכם: "שמן הנחשים הפלאי שלי מסוגל לרפא אתכם מסרטן הריאות בתוך שלושה שבועות בלבד." אתם משיבים: "האם מחקר קליני לא הראה שהטענה הזו אינה אמיתית?" אותו אדם עונה: "המושג הזה של 'אמת' הוא נאיבי למדי. למה אתה מתכוון כשאתה אומר 'אמיתי'?"

אנשים רבים שנסאלים כך, אינם יודעים לענות בפירוט קפדני ומושלם. אף על פי כן, לא יהיה זה נכון מצדם לנטוש את רעיון ה"אמת". הייתה תקופה שבה איש לא ידע את משוואות הכבידה בפירוט קפדני ומושלם, ובכל זאת, לו צעדתם קדימה מראש צוק, הייתם נופלים.

דבר שראיתי קורה לעיתים קרובות, במיוחד ברשימות תפוצה באינטרנט, הוא שבאמצע שיחה אחרת, מישהו אומר "X הוא אמת",

ואז פורץ ויכוח על השימוש במילה "אמת". מאמר זה לא נועד לשמש כאסמכתא אנציקלופדית עבור אותו ויכוח. תחת זאת, אני מקווה ששני הצדדים יקראו את המאמר הזה, ואז יחזרו לנושא בו דנו לפני שמישהו הטיל ספק בטבעה של האמת.

במאמר זה אני מציג שאלות. אם אתם רואים משהו שנראה כמו תשובה טריוויאלית, זו כנראה התשובה אליה אני מתכוון. הבחירה המובנת מאליה אינה הבחירה הטובה ביותר תמיד, אבל בחיי, לפעמים היא כן. אני לא מפסיק לחפש ברוגע שאני מוצא תשובה מובנת מאליה, אבל אם אני ממשיך לחפש, והתשובה שנראתה מובנת מאליה עדיין נראית כזאת, לא ארגיש אשם אם אשאר איתה. הו, כמובן, כולם חושבים ששתיים ועוד שתיים הם ארבע, כולם אומרים ששתיים ועוד שתיים הם ארבע, ובשגרה הארצית והמשעממת של חיי היומיום כולם מתנהגים כאילו שתיים ועוד שתיים הם ארבע, אבל כמה הם שתיים ועוד שתיים באמת, בסופו של דבר? ככל שאני מצליח להבין, ארבע. הם עדיין ארבע גם אם אני הוגה את השאלה בטון חגיגי ורבימשמעות. פשוט מדי, אתם אומרים? אולי, רק הפעם, החיים לא צריכים להיות מסובכים. האין זה מרענן?

אם אתם מאותם בני מזל שלהם נראתה השאלה טריוויאלית מלכתחילה, אני מקווה שהיא עדיין תיראה טריוויאלית בסוף. אם אתם מוצאים את עצמכם נבוכים מול שאלות עמוקות ורבות משמעות, זכרו שאם אתם יודעים בדיוק איך מערכת כלשהי עובדת, ויכולתם לבנות אותה מדליים וחלוקי נחל בעצמכם, אז היא לא אמורה להיות מסתורית עבורכם.

אם הבלבול מאיים כאשר אתם מפרשים מטאפורה כמטאפורה, נסו לפרש הכל באופן מילולי לחלוטין.

דמיינו שבעידן שלפני ההיסטוריה הכתובה או המתמטיקה הפורמלית, אני חי כרועה צאן שמתקשה לעקוב אחר הכבשים שלו. הכבשים שלי ישנות במכלאה, והמכלאה גבוהה מספיק כדי לשמור עליהן מפני

הזאבים המשוטטים בלילות. בכל יום עלי לשחרר את הכבשים שלי מהמכלאה למרעה, ובכל לילה עלי למצוא את הכבשים שלי ולהחזירן למכלאה. אם כבש אחד יישאר בחוץ, אמצא את גופתו המרוטשת בבוקר שלמחרת, אכולה למחצה על ידי זאבים. אבל זה מייאש כל כך, לסרוק את השדות במשך שעות בחיפוש אחר כבש אחרון, כשאני יודע שסביר להניח שכל הכבשים כבר במכלאה. לפעמים אני מוותר מוקדם, ובדרך כלל לא קורה כלום. אבל בערך בעשירית מהמקרים יש כבש מת בבוקר שלמחרת.

אילו רק הייתה דרך כלשהי לגלות אם עדיין ישנן כבשים במרעה, בלי כל אי־הנוחות של החיפוש! אני מנסה כמה שיטות: אני מטיל את מקלות הניחוש של השבט שלי; אני מאמן את כוחי הטלפתי באיתור כבשים באמצעות ראייה על־חושית; אני מתאמן למצוא סיבות להאמין שכל הכבשים נמצאות במכלאה. זה לא משנה דבר. בערך בעשירית מהפעמים שאני פורש מוקדם, אני מוצא כבש מת בבוקר שלמחרת. אולי אני מבין שהשיטות שלי לא עובדות, ואולי אני מתרץ בזהירות כל כישלון, אבל הדילמה שלי היא אותה הדילמה. אני יכול לבלות שעה בחיפוש בכל פינה ובכל חור אפשריים, כשברוב המקרים לא נותרו שום כבשים, או שאני יכול ללכת לישון מוקדם ולאבד, בממוצע, עשירית כבש.

אח־הצהריים אחד אני מרגיש עייף במיוחד. אני מטיל את מקלות הניחוש ומקלות הניחוש אומרים שכל הכבשים חזרו. אני מדמיין כל פינה וכל חור, ובדימוי אני לא רואה שום כבש. אני עדיין לא בטוח, אז אני מביט לתוך המכלאה ונראה שיש שם הרבה כבשים. אני סוקר את מאמצי הקודמים ומחליט שהייתי חרוץ במיוחד. זה מניח את דעתי, ואני הולך לישון. בבוקר שלמחרת אני מגלה שתי כבשים מתות. משהו בתוכי נשבר, ואני מתחיל לחשוב באופן יצירתי.

באותו יום, הלמות פטישים רמות נשמעות משער המכלאה. למחרת בבוקר, אני פותח את שער המכלאה רק מעט, ובכל פעם שיוצא כבש מהמכלאה, אני מטיל חלוק נחל לתוך דלי שממוסמר לעמוד ליד השער. אח־הצהריים, בכל פעם שכבש עובר בחזרה, אני

מוציא חלוק נחל אחד מהדלי. כשלא נותרים עוד חלוקי נחל בדלי, אני יכול להפסיק לחפש וללכת לישון. זהו רעיון מבריק. הוא יחולל מהפכה בתחום רעיית הצאן.

זאת לפחות הייתה התיאוריה. בפועל, נדרש דיוק משמעותי לפני שהשיטה התחילה לעבוד באופן אמין. מספר פעמים חיפשתי במשך שעות ולא מצאתי אף כבש, ולמחרת בבוקר לא היו גופות. בכל אחד מהמקרים הללו נדרשה מחשבה עמוקה כדי להבין היכן כשלה מערכת הדלי. כשחזרתי ערב אחד מחיפוש עקר, שחזרתי את מעשי והבנתי שהדלי כבר הכיל חלוקי נחל כשהתחלתי. זה, מתברר, היה רעיון גרוע. בפעם אחרת הטלתי לדלי חלוקי נחל באקראי בין הבוקר לאחרי-הצהריים על מנת להפיג את השעמום; גם זה היה רעיון גרוע, כפי שהבנתי לאחר חיפוש של כמה שעות. אבל תרגלתי את אומנות החלוקים שלי, והפכתי לאומן-חלוקים מיומן למדי.

אחרי-הצהריים אחד, איש לבוש בגדי לבן מהודרים, זרי דפנה, סנדלים וחליפת עסקים צועד בכבודות בשביל החול המוביל לשדות המרעה שלי.

“אני יכול לעזור לך?” אני שואל.

האיש מוציא ממעילו תג ומציג לי אותו, מוכיח מעל לכל צל של ספק שהוא מרקוס סופיסטיקוס מקסימוס, שליח מטעם הסנאט של רום. (אפשר לתהות אם מישוהו יכול היה לגנוב את התג, אך כוחם של תגים אלה כה גדול, שאם ישתמש בהם אדם אחר, הוא יהפוך למרקוס באותו הרגע.)

“קרא לי מארק,” הוא אומר. “אני כאן כדי להחרים את חלוקי הנחל הקסומים, בשם הסנאט. אסור שחפצים בעלי עוצמה כה גדולה ייפלו לידיהם של עמי ארצות.”

“השוליה הארור,” אני רוטן מתחת לשפמי, “הוא שוב קשקש עם כל הכפר.” אני מסתכל על פניו הרציניים של מארק, ונאנח. “אלה לא חלוקי נחל קסומים,” אני אומר בקול רם. “רק אבנים רגילות שאספתי מהקרקע.”

הבזק של בלבול חולף על פניו של מארק, ואז הם מתבהרים שוב.

“אני כאן בשביל הדלי הקסום!” הוא מכריז.
 “זה לא דלי קסום, אני אומר בעייפות. הייתי משתמש בו
 לגרביים מלוכלכים.”

פניו של מארק מבולבלים. “אז איפה הקסם?” הוא תובע.
 שאלה מעניינת. “קשה להסביר, אני אומר.”

השוליה הנוכחי שלי, אוטרי, נמשך למהומה ומתנדב להסביר:
 “זה המפלס של חלוקי הנחל בדלי, אומר אוטרי. יש מפלס קסום
 של חלוקי נחל, ואתה צריך להגיע בדיוק למפלס הנכון, או שזה לא
 עובד. אם תזרוק עוד חלוקי נחל, או אם תוציא כמה, הדלי כבר לא
 יהיה במפלס הקסום. כרגע, המפלס הקסום הוא, אוטרי מציץ לתוך
 הדלי, “בערך שליש גובה.”

“הבנתי!” אומר מארק בהתלהבות. מהכיס האחורי שלו שולף
 מארק דלי משלו, וערימה של חלוקי נחל. הוא אוסף כמה חופנים של
 חלוקי נחל, ועורם אותם לתוך הדלי. מארק מביט לתוך הדלי, ואומד
 את כמות החלוקים שבתוכו. “הנה, אומר מארק, “המפלס הקסום
 של הדלי הזה הוא חצי גובה. ככה?”

“לא!” אומר אוטרי בחדות. “חצי גובה זה לא המפלס הקסום.
 המפלס הקסום הוא בערך שליש. חצי גובה הוא בהחלט לא-קסום.
 חוץ מזה, אתה משתמש בדלי הלא נכון.”

הוא פונה אליי, מבולבל. “חשבתי שאמרת שהדלי אינו קסום?”
 “הוא לא, אני אומר. כבש עובר דרך השער, ואני משליך לדלי
 חלוק נחל נוסף. “חוץ מזה, אני משגיח על הכבשים. דבר עם אוטרי.”
 מארק מביט בחשדנות בחלוק הנחל שהטלתי, אבל מחליט
 להמתין עם השאלה. מארק פונה לאוטרי ומזדקק ביהירות. “זו מדינה
 חופשית, אומר מארק, “תחת הדיקטטורה הנאורה של הסנאט,
 כמובן. אני יכול להטיל אילו חלוקי נחל שארצה לתוך איזה דלי
 שארצה.”

אוטרי שוקל זאת. “לא, אתה לא יכול, הוא אומר לבסוף, “לא
 יהיה שום קסם.”

“תראה, אומר מארק בסבלנות, “צפיתי בכך בדקדקנות. הסתכלת

אל תוך הדלי שלך, בדקת את מפלס חלוקי הנחל, והכרזת שהוא המפלס הקסום. עשיתי בדיוק את אותו הדבר. "זה לא עובד ככה, " אומר אוטרי.

"הו, אני מבין, " אומר מארק, "המפלס בדלי שלי לא קסום, אבל המפלס בדלי שלך כן קסום, זה מה שאתה אומר? מה הופך את הדלי שלך לכל כך הרבה יותר טוב משלי, הא? "

"ובכן, " אומר אוטרי, "אם היינו מרוקנים את הדלי שלך, ואז שופכים את כל חלוקי הנחל מהדלי שלי לתוך הדלי שלך, אז לדלי שלך היה המפלס הקסום. ישנו גם הליך שבו אנו יכולים להשתמש כדי לבדוק האם לדלי שלך יש את המפלס הקסום, אם אנחנו יודעים שלדלי שלי יש את המפלס הקסום. אנחנו קוראים לזה פעולת השוואת דליים."

כבש נוסף עובר, ואני מטיל עוד חלוק נחל.

"הוא זרק עוד חלוק נחל! " קורא מארק. "ואני מניח שאתה תטען שגם המפלס החדש קסום? אני יכול להטיל חלוקי נחל לדלי שלך עד שהמפלס אצלך יהיה זהה לשלי, ואז הדליים שלנו יסכימו. אתה פשוט משווה את הדלי שלי לדלי שלך כדי לקבוע אם אתה חושב שהמפלס הוא 'קסום' או לא. ובכן, אני חושב שהדלי שלך אינו קסום, כי אין לו את אותו מפלס חלוקי נחל כמו אצלי. אז הנה! "

"חכה, " אומר אוטרי, "אתה לא מבין—"

"כשאתה אומר 'מפלס קסום', אתה פשוט מתכוון למפלס חלוקי הנחל בדלי שלך. וכשאני אומר 'מפלס קסום', אני מתכוון למפלס חלוקי הנחל בדלי שלי. לאור זאת, אתה מסתכל על הדלי שלי ואומר שהוא 'לא קסום', אבל למילה 'קסום' יש משמעויות שונות עבור אנשים שונים. אתה צריך לציין אל הקסם של מי אתה מתכוון. אתה צריך לומר שלדלי שלי אין את 'המפלס הקסום של אוטרי', ואני אומר שלדלי שלך אין את 'המפלס הקסום של מארק'. בניסוח כזה, אין שום סתירה."

"אבל—" אומר אוטרי בחוסר אונים.

"אנשים שונים יכולים להחזיק בדליים שונים עם מפלסים שונים"

של חלוקי נחל, מה שמוכיח שכל העסק הזה של 'קסם' הוא שרירותי וסובייקטיבי לחלוטין.

"מארק, אני אומר, "האם מישהו סיפר לך מה חלוקי הנחל האלה

עושים?"

"עושים?" שואל מארק. "חשבתי שהם פשוט קסומים."

"אם חלוקי הנחל לא היו עושים דבר, "אומר אוטרי, "מפקח

היעילות שלנו היה מוחק את התהליך הזה מהשגרה היומית שלנו לפי תקן ISO 9000."

"מה שמו של המפקח שלך?"

"דארווין, "אומר אוטרי.

"הממ, "אומר מארק. "לצ'ארלס באמת יש מוניטין של מפקח

קפדן. אז האם חלוקי הנחל מברכים את הכבשים, ומעודדים פריה ורבייה בקרב הצאן?"

"לא, "אני אומר. "מעלתם של חלוקי הנחל היא זו: אם אנו

מסתכלים אל תוך הדלי ורואים שהדלי ריק מחלוקים, אנו יודעים שבאותו האופן, שדות המרעה ריקים מכבשים. אם איננו משתמשים בדלי, עלינו לחפש ולחפש עד החשיכה, שמא נותר כבש כלשהו בחוץ. או אם אנו מפסיקים את עבודתנו מוקדם, אזי לעיתים נמצא בבוקר שלמחרת כבש מת, שכן הזאבים טורפים כל כבש שנותר בחוץ. אם אנו מסתכלים בדלי, אנו יודעים מתי כל הכבשים בבית, וניתן לפרוש למיטה ללא חשש."

מארק שוקל זאת. "זה נשמע קלוש למדי, "הוא אומר לבסוף.

"חשבת להשתמש במקלות ניחוש? מקלות ניחוש אינם טועים לעולם, או לפחות, כל מי שאומר שהם טועים מועלה על המוקד. זו דרך כואבת מאוד למות. מכאן נובע שמקלות ניחוש אינם טועים לעולם."

"אתה מוזמן להשתמש במקלות ניחוש אם זה רצונך, "אני אומר.

"הו, אלוהים אדירים, כמובן שלא, "אומר מארק. "הם לא טועים

לעולם, פועלים ללא דופי בכל הזדמנות, בדיוק כפי שמצופה מכלים קדושים שכאלה. אבל מה אם הייתי מוצא כבש מת בבוקר שלמחרת?

אני משתמש במקלות הניחוש רק כשאין להם שום אפשרות לטעות. אחרת עוד עלולים לשרוף אותי חי. אז איך הדלי הקסום שלך עובד?"

איך הדלי עובד...? מוטב שאתחיל עם המקרה הפשוט ביותר. "ובכן", אני אומר, "נניח ששדות המרעה ריקים, והדלי אינו ריק. במקרה כזה נבזזו שעות בחיפושים אחר כבש שכלל לא נמצא שם. ואם יש כבש בשדות המרעה, אבל הדלי ריק, אז אוטרי ואני נלך לישון מוקדם מדי, ונמצא כבש מת בבוקר שלמחרת. אז דלי ריק הוא קסום אם ורק אם שדות המרעה ריקים –"

"רגע אחד, " אומר אוטרי. "זה נשמע לי כמו טאוטולוגיה ריקה מתוכן. האם לא ברור לחלוטין שדלי ריק ושדות מרעה ריקים הם אותו הדבר?"

"ממש לא ריק מתוכן, " אני אומר. "הנה אנלוגיה: הלוגיקן אלפרד טרסקי אמר פעם שהטענה 'שלג הוא לבן' היא אמיתית אם ורק אם שלג הוא לבן. אם תוכל להבין זאת, תהיה מסוגל להבין מדוע דלי ריק הוא קסום אם ורק אם שדות המרעה ריקים מכבשים."

"רגע, " אומר מארק. "אלה דליים. אין להם שום קשר לכבשים. דליים וכבשים הם בביור דברים שונים. אין שום דרך בה יוכלו הכבשים לקיים אינטראקציה עם הדלי."

"אז מאיפה אתה חושב שמגיע הקסם?" שואל אוטרי.

מארק שוקל את השאלה. "אמרת שאפשר להשוות שני דליים כדי לבדוק אם יש להם אותו מפלס... אני יכול לראות איך דליים יכולים לקיים אינטראקציה עם דליים אחרים. אולי כשאתה מקבל אוסף גדול של דליים, ולכולם יש את אותו המפלס, זה מה שיוצר את הקסם. אני אקרא לזה התיאוריה הקוהרנטיסטית של דליים קסומים."

"מעניין, " אומר אוטרי. "אני יודע שהאדון שלי עובד על מערכת עם יותר מדלי אחד – הוא אומר שזה יהיה טוב יותר בגלל 'יתירות' ו'תיקון שגיאות'. נשמע לי כמו קוהרנטיזם."

"זה לא ממש אותו דבר –" אני מתחיל לומר.

"בואו נבחן את התיאוריה הקוהרנטיסטית של הקסם, " אומר אוטרי. "אני רואה שיש לך בכיס האחורי עוד חמישה דליים. אני אתן

לך את הדלי שבו אנחנו משתמשים, ואז תוכל למלא את שאר הדליים לאותו מפלס –

מארק נרתע באימה. "עצור! הדליים האלה עוברים במשפחה שלי במשך דורות, ותמיד עם אותו המפלס! אם אקבל את הדלי שלך, אוסף הדליים שלי יהפוך לפחות קוהרנטי, והקסם ייעלם!"
 "אבל הדליים הנוכחיים שלך לא קשורים לכבשים בשום צורה!"
 מוחה אוטרי.

מארק נראה מרוגז. "תראה, כמו שכבר הסברתי קודם, ברור שאין שום אפשרות שכבשים יקיימו אינטראקציות עם דליים. רק דליים יכולים לקיים אינטראקציות עם דליים אחרים."
 "בכל פעם שעובר כבש אני מטיל חלוק נחל לדלי, אני מציין.
 "כשעובר כבש, אתה מטיל חלוק נחל?" אומר מארק. "מה זה קשור בכלל?"

"זאת אינטראקציה בין הכבשים לבין חלוקי הנחל, אני עונה.
 "לא, זאת אינטראקציה בין חלוקי הנחל לבינק, אומר מארק.
 "הקסם לא מגיע מהכבשים, הוא מגיע ממך. הכבשים כשלעצמן בכירור לא קסומות. הקסם חייב לבוא מאיפשהו, בדרך לדלי."
 אני מצביע על מנגנון עץ המחובר לשער. "אתה רואה את רצועת הברד שמשתלשלת ממתקן העץ הזה? אנחנו עדיין עובדים על זה – המנגנון עדיין לא עובד באופן אמין – אבל בעיקרון, בכל פעם שכבש עובר, הוא מסיט הצידה את הברד. כשהברד זז הצידה, נופל חלוק נחל אחד מהמיכל אל תוך הדלי. וכך אוטרי ואני לא צריכים להטיל את חלוקי הנחל בעצמנו."
 מארק מקמט את מצחו. "אני לא בטוח שהבנתי... האם הברד קסום?"

אני מושך בכתפיי. "הזמנתי אותו באינטרנט מחברה בשם 'ברירות טבעיות'. הברד נקרא 'מערך חישה'. אני עוצר לרגע, למול הבעותיהם הספקניות של מארק ואוטרי. "אני מודה שהשמות קצת ניראייג'יים. הנקודה החשובה היא שכל כבש שעובר שם מפעיל שרשרת של סיבה ותוצאה שמסתיימת בחלוק נחל בדלי. לאחר מכן אתה יכול להשוות

את הדלי לדליים אחרים, וכן הלאה. "אני עדיין לא מבין," אומר מארק. "אי אפשר להכניס כבש לתוך דלי. רק חלוקי נחל נכנסים לדליים, ומובן מאליו שחלוקי נחל מקיימים אינטראקציה רק עם חלוקי נחל אחרים." "הכבשים מקיימות אינטראקציה עם דברים שמקיימים אינטראקציה עם חלוקי נחל..." אני מחפש אנלוגיה. "נניח שאתה מביט מטה אל שרוכי הנעליים שלך. פוטון עוזב את השמש; ואז נע מטה דרך האטמוספירה של כדור הארץ; ואז פוגע בשרוכי הנעליים שלך; ואז עובר דרך האישון של העין שלך; ואז פוגע ברשתית; ואז נספג על ידי קנה או מדוך.* האנרגיה של הפוטון גורמת לנוירון המחובר אליו לירות, מה שגורם לנוירונים אחרים לירות. דפוס פעולה עצבי בקליפת הראייה שלך יכול לקיים אינטראקציה עם האמונות שיש לך לגבי שרוכי נעליך, מכיוון שאמונות לגבי שרוכי נעליים מתקיימות גם הן על מצע עצבי. אם אתה מסוגל להבין את זה, אתה אמור להיות מסוגל לראות איך כבש חולף גורם לחלוק נחל להיכנס לדלי."

"באיזו נקודה בתהליך חלוק הנחל הופך לקסום?" אומר מארק. "זה... אמנם..." עכשיו אני מתחיל להתבלבל. אני מטלטל את ראשי כדי לפזר את הערפל. כל זה נראה פשוט למדי כשהתעוררתי הבוקר, ומערכת חלוקייהנחל-יהדלי לא הסתבכה מאז. "הרבה יותר קל להבין את זה אם זוכרים שהמטרה של המערכת היא לעקוב אחר הכבשים." מארק נאנח בעצב. "לא חשוב... ניכר שאתה פשוט לא יודע. אולי כל חלוקי הנחל קסומים מלכתחילה, עוד לפני שהם נכנסים לדלי. אנחנו יכולים לקרוא לעמדה הזו פֶּאן-חלוקיזם." "הא!" אומר אוטרי, קולו מלא בוז. "הבל וכזב! לא כל חלוקי הנחל נבראו שווים. חלוקי הנחל בדלי שלך אינם קסומים. הם סתם גושי אבן!" מבטו של מארק הופך נוקשה. "עכשיו," הוא קורא, "עכשיו אתה

* קנים ומדוכים (Rods and Cones) הם שני הסוגים של קולטני אור ברשתית העין.

רואה את הסכנה שבדרך בה אתה צועד ! ברגע שתאמר שחלוקי הנחל של אנשים מסוימים הם קסומים ושל אחרים לא, גאוותך תעביר אותך על דעתך ! אתה תראה את עצמך נעלה על כולם, וכך תביא למפלתך ! אנשים רבים לאורך ההיסטוריה עינו ורצחו אחרים משום שחשבו שחלוקי הנחל שלהם הם הנעלים מכל ! " שמץ של התנשאות חודר לקולו של מארק. "הסגידה למפלס של חלוקי נחל כ'קסום' מרמזת על כך שיש מפלס חלוקי נחל מוחלט ב'דלי עליון' כלשהו. איש אינו מאמין בדלי עליון בימינו."

"ראשית," אני אומר. "כבשים הן לא חלוקי נחל מוחלטים. שנית, אני לא חושב שהדלי שלי מכיל את הכבשים הלכה למעשה. שלישית, אני לא סוגד למפלס הדלי שלי כמושלם – אני מתאים אותו מדי פעם – ואני עושה זאת משום שאכפת לי מהכבשים."

"חוץ מזה," אומר אוטרי, "אדם שמאמין שבעלות על חלוקי נחל מוחלטים מאפשרת לו לענות ולרצוח עושה טעות שאין לה שום קשר לדליים. אתה פותר את הבעיה הלא נכונה."

מארק משיב לעצמו את קור רוחו. "אני מניח שיהיה זה נאיבי לצפות ליותר מרועי צאן פשוטים. אתם בטח מאמינים ששלג הוא לבן, נכון?"

"אממ... כן?" אומר אוטרי.

"לא מפריע לך שיוזף סטלין האמין ששלג הוא לבן?"

"אממ... לא?" אומר אוטרי.

מארק נועץ באוטרי מבט לא מאמין, ולבסוף מושך בכתפיו. "בוא נניח, רק לשם הדיון, שחלוקי הנחל שלך קסומים ושלי לא. אתה יכול לומר לי מה ההבדל?"

"חלוקי הנחל שלי מייצגים את הכבשים!" קורא אוטרי בניצחון. "לחלוקי הנחל שלך אין את תכונת הייצוגיות, ולכן הם לא יעבדו. הם ריקים ממשמעות. הבט בהם. אין להם הילה של תוכן סמנטי. הם בסך הכל חלוקי נחל. אתה זקוק לדלי עם כוחות סיבתיים מיוחדים."

"אה!" אומר מארק. "כוחות סיבתיים מיוחדים, במקום קסם." "בדיוק," אומר אוטרי. "אני לא מאמין באמונות טפלות. להניח

את קיומו של קסם, בימינו אנו, לא יתקבל בעין יפה בקהילת רועי הצאן הבינלאומית. מצאנו שהנחת קיומו של קסם פשוט לא עובדת כהסבר לתופעות בתחום רעיית הצאן. אז כשאני נתקל במשהו שאני לא מבין, ואני רוצה להסביר אותו באמצעות מודל נטול פירוט פנימי שאינו מספק תחזיות אפילו בדיעבד, אני מניח את קיומם של כוחות סיבתיים מיוחדים. אם זה לא עובד, אמשיך לשלב הבא ואקרא לזה תופעה של מערכת מתהווה.

"איזה סוג של כוחות מיוחדים יש לדלי?" שואל מארק.

"הממ", אומר אוטרי. "אולי הדלי הזה ניחן ביחס של 'מידע על-כלפי שדות המרעה. זה יסביר מדוע הוא עובד – כשהדלי ריק, זה אומר ששדות המרעה ריקים."

"איפה מצאת את הדלי הזה?" אומר מארק. "ואיך הבנת שיש לו יחס 'מידע על-' לשדות המרעה?"

"זה דלי רגיל", אני אומר. "נהגתי לעמוד עליו כדי לטפס על עצים... אני לא חושב שהשאלה הזו אמורה להיות קשה." "אני מדבר עם אוטרי", אומר מארק.

"עליך לרתום את הדלי לשדות המרעה, ואת חלוקי הנחל לכבשים, באמצעות טקס קסום – סליחה, תהליך מתהווה בעל כוחות סיבתיים מיוחדים שהמאסטר שלי גילה, "מסביר אוטרי.

הוא מנסה לתאר את הטקס, בעוד מארק מהנהן בארשת מלומדת. "אתה צריך להטיל חלוק נחל בכל פעם שכבש עוזב דרך השער?" אומר מארק. "ולהוציא חלוק נחל בכל פעם שכבש חוזר?" אוטרי מהנהן. "כן."

"זה בטח ממש קשה", אומר מארק בהערכה.

אוטרי נמלא חיים, סופג את אהדתו של מארק כמו גשם. "בדיוק!" אומר אוטרי. "זה קשה מאוד מבחינה רגשית. כשהדלי שומר על המפּלס שלו במשך זמן מה, אתה... נוטה להיקשר למפּלס הזה." באותו רגע עובר כבש ויוצא דרך השער. אוטרי רואה אותו. הוא מתכופף, מרים חלוק נחל ומניף אותו באוויר. "הביטו!" מכריז אוטרי. "כבש יצא את השער! זה החלוק-נחל ילך אל הדלי, זה המפּלס ילך"

למיתה, והמפלס החדש ילך לחיים טובים – היה שלום, מפלס ישן ואהוב – “כבש נוסף עובר. אוטרי, שקוע כל כך בדרמה של עצמו, מפספס אותו, אז אני משליך חלוק נחל לדלי. אוטרי עדיין מדבר: –” יען כי זהו המבחן העילאי של הרועה – להטיל את חלוק הנחל, יהיה הדבר מייסר ככל שיהיה, ויהיה המפלס הישן יקר לליבו ככל שיהיה.

שכן, רק הטובים שברועי הצאן עומדים בדרישה כה מחמירה – “אוטרי, אני אומר, “אם ברצונך להיות רועה צאן דגול יום אחד, תלמד לסתום את הפה ולהשליך את חלוק הנחל. בלי בלאגן. בלי דרמה. פשוט תעשה את זה.”

“והטקס הזה, ” אומר מארק, “רותם את חלוקי הנחל לכבשים על ידי חוקי הקסם של זיקה וצימוד, כמו בובת וודו.” אוטרי נרתע ומביט סביבו. “אנא ! אל תקרא לזה זיקה וצימוד. אנו רועי הצאן, מתנגדים לאמונות טפלות. השתמש במילה ‘התכוונות’, או משהו כזה.”

“אני יכול להעיף מבט בחלוק נחל ?” אומר מארק. “בטח, ” אני אומר. אני מוציא את אחד מחלוקי הנחל מהדלי, וזורק אותו למארק. מיד אחר כך אני מושיט יד לאדמה, מרים חלוק נחל אחר, ומשליך אותו לדלי.

אוטרי מביט בי, מבולבל. “לא הרסת את זה הרגע ?” אני מושך בכתפיי. “אני לא חושב. נדע שהרסתי אם יהיה כבש מת מחר בבוקר, או אם נחפש כמה שעות ולא נמצא אף כבש.” “אבל –” אומר אוטרי.

“לימדתי אותך את כל מה שאתה יודע, אבל לא לימדתי אותך את כל מה שאני יודע, ” אני אומר.

מארק בוחן את חלוק הנחל, בוהה בו בריכוז. הוא מעביר את ידו מעל חלוק הנחל וממלמל כמה מילים, ואז מניד בראשו לשלילה. “אני לא חש שום קסם, ” הוא אומר. “סליחה. אני לא חש שום התכוונות.” “לחלוק נחל יש התכוונות רק אם הוא בתוך דלי קס- מתהווה מיוחד, ” אומר אוטרי. “אחרת הוא סתם חלוק נחל.” “אין בעיה, ” אני אומר. אני מוציא חלוק נחל מהדלי, ומשליך אותו

הצידה. מיד לאחר מכן, אני ניגש למקום שבו עומד מארק, טופח על ידו המחזיקה חלוק נחל, ואומר: "אני מכריז בזאת על יד זו כחלק מהדלי הקסום!" ואז אני שב למקומי ליד השער.

אוטרי צוחק. "עכשיו אתה סתם מרושע."

אני מהנהן, שכן זהו אכן המצב.

"אבל זה באמת הולך לעבוד?" אומר אוטרי.

אני מהנהן שוב, בתקווה שאני צודק. עשיתי את זה בעבר עם שני דליים, ובעיקרון, לא אמור להיות שום הבדל בין היד של מארק לבין דלי. גם אם ידו של מארק חדרה באֶלֶן ויטל המבדיל חומר חי מחומר מת, הטריק אמור להצליח ממש כפי שהיה מצליח אילו היה מארק פסל שיש.

מארק מסתכל על ידו בחשש מסוים. "אז... לחלוק הנחל יש עכשיו

התכוונות שוב?"

"יאפ", אני אומר. "אל תוסיף עוד חלוקי נחל לאלה שביד שלך,

ואל תשליך את זה שיש לך, או שתשבור את הטקס."

מארק מהנהן ברצינות. ואז הוא חוזר לבחון את חלוק הנחל. "אני מבין עכשיו איך הצאן שלך גדל כל כך," אומר מארק. "עם דלי בעל כוח שכזה, תוכל להטיל עוד ועוד חלוקי נחל, ועוד ועוד כבשים היו חוזרות מהשדות. יכולת להתחיל עם כבשים בודדות, לתת להן לצאת, ואז למלא את הדלי עד גדותיו לפני שהן חוזרות. ואם הטיפול בכל כך הרבה כבשים היה הופך למייגע, יכולת להניח לכולן לצאת, ואז לרוקן כמעט את כל חלוקי הנחל מהדלי, כך שרק בודדות יחזרו... להגדיל שוב את הצאן כשהגיע הזמן לגזוז... אלוהים אדירים, בן אדם! האם אתה קולט את העוצמה האדירה של הטקס שגילית? אני יכול רק לדמיין את ההשלכות. האנושות עשויה לזנק עשור קדימה – לא, מאה שלמה!"

"זה לא עובד ככה," אני אומר. "אם אתה מוסיף חלוק נחל כשלא

יצא אף כבש, או מסיר חלוק נחל כשלא נכנס שום כבש, הדבר שובר את הטקס. הכוח לא נשאר בחלוקי הנחל, אלא נעלם בבת אחת, כמו בועת סבון שמתבקעת."

פניו של מארק מביעות אכזבה מרה. "אתה בטוח?"
 אני מהנהן. "ניסיתי את זה וזה לא עבד."
 מארק נאנח בכבדות. "ה... מתמטיקה הזו... נראתה כל כך
 עוצמתית ומועילה... טוב נו, הנה הלכה לה הקידמה האנושית."
 "מארק, זה היה רעיון מבריק," אומר אוטרי בהתלהבות. "לא
 חשבתי על כך בעצמי, אבל הרעיון כל כך מתבקש... זה יכול היה
 לחסוך כמות עצומה של משאבים... חייבת להיות דרך להציל את
 התוכנית שלך! אולי אם ננסה דליים שונים נצליח למצוא אחד
 שישמור על הכוח הקסו- על ההתכוונות בחלוקי הנחל, גם בלי
 הטקס. או שאולי ננסה חלוקי נחל אחרים. אולי לחלוקי הנחל שלנו
 פשוט חסרות התכוונות הדרושות להתכוונות אינהרנטית. מה אם
 ננסה להשתמש באבנים מגולפות בדמות כבשים זעירות? או שפשוט
 נכתוב 'כבש' על חלוקי הנחל. אולי זה יספיק."
 "לא הולך לעבוד," אני חוזה ביובש.

אוטרי ממשיך. "אולי אנחנו צריכים חלוקי נחל אורגניים, במקום
 חלוקי נחל מסיליקון... או שאולי אנחנו צריכים להשתמש באבני חן
 יקרות. המחיר של אבני חן מכפיל את עצמו כל שמונה-עשר חודשים,
 אז נוכל לקנות חופן של אבני חן זולות עכשיו, לחכות, ובעוד עשרים
 שנה הן יהיו ממש יקרות."

"ניסית להוסיף חלוקי נחל כדי ליצור עוד כבשים, וזה לא עבד?"
 שואל אותי מארק. "מה בדיוק עשית?"

"לקחתי חופן של שטרות כסף. אחר כך החבאתי את שטרות הכסף
 מתחת לשמיכה שלי, אחד אחד. בכל פעם שהחבאתי שטר נוסף,
 לקחתי מהדק נייר מתוך קופסה, ויצרתי ערימה קטנה של מהדקים.
 נזהרתי לא לספור בלב, כך שכל מה שידעתי הוא שיש 'הרבה'
 שטרות כסף, ו'הרבה' מהדקי נייר. לאחר שכל השטרות הוחבאו
 מתחת לשמיכה, הוספתי מהדק אחד נוסף לערימה. זו היתה הפעולה
 המקבילה להטלת חלוק נחל נוסף לדלי. ואז התחלתי להוציא שטרות
 מתחת לשמיכה, ולהחזיר את אטבי הנייר לקופסה. כשסיימתי, נותר
 מהדק נייר בודד."

"מה משמעות התוצאה הזו?" שואל אוטרי.
 "זה אומר שהטריק לא עבד. ברגע שהטקס נשבר באותו צעד
 שגוי אחד – הכוח לא נשאר, אלא התפוגג מיד. ערימת מהדקי הנייר
 וערימת שטרות הכסף כבר לא התרוקנו באותו הזמן."
 "ממש ניסית את זה?" שואל מארק.

"כן," אני אומר, "ממש ביצעתי את הניסוי, כדי לוודא שהתוצאה
 תואמת את התחזית התיאורטית שלי. יש לי חיבה סנטימנטלית
 לשיטה המדעית, גם כשהיא נראית אבסורדית. חוץ מזה, מה אם הייתי
 טועה?"

"אם זה היה עובד," אומר מארק, "היית אשם בזיוף! תאר לעצמך
 מה היה קורה אילו כולם היו עושים זאת. הכלכלה הייתה קורסת!
 לכולם היו מיליארדי דולרים, אך לא היה נותר דבר שניתן לקנות
 בכסף!"

"בכלל לא," אני עונה. "על פי אותו ההגיון שלפיו הוספת מהדק
 נייר לערימה יוצרת שטר כסף נוסף, יצירת שטר כסף נוסף הייתה
 יוצרת סחורות ושירותים בשווי שטר נוסף."
 מארק מנענע בראשו לשלילה. "זיוף הוא עדיין פשע... לא היית
 צריך לנסות."

"הייתי בטוח במידה סבירה שאיכשל."
 "אהא!" אומר מארק. "אתה ציפית להיכשל! לא האמנת שאתה
 יכול לעשות את זה!"

"אכן," אני מודה. "ניחשת את הציפיות שלי בדיוק מדהים."
 "ובכן, זאת הבעיה," אומר מארק בנחישות. "קסם ניזון מאמונה
 ומכוח רצון. אם אתה לא מאמין שאתה יכול לעשות את זה, אתה
 לא יכול. עליך לשנות את אמונתך לגבי תוצאת הניסוי. זה ישנה את
 התוצאה עצמה."

"מצחיק," אני אומר בנימה נוסטלגית, "זה מה שאוטרי אמר
 כשסיפרתי לו על שיטת חלוקיה הנחליוהדלי. שזה מגוחך מכדי שהוא
 יאמין בזה, ולכן זה לא יעבוד עבורו."
 "איך שכנעת אותו?" שואל מארק.

”אמרתי לו לסתום את הפה ולמלא אחר ההוראות,“ אני אומר,
”וכשהשיטה עבדה, אוטרי התחיל להאמין בה.“
מארק מקמט את מצחו, מבולבל. ”זה לא הגיוני. זה לא פותר את
דילמת הביצה והתרנגולת היסודית.“
”בטח שכן. שיטת הדלי עובדת בין אם אתה מאמין בה ובין אם
לא.“

”זה אבסורד !“ נזעק מארק. ”אני לא מאמין בקסם שעובד בין אם
אתה מאמין בו ובין אם לא !“

”גם אני אמרתי את זה,“ מצטרף אוטרי. ”מסתבר שטעיתי.“
מארק מכווץ את פניו בריכוז. ”אבל... אם לא האמנת בקסם
שעובד בין אם אתה מאמין בו ובין אם לא, אז למה שיטת הדלי עבדה
כשלא האמנת בה ? האם האמנת בקסם שעובד בין אם אתה מאמין בו
ובין אם לא, בין אם האמנת בקסם שעובד בין אם אתה מאמין בו ובין
אם לא ?“

”לא... נראה לי...“ אומר אוטרי בספק.

”אז אם לא האמנת בקסם שעובד בין אם אתה... רגע, אני צריך
נייר ועיפרון—“ מארק משרבט בפראות, בוחן את התוצאה בחדשנות,
הופך את פיסת הנייר, ואז מוותר. ”לא משנה,“ הוא אומר. ”קסם
פשוט הוא קשה מספיק להבנה. מטאקסם הוא מעבר ליכולתי.“
”מארק, אני לא חושב שאתה מבין את אומנות הדליים,“ אני
אומר. ”לא מדובר בשימוש בחלוקי נחל כדי לשלוט בכבשים. מדובר
בלתת לכבשים לשלוט בחלוקי הנחל. באומנות מסוג זה, אין צורך
להאמין מראש שהיא תעבוד. קודם האומנות עובדת, ורק אחר כך
מתחילים להאמין בה.“

”או כך לפחות אתה מאמין,“ אומר מארק.

”כך אני מאמין,“ אני עונה, ”משום שזוהי פשוט עובדה. ההתאמה
בין המציאות לאמונות שלי נובעת מכך שהמציאות מכתובה את
האמונות שלי, ולא להפך.“

כבש נוסף עובר, גורם לי להטיל עוד חלוק נחל.

”אה ! עכשיו אנחנו מגיעים לשורש העניין,“ אומר מארק. ”מה זה

הדבר הזה שנקרא 'מציאות'? אני מבין מה זה אומר שהשערה היא אלגנטית, או ניתנת להפרכה, או תואמת את הראיות. אבל נשמע לי שלתייג אמונה כ'אמיתית' או 'מציאותית' או 'ממשית' זה בסך הכל ההבדל בין לומר שאתה מאמין במשהו, לבין לומר שאתה ממש ממש מאמין במשהו."

אני חושב לרגע. "ובכן..." אני אומר לאט. "בכנות, אני בעצמי לא בטוח מאיפה מגיע כל העסק הזה של 'מציאות'. אני לא יכול ליצור מציאות משלי במעבדה, אז אני כנראה עדיין לא מבין את זה עד הסוף. אבל מדי פעם אני בטוח שמהו הולך לקרות, ואז קורה משהו אחר. אני צריך שם ל'מה-שזה-לא-יהיה' שקובע את תוצאות הניסויים שלי, אז אני קורא לו 'מציאות'. ה'מציאות' הזו נפרדת איכשהו אפילו מההשערות הטובות ביותר שלי. גם כשיש לי השערה פשוטה, שכל הראיות שאני מכיר תומכות בה, קורה שאני עדיין מופתע. אז אני זקוק לשמות שונים עבור הדברים שקובעים את התחזיות שלי ועבור הדבר שקובע את תוצאות הניסויים שלי. אני קורא לדברים הראשונים 'אמונות', ולדבר השני 'מציאות'."

מארק נוחר בבוז. "אני לא יודע למה אני בכלל טורח להקשיב להבלים הללו. כל מה שתגיד על ה'מציאות' הזו, הוא בסך הכל עוד אמונה. אפילו האמונה שלך לגבי כך שהמציאות קודמת לאמונותיך היא אמונה. מכאן נובע, בהכרח, שהמציאות אינה קיימת. רק אמונות קיימות."

"רק רגע," אומר אוטרי, "תוכל לחזור על החלק האחרון? איבדתי אותך בקפיצה הולוגית ההיא שם באמצע."
 "לא משנה מה תגיד על המציאות, זו בסך הכל עוד אמונה," מסביר מארק. "מכאן נובע בהכרח בלתי נמנע שאין מציאות, רק אמונות."
 "אני מבין," אני אומר. "באותו אופן שלא משנה מה אתה אוכל, עליך לאכול את זה עם הפה שלך. מכאן נובע שאין אוכל, רק פיות."
 "בדיוק," אומר מארק. "כל מה שאתה אוכל חייב להיות בפה שלך. איך ייתכן אוכל שמתקיים מחוץ לפה שלך? מחשבה מסוג זה היא אבסורד, מה שמוכיח ש'אוכל' הוא מושג לא קוהרנטי. זו הסיבה

שכולנו גוועים ברעב – האוכל אינו קיים.”
 אוטרי משפיל מבט אל בטנו. “אבל אני לא גווע ברעב.”
 “אהא!” צועק מארק בניצחון. “ואיך הבעת את ההתנגדות הזו
 ממש? עם הפה שלך, ידידי! עם הפה שלך! איזו המחשה טובה יותר
 יכולת לבקש לכך שהאוכל אינו קיים?”
 “מה זה הדיבורים האלה על רעב?” תובע קול צרוד ונוקשה
 ממש מאחורינו. אוטרי ואני נשארים רגועים, שכן אנו כבר מורגלים
 בהפתעות מסוג זה. מארק לעומת זאת, מזנק באוויר, מבוהל עד מוות.
 המפקח דארווין מחייך חיוך קפוץ, מרוצה מההפתעה שיצר,
 ומסמן וי קטן בלוח הכתיבה שבידו.
 “זו רק מטאפורה!” אומר מארק במהירות. “אתה לא צריך להסיר
 לי את הפה, או משהו כזה –”
 “לשם מה אתה זקוק לפה אם אין אוכל?” תובע דארווין בכעס.
 “לא משנה. אין לי זמן לשטויות האלה. אני כאן כדי לפקח על
 הכבשים.”
 “העדר משגשג, אדוני,” אני אומר. “שום כבש לא מת מאז ינואר.”
 “מצוין. אני מעניק לך 0.12 יחידות התאמה. עכשיו, מה האדם
 הזה עושה כאן? האם הוא חלק הכרחי מהפעולות?”
 “למיטב הבנתי, הוא יביא יותר תועלת למין האנושי אם ייתלה על
 כדור פורח בתור משקולת,” אני אומר.
 “אאוץ,” אומר אוטרי בשלווה.
 “לא אכפת לי מהמין האנושי. תן לו לדבר בשם עצמו.”
 מארק מזדקף ביהירות. “רועה הצאן הפשוט הזה,” הוא אומר,
 מצביע לעברי, “טען שיש דבר כזה ‘מציאות’. זה פוגע בי, משום שאני
 יודע בוודאות עמוקה ומוחלטת שאין אמת. המושג ‘אמת’ אינו אלא
 תחבולה של אנשים לכפות את אמונותיהם על אחרים. לכל תרבות יש
 ‘אמת’ משלה, וה‘אמת’ של אף תרבות לא עליונה על זו של אחרת.
 הדברים אשר אמרתי תקפים בכל עת ובכל מקום, ואני עומד על כך
 שתסכים איתי.”
 “רגע אחד,” אומר אוטרי. “אם שום דבר אינו אמת, למה שאאמין

לך כשאתה אומר ששום דבר אינו אמת?"

"לא אמרתי ששום דבר אינו אמת –" אומר מארק.

"כן, אמרת, "קוטע אותו אוטרי, "שמעתי אותך."

"— אמרתי ש'אמת' היא תירוץ המשמש תרבויות מסוימות לאכופ את אמונותיהן על אחרים. אז כשאתה אומר שמהו הוא 'אמיתי', כוונתך אינה אלא שיהיה זה מועיל לקבוצה החברתית שלך שיאמינו בזה."

"ומה שאמרת עכשיו, "אני שואל, "האם זה אמת?"

"אמת לאמיתה, בהחלט!" מכריז מארק בתוקף. "אנשים יוצרים את המציאות של עצמם."

"רגע, "אומר אוטרי, נשמע מבולבל שוב, "לטעון שאנשים יוצרים מציאות משלהם זה דבר שונה לחלוטין מבחינה לוגית מלטעון שאין אמת, מצב עניינים שאני אפילו לא יכול לדמיין באופן קוהרנטי, אולי בגלל שעדיין לא הסברת איך בדיוק זה אמור לעבוד —"

"הנה אתה שוב מתחיל, "אומר מארק בייאוש, "מנסה ליישם את המושגים המערביים שלך לגבי לוגיקה, רציונליות, היגיון, קוהרנטיות ועקביות פנימית."

"נהדר, "ממלמל אוטרי, "עכשיו אני צריך להוסיף לרשימה קטגוריה שלישית, כדי לעקוב אחר הטענה הנפרדת והמובחנת לחלוטין הזו —"

"זו לא טענה נפרדת, "אומר מארק. "תראה, אתה נוקט בגישה הלא נכונה בכך שאתה מתייחס לאמירות שלי כאל השערות, וגוזר את השלכותיהן בזהירות. אתה צריך לחשוב עליהן כעל תירוצים כלליים לחלוטין, שאני מיישם כשמישהו אומר משהו שאינו לרוחי. זה לא ממש מודל של הדרך בה היקום עובד, כמו שזה כרטיס חינם של 'צא מהכלא'. החלק החשוב הוא ליישם את התירוץ באופן סלקטיבי. כשאני אומר שאין דבר כזה אמת, זה חל רק על הטענה שלך שדלי הקסם עובד בין אם אני מאמין בו ובין אם לא. זה לא חל על הטענה שלי שאין דבר כזה אמת."

"אממ... למה לא?" שואל אוטרי.

מארק נאנח בסבלנות. "אוטרי, אתה חושב שאתה האדם הראשון שחשב על השאלה הזו? לשאול אותנו איך האמונות שלנו יכולות להיות בעלות משמעות אם כל האמונות חסרות משמעות? זו אותה הטענה שמשמיעים סטודנטים רבים כשהם פוגשים לראשונה את הפילוסופיה הזו, שיש לה, לידיעתך, חסידים רבים וספרות ענפה."

"אז מה התשובה?" אומר אוטרי.

"קראנו לזה 'בעיית הרפלקסיביות', מסביר מארק.

"אבל מה התשובה?" מתעקש אוטרי.

מארק מחייך בהתנשאות. "תאמין לי, אוטרי, אתה לא האדם הראשון שחשב על שאלה פשוטה כל כך. אין טעם להציג אותה בפנינו כהפרכה ניצחת."

"אבל מה באמת התשובה?"

"כעת, ברצוני לעבור לנושא של איך לוגיקה הורגת גורים חמודים של כלביים—

"אתה מבזבז זמן, נוזף המפקח דארווין.

"שלא לדבר על כך שאתה מאבד את המעקב אחר הכבשים," אני אומר, ומטיל עוד חלוק נחל. המפקח דארווין מביט בשני המתורכחים, ונראה שאיש מהם לא מוכן לוותר על עמדתו. "תקשיבו," אומר דארווין, בטון אדיב יותר כעת, "יש לי רעיון פשוט ליישוב המחלוקת שלכם. אתה אומר, "אומר דארווין ומצביע על מארק, "שהאמונות של אנשים משנות את המציאות האישית שלהם. ואתה מאמין בלהט, "הוא מסובב את אצבעו ומצביע על אוטרי, "שהאמונות של מארק לא יכולות לשנות את המציאות. אז בואו ניתן למארק להאמין ממש חזק שהוא יכול לעוף, ואז לצעוד מקצה הצוק. מארק יראה את עצמו עף משם כמו ציפור, ואוטרי יראה אותו צונח מטה ונמרח על הקרקע, ושניכם תהיו מרוצים."

כולנו עוצרים, שוקלים זאת.

"זה נשמע סביר..." אומר מארק לבסוף.

"יש צוק ממש שם," מציין המפקח דארווין.

על פניו של אוטרי נסוכה הבעה של ריכוז מאומץ. לבסוף הוא

צועק: "חכה! אם זה היה נכון, כולנו היינו עוזבים מזמן ליקומים פרטיים משלנו, ובמקרה כזה האנשים האחרים כאן הם רק פרי דמיונכם – אין טעם לנסות להוכיח לנו שום דבר –"

צרח ארוכה ודועכת נשמעת מהצוק הסמוך, ולאחריה חבטה עמומה ובודדת. המפקח דארווין הופך דף בלוח הכתיבה שלו אל העמוד המציג את מאגר הגנים הנוכחי ורושם בעיפרון תדירות נמוכה במקצת עבור האללים של מארק.

אוטרי נראה מעט חלוש. "זה באמת היה הכרחי?"

"הכרחי?" אומר המפקח דארווין, מבולבל. "זה פשוט קרה... אני לא כל כך מבין את השאלה שלך."

אוטרי ואני פונים בחזרה לדלי שלנו. הגיע הזמן להכניס את הכבשים. עדיף שלא לשכוח את החלק הזה. אחרת, מה הטעם?