רציונליות

מ-א ועד ת



ספר II

איך לשנות את דעתך
הלכה למעשה

אליעזר יודקובסקי

רציונליות: מ-א ועד ת | אליעזר יודקובסקי

ספר II

רציונליות

מ-א ועד ת

מאנגלית: דביר שדה ואליהו פרבר

ספר II

איך לשנות את דעתך
הלכה למעשה

אליעזר יודקובסקי

Eliezer Yudkowsky
Rationality: From AI to Zombies
Book II: How to Actually Change Your Mind

אליעזר יודקובסקי
רציונליות: מ-א׳ ועד ת׳
ספר II: איך לשנות את דעתך הלכה למעשה

תרגום: דביר שדה ואליהו פרבר
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

רציונליות: מבוא 10

תירוצים נוחים מדי

46. מתי עליך להיות עניו 24
47. החלופה השלישית 29
48. הגרלות לוטו: בזבוז של תקווה 33
49. לוטו משודרג 35
50. אבל עדיין יש סיכוי, נכון? 38
51. כשל "הכל אפור" 41
52. סמכות מוחלטת 46
53. איך לשכנע אותי ש- 2+2=3 53
54. ודאות אינסופית 56
55. 0 ו-1 אינם הסתברויות 61
56. הרציונליות שלך היא העסק שלי 66

רציונליות ופוליטיקה

57. פוליטיקה היא אויב התבונה 70
58. דיונים פוליטיים לא אמורים להרגיש חד־צדדיים 73
59. מאזני הצדק, פנקס הרציונליות 77
60. הטיית ההתאמה 80
61. האם האויבים שלכם רעים מטבעם? 84

62. טיפשות הפוכה איננה חוכמה 88
63. הטיעון מייתר את הסמכות 92
64. לחבק את השאלה 98
65. רציונליות והשפה האנגלית 100
66. רוע אנושי ומחשבה מעורפלת 105

נגד רציונליזציה

67. ידע על הטיות יכול להזיק 110
68. עדכנו את עצמכם בהדרגה 114
69. טיעון אחד מול צבא שלם 118
70. השורה התחתונה 121
71. ראיה סלקטיבית היא ראיה ראויה? 125
72. רציונליזציה 129
73. טיעון רציונלי 132
74. להימנע מנקודות התורפה האמיתיות של האמונה שלך 135
75. עצירה מגמתית ומשיכת זמן מגמתית 140
76. הצדקות מזויפות 144
77. האם זו ההתנגדות האמיתית שלך? 147
78. אמיתות שזורות, שקרים מדבקים 151
79. על שקרים וקטסטרופות ברבור שחור 156
80. אפיסטמולוגיה של הצד האפל 157

נגד דוּחוֹשֶׁב

81. חַדחוֹשֶׁב 164

82. דוּחוֹשֵב (הבחירה להיות מוטה) 166
83. תקשיב, אני באמת מצליחה להשלות את עצמי 171
84. אמונה בהונאה עצמית 173
85. פרדוקס מור 178
86. האמן שאינך מסוגל להונות את עצמך 181

לראות בעיניים חדשות

87. עוגנים ותיקונים 186
88. הטרמה וזיהום 189
89. האם אנו מאמינים לכל מה שאומרים לנו? 193
90. מחשבות מטמון 196
91. קופסת ה"מחוץ לקופסה" 200
92. מבט מקורי 204
93. יותר מוזר מההיסטוריה 206
94. הכשל הלוגי של הכללה מראיות בדיוניות 208
95. סוד הצמצום 215
96. כיצד להיראות (ולהיות) עמוקים 219
97. אנו משנים את דעתנו פחות מכפי שנדמה לנו 223
98. המתן לפני שתציע פתרונות 225
99. הכשל הגנטי 229

מעגלי מוות

100. היוריסטיקת הרגש 234
101. יכולת הערכה (וקניות זולות לחגים) 238

102. סולמות אינסופיים, פיצויי עתק של חבר מושבעים ועתידנות 243
103. אפקט ההילה 247
104. הטיית גיבורי־העל 251
105. בסך הכל משיח 255
106. מעגלי מוות רגשיים 259
107. התנגדו למעגל המוות המאושר 262
108. על־קריטיות אל־ביקורתית 268
109. קירור באידוי של אמונות קבוצתיות 273
110. כשאיש אינו מעז לקרוא לאיפוק 278
111. ניסוי מערת השודדים 281
112. כל מטרה שואפת להפוך לכת 285
113. שומרי האמת 289
114. שומרי הגנום 294
115. שומרי איין ראנד 296
116. שני קואנים על כת 302
117. ניסוי הקונפורמיות של אָש 305
118. על הבעת חששות 310
119. התנגדות בודדת 313
120. כת ה־"אנחנו לא כת" 318

לשחרר

121. חשיבותה של המילה "אופס" 330
122. ההצעה להפוך לטרחן 333

123. תתייאשו כבר 336
124. השימוש הנכון בספק 338
125. אתה יכול להתמודד עם המציאות 342
126. מדיטציית סקרנות 343
127. אין מי שיפטור אתכם מחוקי הרציונליות 348
128. השאירו נתיב מילוט 351
129. משבר אמונה 356
130. הטקס 366

רציונליות: מבוא

מאת רוב בנסינגר

במה עליי להאמין?

מתברר שלשאלה הזו יש תשובה נכונה.

יש לה תשובה נכונה גם כשאתם אכולי ספקות, ולא רק כשיש בידיכם הוכחה חותכת. תמיד קיימת רמת ודאות נכונה שיש לייחס לטענה מסוימת, גם כשהיא נראית כמו "אמונה אישית" ולא כמו "עובדה" שאומתה על ידי מומחים.

ובכל זאת, אנו מדברים לעיתים קרובות כאילו חוסר ודאות ומחלוקת הופכים אמונות לעניין של טעם בלבד. אנו אומרים "זו רק דעתי" או "זכותך לחשוב כך," כאילו הקביעות של המדע והמתמטיקה מתקיימות במישור אחר וגבוה יותר מאמונות שהן רק "פרטיות" או "סובייקטיביות". אך כפי שכותב רובין הַנסון:[1]

> לעולם אינך זכאי לאמונות משלך. לעולם לא; אינך זכאי אפילו לומר "אני לא יודע". אתה זכאי לרצונות משלך, ולעיתים לבחירות משלך. ייתכן שבחירה מסוימת היא באמת שלך, ואם אתה יכול לבחור את העדפותיך, ייתכן שיש לך הזכות לעשות זאת. אך האמונות שלך אינן עוסקות בך. אמונות עוסקות בעולם. האמונות שלך צריכות להיות ההערכה הטובה ביותר הזמינה לך לגבי העולם כפי שהוא. כל דבר אחר הוא שקר. [...]
>
> נכון שנושאים מסוימים מספקים למומחים מנגנונים חזקים יותר ליישוב מחלוקות. בנושאים אחרים, ההטיות שלנו ומורכבות העולם מקשות על הסקת מסקנות נחרצות. [...]
>
> אך לעולם אל תשכח שלגבי כל שאלה על האופן שבו העולם מתנהל (או צריך להתנהל), ובכל מצב נתון של מידע זמין, תמיד *קיימת* הערכה מיטבית. זכותך היחידה היא להשקיע את מרב המאמצים, בכנות מלאה, כדי למצוא את ההערכה המיטבית הזו. כל דבר אחר הוא שקר.

נניח שגילית שאחד מתוך שישה אנשים מסוימים דלוק עלייך – אולי קיבלת מכתב ממעריץ אלמוני ואת בטוחה שהוא אחד מהשישה הללו – אך אין לך מושג מי מביניהם. בוב, חברך לספסל הלימודים, הוא אחד משמת המועמדים, אך אין לך כל ראיה מיוחדת המצביעה עליו ככותב המכתב, או פוסלת אותו. במקרה כזה, יחס הסיכויים לכך שבוב דלוק עלייך הוא 1:5.

כיוון שיש שש אפשרויות, ניחוש אקראי יביא לכך שתצדקי פעם אחת על כל חמש פעמים שתטעי, בממוצע. זו כוונתנו כשאנו אומרים ש"יחס הסיכויים הוא 1:5." אינך יכולה לומר, "ובכן, אין לי מושג מי דלוק עליי. אולי זה בוב, ואולי לא, אז פשוט אומר שהסיכויים הם חצי־חצי." גם אם את מעדיפה לומר "איני יודעת" או "אולי" ולעצור שם, התשובה היא עדיין 1:5.[2]

נניח גם ששמת לב שאנשים קורצים לך בתדירות גבוהה פי עשרה כשהם דלוקים עלייך. אם בוב קורץ לך, זוהי ראיה חדשה. במקרה כזה, תהיה זו טעות להישאר סקפטית לגבי האפשרות שבוב הוא המעריץ האלמוני שלך. יחס הנראוּת של 10:1 לטובת הטענה "אדם אקראי שקורץ לי אכן דלוק עליי" גובר על יחס הסיכויים של 1:5 נגד הטענה "בוב דלוק עליי".

תהיה זו גם טעות לומר, "הראיה הזו כל כך חזקה, שזה הימור בטוח שהוא זה שדלוק עליי ! מעכשיו פשוט אניח שבוב בעניין שלי." ביטחון מופרז גרוע בדיוק כמו חוסר ביטחון.

למעשה, קיימת רק תשובה אחת אפשרית לשאלה הזו שהיא עקבית מבחינה מתמטית. כדי לעדכן את אמונתנו מיחס הסיכויים הפריורי של 1:5, על בסיס יחס הנראוּת של הראיה (10:1), עלינו להכפיל את הצדדים השמאליים זה בזה ואת הצדדים הימניים זה בזה. כך נקבל יחס סיכויים פוסטריורי של 10:5, או יחס של 2:1 לטובת האפשרות ש"בוב דלוק עליי."* בהינתן ההנחות שלנו והראיות הזמינות, הניחוש

---

* שני מושגים מרכזיים בסטטיסטיקה בייסיאנית הינם: *הסתברות פריורית* (prior probability) – אמונה סובייקטיבית לגבי ההתפלגות של פרמטר מסוים, ללא נתונים, טרם ביצוע התצפית. *הסתברות פוסטריורית* (posterior probability) –

שבוב דלוק עלייך יתברר כנכון פעמיים על כל פעם אחת שיתברר כשגוי. ובניסוח שקול: ההסתברות שהוא דלוק עלייך היא 2/3. כל רמת ביטחון אחרת תהיה בלתי עקבית.

התרבות שלנו טרם הפנימה את לקחיה של תורת ההסתברות – שהתשובה הנכונה לשאלות כמו "עד כמה אני יכולה להיות בטוחה שבוב דלוק עליי?" כפופה לאילוצים לוגיים נוקשים, ממש כמו התשובה הנכונה לשאלה במבחן באלגברה או בספר לימוד בגאולוגיה. צורת המחשבה המקובלת עדיין נמצאת בפיגור ביחס לתובנה שלשאלה "באילו אמונות עליי להחזיק?" יש תשובה נכונה באופן אובייקטיבי, בין שהשאלה היא "האם חברי לכיתה דלוק עליי?" ובין שהיא "האם יש לי נשמה נצחית?"

אכן קיימת דרך נכונה לעדכן את אמונותיכם, והיא *מדויקת* להפליא.

איך לא לשנות את דעתך, הלכה למעשה

עם זאת, עדכון אמונותינו בדרך שמתקרבת ולו במעט לאידיאל הזה הוא כבר עניין מורכב בהרבה.

בכרך הראשון של *רציונליות: מ-א ועד ת*, דנו בערכן של אמונות "תקינות".* אין שום פסול עקרוני בהבעת תמיכה בדבר־מה שחשוב לכם – כגון קבוצה שאתם מזדהים עמה, או חוויה רוחנית שמעוררת בכם התעלות. אולם כאשר אנו מבלבלים בין קריאות עידוד לבין אמונות לגבי המציאות, אותן קריאות עידוד – אם יפורשו בטעות כאמונות – עלולות ליצור שריון שיחסום ראיות שבכוחן לחדור לאידיאולוגיה שלנו ולערער אותה.

אפילו אמונות המעניקות לכאורה הסבר אלגנטי לתצפיות שלנו

ההתפלגות שהתקבלה מעדכון של ההסתברות הפריורית על סמך מידע שהתקבל מהתצפית.

* ראו תחילת פרק 18.

אינן חסינות מפני הבעיה הזו. קל מדי ללכת שולל אחר ביטוי הנשמע מדעי (או סמכותי בדרך אחרת) ולהסיק שהוא "הסביר" משהו, גם כאשר אינו משפיע כלל על ההסתברויות שאנו מייחסים במובלע לחוויות עתידיות אפשריות.

גרוע מכך, אמונות רגילות – אמונות שבאופן עקרוני *ניתנות* להפרכה, ואכן *משפיעות* על מה שאנו מצפים לראות – עדיין עלולות להשתרש בתודעתנו כשהן מבוצרות על ידי מערך שלם של אשליות והטיות.

בשנת 1951, משחק פוטבול בין דארטמות׳ לפרינסטון הפך לאלים במיוחד. הפסיכולוגים הַאסטוֹרף וקַנטריל שאלו סטודנטים משני המוסדות מי התחיל באלימות. כמעט כולם הסכימו שפרינסטון לא התחילה בה, אך 86% מהסטודנטים מפרינסטון האמינו שדארטמות׳ התחילה, בעוד שרק 36% מהסטודנטים מדארטמות׳ האשימו את קבוצתם. (רוב הסטודנטים מדארטמות׳ סברו כי "שתי הקבוצות התחילו".)

אין סיבה להניח שמדובר היה בקריאת עידוד בלבד, להבדיל מאמונה אמיתית. סביר להניח שאמונותיהם השונות הובילו את הסטודנטים לתחזיות שונות בנוגע להתנהגות השחקנים במשחקים עתידיים. ואף על פי כן, בדרך כלשהי, האמונות העובדתיות והרגילות לחלוטין של הסטודנטים בדארטמות׳ היו שונות בתכלית מהאמונות העובדתיות והרגילות לחלוטין של הסטודנטים בפרינסטון.

האם ניתן לייחס זאת למקורות המידע השונים שהיו נגישים לסטודנטים בשני המוסדות? בפני עצמה, הטיה במקורות המידע שעליהם נסמכות קבוצות שונות היא בעיה חמורה למדי.

ואולם, במקרה זה היה מעורב דבר־מה עמוק יותר. כאשר הוקרן בפניהם מאוחר יותר סרטון של המשחק והם התבקשו למנות את העבירות שבוצעו, הסטודנטים מדארטמות׳ טענו שראו בממוצע 4.3 עבירות מצד קבוצתם (והגדירו את מחציתן כ"קלות"), בעוד שהסטודנטים מפרינסטון דיווחו שראו בממוצע 9.8 עבירות מצד דארטמות׳ (ורק שליש מהן הוגדרו כ"קלות").

הַניחו בצד את הניסיונות לגרום למחנות יריבים להסכים על סוגיות סבוכות בפוליטיקה או בפילוסופיה של המוסר: סטודנטים מקבוצות שונות לא הצליחו להסכים אפילו על העובדות *שראו במו עיניהם*.[3]

כאשר דבר־מה יקר לליבנו מאוים – תהא זו תפיסת עולמנו, הקבוצה שאליה אנו משתייכים, מעמדנו החברתי או כל דבר אחר – המחשבות והתפיסות שלנו מתגייסות להגנתו.[4,5] ישנם כיום פסיכולוגים המרחיקים לכת ומשערים כי יכולתנו לנסח הנמקות מפורשות למסקנותינו התפתחה *במיוחד* כדי לסייע לנו לנצח בוויכוחים.[6]

אחת מתובנות היסוד של הפסיכולוגיה במאה העשרים, המשותפת לכולם – מפרוידיאנים ועד פסיכולוגים קוגניטיביים בני ימינו – היא כי תהליכים לא־מודעים ומתוחכמים הם אלה שלרוב מכתיבים את התנהגותנו, וכי הסיפורים שאנו מספרים לעצמנו על המניעים והסיבות שלנו מוטים ומומצאים הרבה יותר מכפי שנדמה לנו.

למעשה, לעיתים קרובות איננו מודעים כלל לכך שאנו מספרים לעצמנו סיפור. כאשר נדמה לנו שאנו "תופסים ישירות" אמיתות על עצמנו באמצעות התבוננות פנימית, מתברר לא פעם כי תפיסה זו נשענת על מודלים סיבתיים סמויים ורופפים.[7,8] כשאנו מנסים לנמק את אמונותינו, אנו עשויים להציג היגיון רעוע שאינו רלוונטי כלל לדרך שבה התגבשה האמונה מלכתחילה.[9] במקום לשפוט את ההסברים שלנו על פי כוח הניבוי שלהם, אנו טווים סיפורים כדי להעניק משמעות למה שאנו חושבים שאנו כבר יודעים.

כיצד ניתן להשתפר? כיצד נוכל להגיע לתפיסת עולם מציאותית, כאשר מוחנו כה נוטה לרציונליזציה? כיצד נוכל לגבש הבנה מציאותית של תודעתנו שלנו, כאשר מחשבותינו *על החשיבה עצמה* גם הן חשודות? כיצד נוכל להיות פחות מוטים, כאשר המאמצים שלנו לנטרל הטיות עלולים להיות נגועים בהטיות בעצמם?

מהו המקום הכי פחות רעוע שעליו נוכל להישען?

המתמטיקה של הרציונליות

בפתחה של המאה ה-20, ניסוחן של אקסיומות פשוטות העומדות ביסוד האריתמטיקה (לדוגמה, בתורת הקבוצות) העניק למתמטיקאים אמת מידה ברורה יותר לבחינת תקפותן של מסקנותיהם. אם אדם או מחשבון קובעים כי "2+2=4", אנו יכולים כעת לעשות יותר מאשר לומר "זה נראה נכון אינטואיטיבית." אנו יכולים להסביר *מדוע* הדבר נכון, ולהוכיח כי אמיתות זו קשורה באופן שיטתי לנכונותם של חלקים אחרים באריתמטיקה.

אך מתמטיקה ולוגיקה מאפשרות לנו לעשות הרבה יותר מאשר רק למדל מחשבונים פשוטים. הן מאפשרות לנו ליצור מודלים של מערכות פיזיקליות מרתקות בהרבה. אנו יכולים לנסח *אמונה רציונלית באופן כללי* ופורמלי, באמצעות תורת ההסתברות, כדי לזהות מאפיינים המשותפים לכל תהליך של הסקה תקפה. אנו יכולים אף לנסח באופן פורמלי *התנהגות רציונלית כעיקרון כללי*, על בסיס תורת ההחלטות.

תורת ההסתברות מגדירה כיצד היינו מסיקים מסקנות באופן אידיאלי לנוכח אי־ודאות, אילו רק עמדו לרשותנו הזמן, כוח החישוב והשליטה העצמית הדרושים. בהינתן ידע מוקדם (*הסתברות פריורית*) וראיות חדשות, תורת ההסתברות מגדירה באופן חד־משמעי את האמונות החדשות המיטביות (*הסתברות פוסטריורית*) שביכולתנו לאמץ. בדומה לכך, תורת ההחלטות מגדירה איזו פעולה עלינו לנקוט על סמך אמונותינו. עבור כל מערך עקבי של אמונות והעדפות שעשויות להיות לי בנוגע לבוב, קיימת תשובה במסגרת תורת ההחלטות המורה כיצד עליי לפעול כדי לממש את העדפותיי.

בני אדם אינם מכונות חשיבה מושלמות או מקבלי החלטות ללא דופי, בדיוק כפי שאינם מחשבונים מושלמים. מוחנו הוא אוסף של פתרונות מאולתרים שקובצו יחדיו על ידי הברירה הטבעית. אפילו במיטבנו, איננו מחשבים את התשובה המדויקת והנכונה לשאלות "מה עליי לחשוב?" ו"מה עליי לעשות?" חסרים לנו הזמן וכוח

החישוב, ולאבולוציה חסרו המומחיות ההנדסית וראיית הנולד הנדרשות כדי לתקן את כל התקלות במערכת שלנו.

למעשה, אפילו מכונת חשיבה מושלמת ויעילה להפליא בעולם האמיתי עדיין תיאלץ להסתמך על היוריסטיקות וקירובים. האלגוריתמים הטובים ביותר שניתן להריץ בפועל לשם עדכון אמונות אינם מגיעים לרמת העקביות של תורת ההסתברות.

עם זאת, אף שאנו יודעים כי עקביות *מושלמת* אינה בהישג ידינו, אנו בהחלט יכולים להשתפר. הידיעה שקיים תקן אידיאלי שנוכל להשוות עצמנו אליו – מה שחוקרים מכנים "רציונליות בייסיאנית" – יכולה להוות עבורנו מצפן בעודנו שואפים לשפר את מחשבותינו ומעשינו. אף שלעולם לא נהיה בייסיאנים מושלמים, המתמטיקה של הרציונליות עשויה לעזור לנו להבין *מדוע* תשובה מסוימת נכונה, ולסייע לנו לשים את האצבע על המקום שבו שגינו.

תארו לעצמכם שאתם מנסים ללמוד מתמטיקה באמצעות שינון בלבד. אולי יאמרו לכם ש-"13=3+10", ש-"139=108+31", וכן הלאה, אך לא תפיקו מכך תועלת רבה אלא אם כן תבינו את החוקיות העומדת מאחורי הסמלים. באופן דומה, קשה הרבה יותר לחפש דרכים להשתפר בחשיבה רציונלית כשאין לכם מודל כללי שיעזור להבין אם השיטה באמת עובדת. מטרת הספר הזה היא לסייע לכם לבנות מודלים כאלה.

רציונליות יישומית

במאמר שעסק בהבדל בין "רציונליסטים" שוחרי רציונליות לבין "רציונליסטים" המתנגדים לגישה האמפירית, כתב סקוט אלכסנדר:[10]

> כמובן, יש תועלת בצבירת ראיות רבות ככל האפשר, כפי שיש תועלת בצבירת הון רב ככל האפשר. אך באותה מידה, יש תועלת ביכולת להשתמש בכמות מוגבלת של ראיות בתבונה, בדיוק כפי שיש תועלת ביכולת להשתמש בכמות מוגבלת של כסף בתבונה.

טכניקות רציונליות מסייעות לנו להפיק יותר מהראיות שבידינו, במקרים שבהם הראיות אינן חד־משמעיות, או כאשר ההטיות והנאמנויות שלנו מעוותות את האופן שבו אנו מפרשים אותן. הדבר תקף לחיינו האישיים, כמו בסיפורו של בוב, למחלוקות בין מחנות פוליטיים (ובין אוהדי ספורט), ולסוגיות טכנולוגיות ופילוסופיות, כמו בוויכוחים על טרנס־הומניזם – העמדה שלפיה עלינו להשתמש בטכנולוגיה כדי לשדרג מן היסוד את הקיום האנושי. מתוך הכרה בכך שאותם כללים מתמטיים חלים על כל אחד מהתחומים הללו – ושאותן הטיות קוגניטיביות משפיעות עלינו במרבית המקרים – כרך זה, *איך לשנות את דעתך הלכה למעשה*, נשען על מגוון רחב של דוגמאות ובעיות.

רצף המאמרים הראשון בכרך זה, "תירוצים נוחים מדי", מתמקד בשאלות שהתשובה ההסתברותית עליהן היא הברורה ביותר האפשרית. לעיתים קרובות התשובה הבייסיאנית האופטימלית אינה ניתנת לחישוב בפועל, אך שגיאות כמו הטיית האישוש עלולות להשתרש גם במקרים שבהם הראיות הזמינות הן חד משמעיות, ויש לנו שפע של זמן לשקול אותן.

משם אנו עוברים למים עכורים יותר, עם אסופת המאמרים "רציונליות ופוליטיקה". פוליטיקה ארצית, כפי שהיא נדונה בפי פרשני טלוויזיה, ידועה לשמצה בשיח הזועם והעקר המאפיין אותה. על פניו, יש בכך משהו מפתיע. מדוע אנו לוקחים מחלוקות פוליטיות באופן אישי כל כך, אפילו כאשר המנגנונים וההשפעות של הפוליטיקה רחוקים מאיתנו כל כך בזמן או במרחב? יתרה מזאת, מדוע איננו נעשים זהירים ויסודיים *יותר* בבחינת הראיות דווקא כאשר אנו עוסקים בסוגיות כה כבדות משקל?

המשחק בין דארטמות' לפרינסטון רומז על התשובה. חלק ניכר מתהליך ההסקה שלנו הוא למעשה רציונליזציה – טוויית סיפורים שגורמים לאמונותינו הנוכחיות להיראות עקביות ומוצדקות יותר, מבלי לשפר בהכרח את מידת הדיוק שלהן. הרצף "נגד רציונליזציה" עוסק בבעיה זו, ואחריו הרצפים "נגד דוחושב" (על הונאה עצמית)

ו"לראות בעיניים חדשות" (על האתגר שבזיהוי ראיות שאינן תואמות את הציפיות וההנחות שלנו).

התקדמות מבחינה רציונלית משמעה מפגש עם רעיונות רבים חדשים, מעניינים ועוצמתיים. במקרים רבים, משמעות הדבר היא גם רכישת חברים שעימם ניתן להחליף דעות, ומציאת קהילות המעודדות אתכם להשתפר. "מעגלי מוות" דן בכמה סכנות מרכזיות שעלולות לפקוד קבוצות המאוחדות סביב תחומי עניין משותפים ורעיונות מדהימים ומסנוורים – סכנות שיש להתגבר עליהן אם ברצוננו להפיק את מלוא התועלת מקהילות רציונליסטיות. הכרך *איך לשנות את דעתך הלכה למעשה* נחתם ברצף "לשחרר".

הנטייה הטבעית שלנו *אינה* לשנות את דעתנו כפי שבייסיאני אמור לעשות. לגרום לסטודנטים מדארטמות' ומפרינסטון *להבחין במה שהם רואים באמת*, לא יהיה פשוט כמו לדקלם באוזניהם את האקסיומות של תורת ההסתברות. כפי שכתב לוק מולהאוזר:[11]

> אינך הומונקולוס* בייסיאני שההסקה שלו "התקלקלה" בשל הטיות קוגניטיביות.
> אתה *הינך* הטיות קוגניטיביות.

הטיית האישוש, הטיית הסטטוס קוו, הטיית הייחוס ודומיהן אינן סרח עודף המודבק לתהליך ההסקה שלנו. הן עצם מהותו.

אין פירוש הדבר שנטרול הטיות הוא משימה בלתי אפשרית. מנגד, איננו מחשבונים מושלמים המסתתרים מאחורי השגיאות האריתמטיות שלנו. רבות מהמגבלות המתמטיות שלנו נובעות מעובדות יסודיות מאוד על אופן פעולתו של המוח האנושי. ובכל זאת, אנו יכולים לאמן את היכולות המתמטיות שלנו. אנו יכולים

---

* *הומונקולוס* (מלטינית: "איש קטן") הוא כשל הסברי במדעי הקוגניציה, המנסה לבאר יכולת מנטלית באמצעות סוכן פנימי הניחן באותה היכולת עצמה – למשל, הסברת יכולת הראיה באמצעות ההנחה שהמוח "מביט" בתמונות שנקלטות ברשתית העין. הסבר כזה רק דוחה את השאלה במקום לענות עליה, שכן כעת יש להסביר כיצד המוח עצמו רואה.

ללמוד מתי לבטוח ומתי להטיל ספק באינטואיציות המתמטיות שלנו, לחלוק את הידע שלנו ולעזור זה לזה. אנו יכולים לעצב את הסביבה שלנו כדי להקל על עצמנו, ולבנות כלים שיחסכו מאיתנו חלק ניכר מהעבודה.

ההטיות שלנו הן חלק מאיתנו. אך קיים בתוכנו גם צל של חשיבה בייסיאנית, מנגנון בלתי־מושלם שאכן מסוגל לקרב אותנו אל האמת. אין שם הומונקולוס – אך בכל זאת, יש שם מעט אמת. וייתכן שדי בכך כדי להתחיל.

1. Robin Hanson, "You Are Never Entitled to Your Opinion," *Overcoming Bias* (blog) (2006), http://www.overcomingbias.com/2006/12/you_are_never_e.html.

2. הדבר נובע מן ההנחה שקיימות שש אפשרויות, ושאין לך כל סיבה להעדיף אחת מהן על פני האחרות. אנו גם מניחים, באופן בלתי מציאותי, שביכולתך לדעת בוודאות כי המעריץ הוא אכן אחד מאותם שישה אנשים, ושאינך מתעלמת מאפשרויות נוספות. (מה אם יותר מאחד מאותם שישה אנשים דלוק עלייך?)

3. Albert Hastorf and Hadley Cantril, "They Saw a Game: A Case Study," *Journal of Abnormal and Social Psychology* 49 (1954): 129–134, http://www2.psych.ubc.ca/~schaller/Psyc590Readings/ Hastorf1954.pdf.

4. Pronin, "How We See Ourselves and How We See Others."

5. Robert P. Vallone, Lee Ross, and Mark R. Lepper, "The Hostile Media Phenomenon: Biased Perception and Perceptions of Media Bias in Coverage of the Beirut Massacre," *Journal of Personality and Social Psychology* 49 (1985): 577–585, http://ssc.wisc.edu/~jpiliavi/965/hwang.pdf.

6. Hugo Mercier and Dan Sperber, "Why Do Humans Reason? Arguments for an Argumentative Theory," *Behavioral and Brain Sciences* 34 (2011): 57–74, https://hal.archives-ouvertes.fr/file/index/docid/904097/filename/MercierSperberWhydohumansreason.pdf.

7. Richard E. Nisbett and Timothy D. Wilson, "Telling More than We Can Know: Verbal Reports on Mental Processes," *Psychological Review* 84 (1977): 231–259, http://people.virginia.edu/~tdw/nisbett&wilson.pdf.

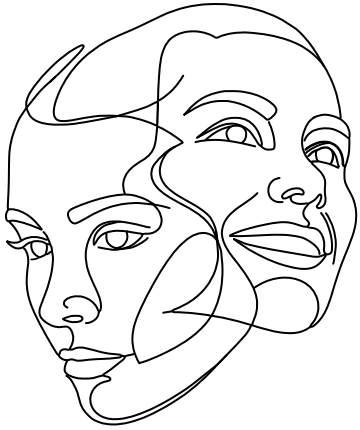
8. Eric Schwitzgebel, *Perplexities of Consciousness* (MIT Press, 2011).

9. Jonathan Haidt, "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* 108, no. 4 (2001): 814–834, doi:10.1037/0033-295X.108.4.814.

10. Scott Alexander, "Why I Am Not Rene Descartes," *Slate Star Codex* (blog) (2014), http://slatestarcodex.com/2014/11/27/why-i-am-not-rene-descartes/.

11. Luke Muehlhauser, "The Power of Agency," *Less Wrong* (blog) (2011), https://www.greaterwrong.com/lw/5i8/the_power_of_agency/.

איך לשנות את דעתך
הלכה למעשה



חלק ה'

תירוצים נוחים מדי

רצף המאמרים הראשון בכרך זה, "תירוצים נוחים מדי", מתמקד בשאלות שהתשובה ההסתברותית עליהן היא הברורה ביותר האפשרית. לעיתים קרובות התשובה הבייסיאנית האופטימלית אינה ניתנת לחישוב בפועל, אך שגיאות כמו הטיית האישוש עלולות להשתרש גם במקרים שבהם הראיות הזמינות הן חד־משמעיות, ויש לנו שפע של זמן לשקול אותן.

— רוב בנסינגר,
רציונליות: מבוא

46

מתי עליך להיות עניו

מקובל על הכל כי מדע ראוי לשמו דורש סוג מסוים של ענווה. אולם יש פחות תמימות דעים בנוגע לשאלה *איזה סוג* של ענווה.

קחו למשל בריאתן האומר: "אבל מי באמת יכול לדעת אם האבולוציה נכונה? הרי זו *רק תיאוריה*. עליכם להיות ענווים ופתוחים יותר לדעות אחרות." האם זו ענווה? הבריאתן נוקט בחוסר־ביטחון סלקטיבי מאוד: הוא מסרב להביא בחשבון מסה אדירה של ראיות לטובת מסקנה שאינה נוחה לו. הייתי אומר שבין אם תקראו לזה "ענווה" ובין אם לאו, זהו הצעד הלא־נכון בריקוד.

מה לגבי המהנדס שמתכנן בענווה מנגנוני אל־כשל במכונה, למרות שהוא בטוח לחלוטין שהמכונה לא תיכשל? זו נראית לי כמו ענווה מסוג טוב. מבחינה היסטורית, אין זה חסר תקדים שמהנדס יהיה משוכנע לגמרי שמכונה חדשה לא תתקלקל, ואז היא מתקלקלת בכל זאת.

ומה לגבי הסטודנט שבודק בענווה את תשובותיו במבחן במתמטיקה פעם נוספת? גם זאת הייתי מסווג כענווה טובה.

ומה לגבי סטודנט שאומר: "ובכן, לא משנה כמה פעמים אבדוק, לעולם לא אוכל להיות *בטוח* שתשובותיי נכונות," ולכן אינו בודק אפילו פעם אחת? גם אם בחירה זו נובעת מרגש דומה לרגש שחש הסטודנט הקודם, היא פחות נבונה.

אם תציעו לסטודנט ללמוד קשה יותר, הוא עשוי לענות:

״לא, עזוב אותי. אני לא אחד מהגאונים של הכיתה כמוך. למישהו ברמה שלי אין סיכוי להצליח.״ זוהי צניעות חברתית, לא ענווה. זהו עניין של ויסות סטטוס בתוך השבט, לא של תהליך מדעי. אם תבקשו ממישהו ״*להיות עניו יותר*,״ באופן טבעי הוא יקשר את המילים לצניעות חברתית – מושג אינטואיטיבי, יומיומי ורלוונטי מבחינה אבולוציונית. ענווה מדעית היא המצאה מאוחרת ונדירה יותר, והיא אינה חברתית במהותה. ענווה מדעית היא משהו שהייתם מתרגלים גם אם הייתם לבדכם בתוך חליפת חלל, במרחק שנות אור מכדור הארץ, כשאיש אינו צופה בכם. או אפילו אם הייתם מקבלים ערובה מוחלטת לכך שאיש לא ימתח עליכם ביקורת לעולם, לא משנה מה תגידו או מה תחשבו על עצמכם. אם הייתם נבונים, עדיין הייתם בודקים שוב את החישובים שלכם.

הסטודנט אומר: ״אבל ראיתי סטודנטים אחרים שבדקו את תשובותיהם שוב ושוב ובסוף הן בכל זאת היו שגויות. ומה אם, בגלל בעיית האינדוקציה,* הפעם $2 + 2 = 5$? לא משנה מה אעשה, לא אהיה בטוח בעצמי.״ זה נשמע עמוק מאוד, וצנוע מאוד. אך אין זה מקרה שהסטודנט רוצה להגיש כבר את המבחן וללכת הביתה לשחק במחשב.

סופה של תקופה בפיזיקה לא תמיד מכריז על עצמו בתרועות חצוצרה. לרוב הוא מתחיל במה שנראה כפגם מזערי... אך מכיוון שפיזיקאים מחזיקים ברעיון היהיר שהמודלים שלהם אמורים לעבוד *כל* הזמן, ולא רק *רוב* הזמן, הם עוקבים אחר פגמים קטנים. בדרך כלל, הפגם הזעיר נעלם תחת בחינה מדוקדקת. לעיתים רחוקות, הפגם מתרחב עד לרגע שבו הוא מנפץ את התיאוריה כולה. לכן נכתב: ״אם לא תשאפו לשלמות, תיעצרו עוד לפני שתצעדו את הצעד הראשון.״

אך חשבו על עזות המצח החברתית שבניסיון להיות צודק כל הזמן! אני חושד ברצינות שאילו המדע היה טוען שתאוריית האבולוציה

* לפי *בעיית האינדוקציה* (problem of induction), שניסח הפילוסוף דייוויד יום, רק משום שדבר מסוים התרחש שוב ושוב בעבר – כמו זריחת השמש – לא ניתן להסיק לוגית שהוא יקרה שוב גם בעתיד.

נכונה ברוב הזמן אך לא בכולו – או אילו המדע היה מודה שאולי, בימים מסוימים, העולם *אכן* שטוח, מי יכול לדעת? – אזי למדענים היה מוניטין חברתי טוב יותר. המדע היה נתפס כפחות לעומתי, משום שלא היינו נאלצים להתווכח עם אנשים הטוענים שהעולם שטוח – היה מקום לפשרה. כאשר אתם מתווכחים רבות, אנשים רואים בכם טיפוסים וכחנים. אם אתם מסרבים שוב ושוב להתפשר, המצב גרוע עוד יותר. ראו זאת כסוגיה של סטטוס שבטי: המדענים בהחלט הרוויחו מעמד עודף בתמורה לכלים שימושיים חברתית כגון רפואה וטלפונים סלולריים. אך הסטטוס החברתי הזה אינו מצדיק את התעקשותם על כך *שרק* רעיונות מדעיים על אבולוציה יילמדו בבתי הספר הציבוריים. גם לכמרים יש סטטוס חברתי גבוה, אחרי הכל. המדענים הרימו את האף – הם זכו במעט מעמד, וכעת הם סבורים שהם ראשי השבט כולו! עליהם להיות ענווים יותר, ולהתפשר קצת.

נראה שלאנשים רבים יש תפיסות מעורפלות למדי לגבי "ענווה רציונליסטית". מסוכן להחזיק בעיקרון נורמטיבי שאותו אתם מבינים רק במעומעם. הייצוג המנטלי שלכם עשוי לכלול דרגות חופש כה רבות, עד שיוכל להתגמש ולהצדיק כמעט כל מעשה. כאשר אנשים מחזיקים במודל מנטלי מעורפל שניתן להשתמש בו כדי לטעון כל דבר, התוצאה היא בדרך כלל אימוץ האמונות בהן הם רצו להאמין מלכתחילה. הדבר נוח כל כך, עד שאנשים נרתעים לרוב מלוותר על הערפול. ואולם, מטרתה של האתיקה שלנו היא לכוון אותנו, לא להיות מוכוונת על ידינו.

"ענווה" היא מידה טובה שלעיתים קרובות אינה מובנת כהלכה. אין פירוש הדבר שעלינו להשליך את מושג הענווה, אלא שעלינו להיות זהירים בשימוש בו. עשוי להועיל אם נבחן את *הפעולות* הנגזרות מקו מחשבה "עניו", ונשאל: "האם פעולה בדרך זו הופכת אתכם *לחזקים* יותר, או *לחלשים* יותר?" אם תחשבו על בעיית האינדוקציה כפי שהיא חלה על גשר שאמור להחזיק מעמד, המסקנה ששום דבר אינו ודאי עשויה להישמע הגיונית, ללא קשר לאמצעי הזהירות הננקטים. אך אם תבחנו את ההבדל המעשי בעולם האמיתי

בין הוספת מספר כבלי חיזוק לבין משיכת כתפיים, הרי שברור למדי איזו מן הפעולות יוצרת גשר חזק יותר.

רוב הנטיות שאני עד להן ל"ענווה רציונליסטית" הן תירוצים למשיכת כתפיים. זהו האדם שקונה כרטיס לוטו ואומר: "אבל אתה לא יכול *לדעת* שאפסיד." זהו האדם שאינו מאמין באבולוציה ואומר: "אבל אתה לא יכול *להוכיח* לי שזה נכון." זהו האדם שמסרב להתמודד עם בעיה שנראית קשה ואומר: "זה כנראה קשה מדי לפתרון." הבעיה היא ספקנות מגמתית (או, הטיית אי־האישוש) – בחינה קפדנית יותר של טענות שאיננו רוצים להאמין בהן. ענווה, בצורתה המובנת־שלא־כהלכה הנפוצה ביותר, היא תירוץ גורף לחלוטין שלא להאמין במשהו. שכן, אחרי ככלות הכל, אי אפשר להיות *בטוחים*. היזהרו מתירוצים גורפים לחלוטין!

בעיה נוספת היא שקל מדי להצהיר על ענווה. דֶנֶט, בספרו "Breaking the Spell: Religion as a Natural Phenomenon" ("לשבור את הקסם: הדת כתופעה טבעית"), מציין שבעוד שקשה מאוד להאמין בטענות דתיות רבות, קל לאנשים להאמין שהם *צריכים* להאמין בהן. דנט מכנה זאת "אמונה באמונה". מה זה אומר באמת להניח, באמת להאמין, ששלוש שווה לאחד? קל הרבה יותר להאמין שאתה *צריך*, איכשהו, להאמין ששלוש שווה לאחד, ולדקלם את התשובה הזו ברגעים הנכונים בכנסייה. דנט מציע ש"אמונה דתית" רבה צריכה להיחקר כ"הצהרה דתית" – מה שאנשים חושבים שהם צריכים להאמין בו ומה שהם יודעים שעליהם לומר.

זה קל מדי לענות לכל טיעון נגד באמירה: "ובכן, כמובן שאני *יכול* לטעות." ואז, לאחר שכרעת ברך כנדרש לכיוון ה"צניעות" ולאחר שקדת קידה כמתבקש, אתה יכול להמשיך בדרכך מבלי לשנות דבר.

הפיתוי הוא תמיד לזכות במירב הנקודות במינימום מאמץ. הפיתוי הוא לשקול בזהירות כל מידע חדש שנכנס בדרך שתצמצם למינימום את הצורך לשנות את אמונותינו, ובעיקר את *הפעולות* שלנו. ג׳ון קֶנֶת גַלבריית׳ אמר: "בפני הבחירה בין לשנות את דעתנו ובין להוכיח שאין בכך צורך, כמעט כולם משנסים מותניים לטובת

מאמץ ההוכחה.״[1] וככל שחוסר הנוחות שבשינוי הדעה גדול יותר, כך אנשים ישקיעו מאמץ רב יותר בהוכחה.

אבל אתם יודעים, אם אתם הולכים לעשות את אותו הדבר בכל מקרה, אין טעם להתאמץ כל כך ברציונליזציה שלו. לעיתים קרובות הייתי עד לאנשים שנתקלים במידע חדש, מקבלים אותו לכאורה, ואז מסבירים בזהירות מדוע הם עומדים לעשות בדיוק את אותו הדבר שתכננו לעשות קודם לכן, אך עם הצדקה שונה. מטרת החשיבה היא *לעצב* את התוכניות שלנו. אם אתם הולכים לדבוק באותן תוכניות בכל מקרה, למה לטרוח כל כך על ההצדקה? כשאתם נתקלים במידע חדש, החלק הקשה הוא *לעדכן*, *להגיב*, ולא רק לתת למידע להיעלם בתוך חור שחור. וענווה, כשהיא מובנת שלא כהלכה, יוצרת חור שחור נפלא – כל שעליכם לעשות הוא להודות שאתם עשויים לטעות.

לפיכך נכתב: ״להיות עניו משמעו לנקוט בפעולות ספציפיות מתוך ציפייה לטעויות של עצמך. להודות במוגבלותך ואז לא לעשות דבר בנידון זו אינה ענווה – זוהי התרברבות בצניעותך.״

---

1. John Kenneth Galbraith, *Economics, Peace and Laughter* (Plume, 1981), 50.

47

החלופה השלישית

האמונה בסנטה קלאוס מעניקה לילדים תחושת פליאה ומעודדת אותם להתנהג היטב בתקווה לקבל מתנות. אם האמונה בסנטה תושמד על ידי האמת, הילדים יאבדו את תחושת הפליאה שלהם ויפסיקו להתנהג יפה. לפיכך, אף על פי שהאמונה בסנטה כוזבת מבחינה עובדתית, היא בגדר 'שקר אציל' שראוי לשמר את מאזן התועלת שלו משיקולים תועלתניים.

באופן קלאסי, טיעון מסוג זה מוכר כדילמה כוזבת, כשל "האמצע המוּדר" או כשל "עסקת חבילה". גם אם נקבל את הנחות היסוד העובדתיות והמוסריות של הטיעון לעיל, הוא לא מחזיק מעמד. גם אם נניח ש"אסטרטגיית־סנטה" (עידוד ילדים להאמין בסנטה קלאוס) טובה יותר מ"אסטרטגיית האפס" (חוסר מעשה), אין משמעות הדבר שאסטרטגיית־סנטה היא *הטובה מבין כל החלופות האפשריות*. ישנן דרכי פעולה נוספות שיכולות לספק לילדים תחושת פליאה, כגון לקיחתם לצפות בשיגור של מעבורת חלל או אספקת ספרי מדע בדיוני. באופן דומה (אם זכרוני אינו מטעה אותי), מתן שוחד לילדים בתמורה להתנהגות טובה מעודד אותם להתנהג היטב רק כאשר מבוגרים צופים בהם, בעוד שמתן מחמאות ללא שוחד מוביל להתנהגות טובה ללא תנאי. שקרים אצילים* הם בדרך כלל כשלי

* *שקר אציל* (noble lie) הוא מיתוס או כזב שמנהיגים מפיצים במכוון, מתוך אמונה שהוא חיוני לשמירת הסדר החברתי. המונח מופיע לראשונה ב*המדינה* של אפלטון.

עסקת־חבילה. התגובה לכשל כזה היא שאם אנו באמת זקוקים לרווח המשוער, אנו יכולים לבנות *חלופה שלישית* כדי להשיגו.

כיצד נוכל להשיג *חלופות שלישיות*? הצעד הראשון בהשגת *חלופה שלישית* הוא ההחלטה לחפש אחת כזו, והצעד האחרון הוא ההחלטה לקבל אותה. זה נשמע מובן מאליו, ובכל זאת רוב האנשים נכשלים בשני הצעדים הללו, ולאו דווקא בתהליך החיפוש עצמו. מניין מגיעות דילמות כוזבות? חלקן צצות בתום לב, פשוט משום שחלופות עדיפות קשות לזיהוי מבחינה קוגניטיבית. אך מפעל פורה אחד לייצור דילמות כוזבות הוא הניסיון להצדיק דרך פעולה מפוקפקת באמצעות הצבעה על התועלת המשוערת שבה לעומת חוסר מעשה. במקרה כזה, המצדיק *אינו רוצה* בחלופה שלישית – מציאת חלופה כזו תשמיד את ההצדקה שלו. הדבר האחרון שתומך־סנטה רוצה לשמוע הוא שדברי שבח עובדים טוב יותר משוחד, או שחלליות יכולות לעורר השראה באותה מידה כמו איילים מעופפים.

הטוב ביותר הוא אויבו של הטוב. אם המטרה היא *באמת* לעזור לאנשים, הרי שחלופה עדיפה היא סיבה לחגיגה – ברגע שנמצא אסטרטגיה טובה יותר, נוכל לעזור לאנשים ביעילות רבה יותר. אך אם המטרה היא אך ורק להצדיק אסטרטגיה מסוימת *בטענה שהיא עוזרת לאנשים*, הרי שחלופה שלישית היא טיעון אויב,* היא מתחרה.

הפסיכולוגיה הקוגניטיבית המודרנית רואה את תהליך קבלת ההחלטות כחיפוש אחר חלופות. בחיים האמיתיים אין די בהשוואה בין אפשרויות – עליכם ליצור את האפשרויות מלכתחילה. בבעיות רבות מספר החלופות הוא עצום, ולכן נדרש "קריטריון עצירה" לחיפוש. כשאתם מבקשים לקנות בית, אינכם יכולים להשוות כל בית בעיר. בנקודה מסוימת עליכם להפסיק לחפש, ולהחליט.

אך מה קורה כאשר המניעים המודעים שלנו לחיפוש – הקריטריונים שאנו יכולים להודות בהם בפני עצמנו – אינם עולים בקנה אחד עם השפעות תת־מודעות? כאשר אנו מבצעים חיפוש

* ראו פרק 57.

אלטרואיסטי לכאורה, חיפוש אחר דרך פעולה אלטרואיסטית, ואנו מוצאים אסטרטגיה המועילה לאחרים אך מזיקה לנו – ובכן, אנחנו לא מפסיקים לחפש *שם*, אנחנו ממשיכים לחפש. כל זאת, תוך שאנו מספרים לעצמנו שאנו מחפשים אסטרטגיה שתביא תועלת אלטרואיסטית גדולה אף יותר, כמובן. אך נניח שאנו מוצאים דרך פעולה *שיש* לה תועלת שניתן להגן עליה, ושהיא *במקרה* גם נוחה לנו אישית? הרי אז נפסיק את החיפוש מיד! למעשה, אנחנו כנראה *נתנגד* לכל הצעה להתחיל לחפש שוב – בטענה של חוסר זמן, אולי. (ועם זאת, איכשהו, תמיד יש לנו דַי משאבים קוגניטיביים להמצאת הצדקות לדרך הפעולה הנוכחית שלנו.)

עליכם לחשוד כאשר אתם מוצאים את עצמכם טוענים שמדיניות מסוימת *ניתנת לצדקה* במקום שהיא *אופטימלית*, או שיש בה יתרון כלשהו בהשוואה לחוסר מעשה, במקום התועלת המירבית מבין כל הפעולות האפשריות.

דילמות כוזבות מוצגות לעיתים קרובות כדי להצדיק דרך פעולה לא אתית, שהיא – בצירוף מקרים מופלא כלשהו – גם נוחה מאוד. שקר, למשל, הוא לעיתים קרובות נוח הרבה יותר מאמירת האמת, ולהמשיך להאמין בכל מה שהתחלתם איתו זה נוח יותר מאשר לעדכן את אמונותיכם. מכאן נובעת הפופולריות של טיעונים לטובת שקרים אצילים. הם משמשים כהגנה על אמונה קיימת מראש – לא מוצאים "שקרנים אצילים" המנסחים שקר אצילי חדש ואופטימלי. הם נצמדים לכל שקר שהתחילו איתו. עדיף להפסיק את החיפוש הזה מהר!

כדי להצליח יותר, שאלו את עצמכם ישירות: *אילו ראיתי שקיימת חלופה עדיפה לדרך הפעולה הנוכחית שלי, האם הייתי שמח מעומק לבי, או שמא הייתי חש זיק קטן של רתיעה לפני שהייתי מרפה?* אם התשובות הן "לא" ו-"כן", היזהרו שמא לא באמת חיפשתם חלופה שלישית.

מה שמוביל לשאלה טובה נוספת שעליכם לשאול את עצמכם ישירות: האם הקדשתי חמש דקות, בעיניים עצומות, לסיעור מוחות של אפשרויות פרועות ויצירתיות בניסיון לחשוב על חלופה טובה

יותר? אלו חייבות להיות חמש דקות על השעון, כי אחרת אתם ממצמצים – עוצמים עיניים ופוקחים אותן שוב – ואומרים, "ובכן, כן, חיפשתי חלופות, אך לא היו כאלו." המצמוץ משמש *כחוד שחור* מוצלח שלתוכו אפשר להשליך את החובות שלכם. מומלץ להשתמש בשעון ממשי ופיזי לחלוטין.

ואותן אפשרויות פרועות ויצירתיות – האם נזהרתם שלא לחשוב על אחת טובה? האם היה מאמץ סמוי מפינת תודעתכם להבטיח שכל אפשרות שתישקל תהיה גרועה בעליל?

זה מדהים כמה שקרנים אצילים ובני מינם שׂשׂים לאמץ עבירות מוסריות – תוך קינה כנדרש על ייסורי המצפון שלהם – בשעה שהם לא הקדישו אפילו חמש דקות על השעון בחיפוש אחר חלופה. ישנם חיפושים מנטליים שאנו מייחלים בסתר ליבנו שייכשלו, וכאשר הסיכוי להצלחה אינו נוח, אנשים נאחזים בתירוץ הראשון שנקרה בדרכם כדי לוותר.

48

הגרלות לוטו: בזבוז של תקווה

הביקורת הקלאסית על הגרלות הלוטו היא שהאנשים שמשחקים הם אלו שידם אינה משגת להפסיד. שההגרלה היא בור של כסף, המנקז הון מאלו הזקוקים לו ביותר. חלק מחסידי ההגרלות, ואפילו כמה מגיבים בבלוג *Overcoming Bias*, ניסו להגן על רכישת כרטיס הגרלה בטענה שמדובר *ברכישה רציונלית של פנטזיה* – תשלום של דולר עבור יום של ציפייה נעימה ודמיון עצמי כמיליונר.

אך חִשבו מה בדיוק משתמע מכך. משמעות הדבר היא שאתם מעסיקים את מוחכם היקר בפנטזיה שההסתברות הממשית שלה היא כמעט אפס – חוט דקיק של סבירות שאתם עצמכם אינכם יכולים לעשות דבר כדי לממשו. כדורי ההגרלה הם שיכריעו את עתידכם. הפנטזיה היא על עושר שמגיע ללא מאמץ – ללא חריצות, למידה, כריזמה או אפילו סבלנות.

מה שהופך את ההגרלה לסוג אחר של בור: בור של אנרגיה רגשית. היא מעודדת אנשים להשקיע את חלומותיהם ואת תקוותיהם לעתיד טוב יותר בהסתברות מזערית. לולא ההגרלה, אולי היו מפנטזים על לימודים בבית ספר מקצועי, על פתיחת עסק משלהם, או על קבלת קידום בעבודה – דברים שייתכן שהיו מסוגלים באמת *לעשות*, תקוות שהיו מדרבנות אותם לרצות להתחזק. המוח החולם שלהם עשוי היה, בפעם העשרים שבה היו מדמיינים את אותה פנטזיה נעימה, להבחין בדרך לבצע אותה באמת. האם לא לשם כך *נועדו* חלומות ומוחות?

אך כיצד יכול תפריט כזה, המוגבל על ידי המציאות, להתחרות בסיכוי הממותק מלאכותית לעושר מיידי – לא לאחר הובלת סטארט־אפ להנפקה בבורסה, אלא סתם כך ביום שלישי בהיר אחד?

ברצינות, מדוע איננו יכולים פשוט לומר שקניית כרטיסי הגרלה היא דבר טיפשי? *בני אדם* הם טיפשים מעת לעת – זו לא אמורה להיות השערה מפתיעה כל כך.

כצפוי, המוח האנושי אינו מבצע חישוב אריתמטי של נקודה צפה ב-64 סיביות, והוא אינו יכול להפחית את ערכו של הכוח הרגשי הטמון בציפייה נעימה בפקטור של 0.00000001 מבלי לזנוח את קו המחשבה לחלוטין. כצפוי, אנשים רבים אינם מבינים שחישוב מספרי של "תוחלת התועלת" אמור *לבטל* או *להחליף* את האינסטינקטים הפיננסיים הלא־מדויקים שלהם, ובמקום זאת מתייחסים לחישוב כאל *טיעון* בלבד שיש לאזן מול הציפיות הנעימות שלהם – טיעון חלש מבחינה רגשית, שכן הוא מורכב משרבוטים גרידא על נייר, במקום חזיונות של עושר מופלג.

נראה שדי בכך כדי להסביר את הפופולריות של ההגרלות. אם כך, מדוע כה רבים חשים דחף להגן על הצורה הקלאסית הזו של *הדס עצמי*?

התהליך של התגברות על הטיות דורש (1) הבחנה בהטיה, (2) ניתוח ההטיה לפרטיה, (3) החלטה שההטיה היא רעה, (4) מציאת מעקף, ורק אז (5) יישומו. מצער לראות כמה אנשים צולחים את שלבים 1 ו-2 ואז נתקעים בשלב 3, שאמור לכאורה להיות הקל מבין החמישה. הטיות הן לימונים, לא לימונדה, ואסור לנו לנסות להכין מהן לימונדה – פשוט שִׁרפו את הלימונים האלו *עד היסוד*.

49

לוטו משודרג

יש המוסיפים לטעון כי ההגרלה אינה בזבוז של תקווה, אלא שירות המאפשר רכישת פנטזיה – "לפנטז על עושר תמורת הרבה פחות כסף מכפי שעולה לנו לפנטז על כוכבי הוליווד בקולנוע." אחד המגיבים כתב: "יש הבדל עצום בין אפס סיכוי להתעשר לבין אפסילון."* קניית כרטיס מאפשרת לחלום העושר שלכם לגשר על הפער הזה".

למעשה, אחת הנקודות שניסיתי להבהיר היא שבין סיכוי אפס להתעשר לבין סיכוי של אפסילון, קיים הבדל *בסדר גודל של אפסילון*. אם אתם מטילים בכך ספק, פשוט הגדירו את אפסילון כ-1 חלקי גוגולפלקס.

מכל מקום, אם נעמיד פנים שכרטיס הגרלה אכן מוכר "תקווה אפסילונית", הרי שהדבר מזמין אותנו לתכנן את *הלוטו המשודרג*. הלוטו המשודרג מחלק פרס אחת לחמש שנים בממוצע, בעיתוי אקראי לחלוטין – שנקבע, נאמר, על ידי דעיכה של יסוד לא־כל־כך־רדיואקטיבי. אתם קונים כרטיס פעם אחת, תמורת דולר בודד, ומקבלים בתמורה לא רק כמה ימים של סיכוי אפסילוני להתעשר, אלא כמה *שנים טובות* של אפסילון. ולא זו בלבד, אלא שהעושר יכול לנחות עליכם בכל רגע נתון! בכל שנייה, הטלפון עשוי לצלצל ולבשר לכם *שאתם, כן, אתם* מיליונרים!

---

* בחשבון דיפרנציאלי ואינטגרלי, האות היוונית אפסילון (ε) מסמנת מספר חיובי קטן כרצוננו.

תארו לעצמכם עד כמה זה יהיה טוב יותר מהגרלת פיס רגילה, המתקיימת רק במועדים מוגדרים פעמים ספורות בשבוע. נניח שהבוס נכנס ודורש מכם לשכתב הצעה, לחדש את המלאי, או כל מטלה מעצבנת אחרת. במקום לגשת לעבודה, תוכלו לפנות לטלפון הנייד ולבהות בו בציפייה לשיחה הגורלית – כי ישנו סיכוי אפסילוני שבדיוק ברגע זה *אתם, כן, אתם* תזכו בפרס הגדול! וגם אם זה לא קורה בדקה *הזו*, ובכן, אין סיבה להתאכזב – זה עשוי לקרות בדקה *הבאה*!

חשבו על שפע הפנטזיות הנוספות שההגרלה החדשה והמשופרת הזו תאפשר. תוכלו לערוך קניות בחנות, להעמיס פריטים יקרים לעגלת הקניות שלכם – ואם הטלפון הנייד שלכם לא יצלצל ויבשר על זכייה בהגרלה, תמיד תוכלו להחזיר את הפריטים למדף, נכון?

אולי הלוטו החדש והמשופר יוכל אפילו להציג התפלגות הסתברות המשתנה ללא הרף באשר לסיכויי הזכייה, ולסבירות שייבחרו מספרים מסוימים, כאשר התוחלת הכוללת מצטברת לכדי התפלגות פּוֹאָסוֹן האמורה. חשבו כמה כיף *זה* יהיה! הו אלוהים, ממש ברגע זה הסיכוי לזכייה גבוה כמעט פי עשרה מהרגיל! והביטו, למספר 42 שבחרתי כ׳מגה־בול׳ יש סיכוי זכייה כפול מהרגיל! ניתן יהיה להזין את הנתונים הללו היישר לצג הטלפונים הניידים של האנשים, כך שהם יוכלו פשוט לפתוח את המכשיר ולראות את סיכויי הזכייה שלהם מתעדכנים בזמן אמת. חשבו כמה מרגש זה יהיה! הרבה יותר מרגש מלנסות לאזן את חשבון הבנק שלכם! הרבה יותר מרגש מלהכין שיעורי בית! החלום החדש הזה יהיה *כל כך הרבה יותר מתוק*, שהוא יתחרה לא רק בתקווה ללמוד בבית ספר מקצועי, אלא אפילו בתקווה לחזור הביתה מוקדם מהעבודה. אנשים יוכלו פשוט להישאר דבוקים למסך כל היום. אחרי ככלות הכל, הם כבר לא יצטרכו לחלום על שום דבר אחר!

אכן, להציע לאנשים חלומות בהקיץ מפתים *שלעולם לא יתגשמו* זהו בהחלט שירות בעל ערך. אנשים מוכנים לשלם על כך, ולכן זה ודאי בעל ערך. החלופה היחידה היא שהצרכנים טועים, וכולנו

יודעים שזה הרי לא ייתכן.

ואף על פי כן, הממשלות הנוכחיות, עם המונופול הנתעב שלהן על ההגרלות, אינן מציעות את השירות הפשוט והמתבקש הזה. מדוע? מפני שהן רוצות לגבות מאנשים מחיר מופקע. הן רוצות שהם ימשיכו להוציא כסף מדי שבוע. הן רוצות שהאזרחים יוציאו מאה דולר עבור הריגוש שבאמונה שסיכוי הזכייה שלהם גדל פי מאה, במקום שיוכלו פשוט לבהות במסך הטלפון הנייד בהמתנה לזינוק בסבירות. לכן, אם אתם אכן מאמינים שהגרלת הלוטו היא שירות, הרי שזהו בבירור שירות יקר להחריד – הנגבה מהחברים העניים ביותר בחברה. חובתכם החגיגית כאזרחים היא לדרוש את הלוטו החדש והמשופר במקומו.

50

אבל עדיין יש סיכוי, נכון?

לפני שנים, שוחחתי עם אדם שהעיר כבדרך אגב שאינו מאמין באבולוציה. אמרתי לו: "אנחנו לא במאה התשע־עשרה. כשדארווין הציע לראשונה את תורת האבולוציה, אולי היה סביר להטיל בה ספק. אבל אנחנו במאה העשרים ואחת. אנחנו יכולים *לקרוא את הגנים*. לבני אדם ולשימפנזים יש 98% מולקולות DNA משותפות. אנחנו *יודעים* שבני אדם ושימפנזים הם קרובי משפחה. זה *נגמר*."

הוא אמר: "אולי ה־DNA פשוט דומה במקרה."

אמרתי: "הסיכויים לכך הם משהו כמו אחד לשתיים בחזקת שבע מאות וחמישים מיליון."

הוא אמר: "אבל עדיין יש סיכוי, נכון?"

ובכן, ישנן כמה סיבות לכך שהאני שלי מן העבר אינו יכול לתבוע לעצמו ניצחון מוסרי מובהק בשיחה הזו. סיבה אחת היא שאין לי שום זיכרון מניין שלפתי את המספר $2^{750,000,000}$, אם כי הוא כנראה במטא־סדר־גודל הנכון. הסיבה השנייה היא שהאני שלי מן העבר לא יישם את המושג של ביטחון מכויל.* מבין כל הפעמים בהיסטוריה האנושית שבהן אדם חישב סיכויים של $2^{750,000,000}$ נגד משהו, הוא ללא ספק טעה בתדירות גבוהה הרבה יותר מאשר פעם אחת

* *ביטחון מכויל* (calibrated confidence) הוא התאמה בין מידת הביטחון המוצהרת לבין שיעור הדיוק בפועל. אם אמרתם "70 אחוז" במאה מקרים שונים, כ־70 מהם אמורים להתברר כנכונים.

לכל $2^{750,000,000}$ פעמים. למשל, הערכת הגנים המשותפים עודכנה בהמשך ל-95%, ולא 98% – וייתכן שזה חל רק על 30,000 הגנים המוכרים ולא על הגנום כולו, ובמקרה כזה מדובר במטא־סדר־גודל שגוי.

אך אני חושב שתשובתו של האיש ההוא עדיין משעשעת למדי.

אינני זוכר מה השבתי לו בהמשך – כנראה משהו כמו "לא" – אך אני זוכר את האירוע הזה, משום שהוא העניק לי כמה תובנות לגבי חוקי המחשבה כפי שהם נתפסים בעיני אלה שאינם מוארים.

ראשית, עלה בדעתי שהאינטואיציות האנושיות מבצעות הבחנה איכותית בין *"אין סיכוי"* לבין *"סיכוי זעיר מאוד, אך כזה ששווה לעקוב אחריו"*. ניתן לראות זאת בדיון על ההגרלות באתר *Overcoming Bias*, שבו מישהו אמר: "יש הבדל גדול בין סיכוי אפס לזכות לבין סיכוי אפסילון לזכות", ואני השבתי: "לא, יש הבדל בסדר גודל של אפסילון. אם אתם מטילים בכך ספק, הגדירו את אפסילון כ-1 חלקי גוגולפלקס."

הבעיה היא שתורת ההסתברות מאפשרת לנו לעיתים לחשב סיכוי שאכן קטן מכדי להקצות לו מרחב מחשבתי – אך עד שזה קורה, הרי שכבר חישבתם אותו. אנשים מבלבלים בין המפה לשטח, כך שברמה האינסטינקטיבית, מעקב אחר הסתברות המתוארת באופן סמלי מרגיש כמו "סיכוי ששווה לעקוב אחריו", גם אם *המושא* של התיאור הסמלי הוא מספר כה זעיר שאילו היה גרגר אבק, לא הייתם מסוגלים לראות אותו. אנו יכולים להשתמש במילים כדי לתאר מספרים כה קטנים, אך לא ברגשות. רגש כה קטן פשוט אינו קיים, הוא אינו מצית מספיק נוירונים או משחרר מספיק מוליכים עצביים כדי שנחוש בו. זו הסיבה שאנשים קונים כרטיסי הגרלה – איש אינו מסוגל לחוש את קטנותה של הסתברות כה זעירה.

אך מה שמצאתי כמרתק אף יותר היה ההבחנה האיכותית בין טיעונים "ודאיים" ל"בלתי ודאיים", שלפיה אם טיעון אינו ודאי לחלוטין, מותר לכם להתעלם ממנו. כלומר, אם ההסתברות היא אפס, עליכם לזנוח את האמונה, אך אם הסבירות היא 1 חלקי גוגול, מותר

לכם לדבוק בה.

ובכן, זוהי מדינה חופשית ואיש לא אמור להשליך אתכם לכלא על חשיבה בלתי חוקית. אך אם בכוונתכם להתעלם מטיעון הקובע שהסבירות היא 1 חלקי גוגול, מדוע שלא תתעלמו גם מטיעון הקובע שהסבירות היא אפס? כוונתי היא, שכל עוד אתם מתעלמים מהראיות ממילא, מדוע זה גרוע כל כך להתעלם מראיות ודאיות יותר מאשר מראיות בלתי ודאיות?

גיליתי לעיתים קרובות בחיי שלמדתי מדוגמאות גרועות ובוטות להפליא של אנשים אחרים, לאחר שאלו הוכללו כראוי למקרים מעודנים יותר. במקרה זה, הלקח המשלים הוא שאם אינכם יכולים להתעלם מסבירות של 1 חלקי גוגול רק כי מתחשק לכם, אינכם יכולים להתעלם גם מסבירות של 0.9 רק כי מתחשק לכם. הכל הוא אותו מדרון חלקלק.

חשבו על הדוגמה של אותו אדם בכל פעם שאתם מוצאים את עצמכם מהרהרים: "אבל אתם לא יכולים *להוכיח* שאני טועה." אם אתם עומדים להתעלם מטיעון־נגד הסתברותי, מדוע שלא תתעלמו גם מהוכחה מוחלטת?

51

כשל "הכל אפור"

המתחכם: "העולם אינו שחור ולבן. אף אחד אינו טוב לגמרי או רע לגמרי. הכל אפור. לכן, איש אינו טוב יותר מרעהו."
החוקר: "מאחר שאינך רואה אלא אפור, אתה מסיק שכל גוני האפור הם אותו הגוון. אתה לועג לפשטנות שבראייה הדו־צבעית, אך מחליף אותה בראייה של צבע אחד בלבד..."

— מארק סטיגלר, *הקלע של דוד*[1]

אינני יודע אם לטעותו של "המתחכם" יש שם רשמי, אך אני מכנה אותה *כשל "הכל אפור"*. ראינו אותה מתבטאת בפרק הקודם — זה שהאמין שהסתברות של אחד לשתיים בחזקת שבע מאות וחמישים מיליון עדיין אומרת ש"יש סיכוי". כל ההסתברויות, בעיניו, היו פשוט "אי־ודאות", וזה מבחינתו נתן לו רישיון להתעלם מהן כרצונו. "הירח עשוי מגבינה ירוקה" ו"השמש עשויה בעיקר ממימן ומהליום" הן שתיהן אי־ודאויות, אך הן אינן אותה אי־ודאות.

הכל הוא אמנם גוונים של אפור, אך ישנם גוונים של אפור שהם כה בהירים עד שהם כמעט לבנים, וגוונים של אפור שהם כה כהים עד שהם כמעט שחורים. וגם אם לא, אנחנו עדיין יכולים להשוות בין גוונים ולומר "זה כהה יותר" או "זה בהיר יותר."

לפני שנים, אחד הרגעים המכוננים והמוזרים בקריירה שלי כרציונליסט היה קריאת הפסקה הבאה מתוך הספר *השחקן* מאת אִיאַן

מ. בַּנקס, ובמיוחד המשפט המודגש:[2]

> מערכת אַשְׁמָה אינה מכירה בחפים מפשע. כמו כל מנגנון כוח שחושב שכולם הם או בעדו או נגדו, אנחנו נגדו. גם אתה היית נגדו, לו היית חושב על זה. עצם צורת המחשבה שלך מציבה אותך בין אויביו. ייתכן שאין זו אשמתך, כיוון ש**כל חברה כופה חלק מערכיה על אלו שגדלו בתוכה, אך הנקודה היא שישנן חברות שמנסות למקסם את ההשפעה הזו, וישנן כאלו שמנסות למזער אותה.** אתה מגיע מאחת האחרונות ומתבקש להסביר את עצמך בפני אחת הראשונות. התחמקות תהיה קשה מכפי שאתה יכול לדמיין. ניטרליות היא ככל הנראה בלתי אפשרית. אינך יכול להתנער מהזהות הפוליטית שלך. היא אינה ישות נפרדת שניתן לנתק איכשהו משאר הווייתך, היא פונקציה של קיומך. אני יודע זאת והם יודעים זאת. מוטב שתקבל זאת.

אל תכתבו לי תגובות זועמות על כך שאם חברות כופות פחות מערכיהן, אזי לכל דור עוקב יש יותר עבודה להתחיל הכל מאפס. לא זו התובנה שלי מהפסקה.

התובנה שלי מהפסקה היא משהו שנראה כל כך מובן מאליו בדיעבד, שיכולתי להבין אותו במאה מקומות אחרים, אך משהו בפסקה הזו הפיל לי את האסימון.

היה זה עצם המושג של *הגישה הכמותית* כשהיא מיושמת על בעיות חיים כמו שיפוט מוסרי והשאיפה לשיפור עצמי — העובדה שגם אם אינך יכול לבטל משהו לגמרי, אתה עדיין יכול לשאוף להגדיל או להקטין אותו.

האם זה מובן מאליו מכדי שאצטרך לומר זאת? אני טוען שלא. בלוגרים רבים כתבו על *Overcoming Bias*: "זה בלתי אפשרי. איש אינו יכול לחסל את ההטיות לחלוטין." לא אכפת לי אם הכותב הוא כלכלן במקצועו, ברור שהוא עדיין לא הפנים את *החשיבה הכמותית* כפי שהיא חלה על חיי היומיום ועל נושאים כמו שיפור עצמי. מה שאינני יכול *למגר לחלוטין*, עדיין עשוי בהחלט להיות שווה *צמצום*.

או קחו למשל את חילופי הדברים הללו בין רובין הַנסון

לטַיילר קאוֶן. הנסון אמר שהוא מעדיף לתת משקל של לפחות 75% להמלצות התיאוריה הכלכלית לעומת האינטואיציות שלו: "אני מנסה פשוט ליישם תיאוריה כלכלית באופן ישיר, תוך הוספת מעט מאוד שיפוט אישי או תרבותי." טיילר קאון השיב:

> "לדעתי אין דבר כזה 'ליישם תיאוריה כלכלית באופן ישיר'... תאוריות תמיד מיושמות דרך המסננים האישיים והתרבותיים שלנו, ואין שום דרך אחרת שבה זה יכול לקרות."

זה נכון, אבל אתה יכול לנסות למזער את ההשפעה הזו, או שאפשר לעשות דברים שבוודאות יגדילו אותה. *ואם* אתה מנסה למזער אותה, הרי שבמקרים רבים אני לא חושב שזה יהיה בלתי סביר לקרוא לתוצאה "יישום ישיר" – אפילו בכלכלה.

"אף אחד לא מושלם." מָהַטמָה גַנדי היה לא מושלם ויוסף סטָלין היה לא מושלם, אך הם לא היו באותו גוון של חוסר שלמות. האמירה "אף אחד לא מושלם" היא דוגמה מצוינת להחלפת ראייה דו־צבעית בראייה של צבע אחד בלבד. אם תאמרו "אף אחד אינו מושלם, אך *ישנם אנשים שהם פחות בלתי־מושלמים מאחרים*," אולי לא תזכו למחיאות כפיים, אך הענקתם תקווה עבור אלו השואפים להשתפר. הרי אף אחד אינו בלתי־מושלם באופן *מושלם*.

(בכל פעם שמישהו אומר לי: "פרפקציוניזם זה רע בשבילך," אני עונה: "אני חושב שזה בסדר להיות לא מושלם, אבל עדיף לא להיות כל־כך לא מושלם עד שאנשים אחרים ישימו לב.")

כך גם לגבי האיוולת של אלו האומרים: "כל פרדיגמה מדעית כופה חלק מהנחות היסוד שלה על האופן שבו היא מפרשת ניסויים," ואז מתנהגים כאילו הוכיחו שהמדע נמצא באותה רמה כמו רפואת אליל. כל השקפת עולם כופה חלק מהמבנה שלה על התצפיות שלה, אך הנקודה היא שישנן השקפות עולם המנסות למזער את הכפייה הזו, לעומת השקפות עולם שמתענגות עליה. אין לָבָן, אך ישנם גוונים של אפור שהם הרבה יותר בהירים מאחרים, וזו איוולת להתייחס לכולם כאילו היו באותה הרמה.

אם הירח סבב את כדור הארץ במיליארדי השנים האחרונות, אם ראיתם אותו בשמיים בשנים האחרונות, ואתם מצפים לראות אותו במקומו ובמועדו מחר, הרי שזו אינה ודאות מוחלטת. ואם אתם מצפים שדרקון בלתי־נראה ירפא את בתכם מסרטן, גם זו אינה ודאות. אך אלו דרגות שונות למדי של אי־ודאות – העניין הזה של לצפות לדברים שיקרו שוב באותו אופן שחזיתם בעבר, בדיוק של שתים־עשרה ספרות אחרי הנקודה, לעומת הציפייה למשהו שיקרה תוך *הפרה בוטה* של הסדר שנצפה בעבר. לקרוא לשניהם בשם "אמונה" נראה לי כחוסר הבחנה קל.

הפסיכולוגיה סביב האמירה "גם המדע מבוסס על אמונה, אז הנה לכם!" היא מוזרה ביותר. בדרך כלל הדבר נאמר על ידי אנשים שטוענים שאמונה היא דבר *טוב*. אם כך, מדוע הם אומרים "גם המדע מבוסס על אמונה!" בטון מנצח וזועם, ולא כמחמאה? לדעתי, זוהי מחמאה מסוכנת למדי לתת מנקודת המבט שלהם. אם המדע מבוסס על "אמונה", הרי שהמדע נמצא באותה הקטגוריה כמו הדת – וניתן להשוואה ישירה אליה. אם המדע הוא דת, הרי שזו הדת שמרפאת חולים וחושפת את סודות הכוכבים. יהיה הגיוני יותר לומר: "כוהני המדע יכולים לצעוד על הירח בגלוי, באופן פומבי וניתן לאימות, כְּנֶגֶד מבוססי־אמונה, ואילו האמונה של הכוהנים שלכם לא מסוגלת לעשות את אותו הדבר." האם אתה בטוח שאתה רוצה ללכת לשם, איש האמונה? אולי, במחשבה שנייה, תעדיף לחזור בך מכל העניין הזה של "גם המדע הוא דת!"

יש כאן דינמיקה מוזרה: אתה מנסה לזקק את גוון האפור שלך ומביא אותו לנקודה שבה הוא בגוון די בהיר, ומישהו קם ואומר בטון נעלב עמוקות, "אבל זה לא לבן! זה אפור!" זה דבר אחד כשמישהו אומר: "זה לא בהיר כמו שאתה חושב, בגלל בעיות ספציפיות א' ב' ג'." זה עניין אחר לגמרי כשמישהו אומר בכעס: "זה לא לבן! זה אפור!" מבלי להצביע על שום כתמים כהים ספציפיים.

במקרה כזה, אני מתחיל לחשוד בפסיכולוגיה פגומה מהרגיל – שייתכן שמישהו כרת "ברית עם השטן" עם הטעויות שלו, וכעת הוא

מסרב לשמוע על כל אפשרות לשיפור. כשמישהו מוצא תירוץ כדי לא לנסות להשתפר, הוא לעיתים קרובות מסרב להודות שמישהו אחר יכול לנסות להשתפר. מאותו רגע כל דרך של שיפור הופכת לאויב שלו, וכל טענה שאפשר להתקדם היא מבחינתו עלבון. לכן הוא מכריז בנשימה אחת בגאווה: "אני שמח להיות אפור," ובנשימה הבאה תוקף בכעס: *"וגם אתה אפור!"*

גם אם אין שחור ולבן, עדיין יש בהיר וכהה. לא כל גווני האפור זהים.

אסימוב מצביע על כך בחיבורו *יחסיותה של הטעות*:[3]

> כשחשבו שהעולם שטוח, הם טעו. כשחשבו שהעולם כדורי, הם טעו. אבל אם אתה חושב שלחשוב שהעולם כדורי זה שגוי באותה מידה כמו לחשוב שהעולם שטוח, הרי שהטעות שלך גדולה משתי הטעויות האחרות גם יחד.

---

1. Marc Stiegler, David's Sling (Baen, 1988).
2. Iain Banks, The Player of Games (Orbit, 1989).
3. Isaac Asimov, The Relativity of Wrong (Oxford University Press, 1989).

52

סמכות מוחלטת

מישהו בא אליכם ואומר ברהבתנות: "המדע לא באמת *יודע* שום דבר. כל מה שיש לכם זה *תאוריות* – אתם לא יכולים לדעת *בוודאות* שאתם צודקים. אתם המדענים שיניתם את דעתכם לגבי האופן שבו כוח הכבידה עובד – מי אמר שמחר לא תשנו את דעתכם לגבי האבולוציה?"

חזו בפער התרבות התהומי. אם אתם חושבים שתוכלו לגשר עליו בכמה משפטים, אתם צפויים לאכזבה מרה.

בעולמם של אלו שאינם מוארים, ישנה "סמכות" ו"אי־סמכות". מה שניתן לסמוך עליו, ניתן לסמוך עליו. מה שלא ניתן לסמוך עליו, מוטב להשליך לפח. ישנם מקורות מידע טובים ומקורות מידע רעים. אם מדענים שינו אי־פעם את הגרסה שלהם בהיסטוריה, אזי המדע אינו יכול לשמש כסמכות אמיתית, ולעולם לא ניתן יהיה לסמוך עליו שוב – ממש כמו עד שנתפס משקר, או עובד שנתפס גונב מהקופה.

בנוסף, אותו אדם מניח כמובן מאליו שמצופה ממי שמציג רעיון להגן עליו מפני כל טיעון נגד אפשרי, ולא להודות בכלום. כל הטענות מאבדות ממשקלן בהתאם. אם אפילו *תומכי* המדע מודים שהמדע אינו מושלם, ובכן, הוא ודאי כמעט חסר ערך.

כשמישהו חי את חייו כשהוא רגיל לוודאות, אינכם יכולים פשוט לומר לו: "המדע הוא הסתברותי, בדיוק כמו כל ידע אחר." הם יקבלו את החצי הראשון של המשפט כהודאה באשמה, ויפטרו את החצי

השני כניסיון נואש להאשים את כל השאר רק כדי לחמוק מביקורת.

״הודית שאינך ראוי לאמון – אז הסתלק לך, מדע, והנח לנו לנפשנו!״

מקור ברור אחד לדפוס החשיבה הזה הוא הדת. בדת, כתבי הקודש אמורים להגיע מאלוהים, ולכן הודאה ולו בפגם קטן תהרוס את סמכותם כליל. משום כך, כל שמץ של ספק הוא חטא, והצהרה על ודאות היא *חובה*, בין שאתה ודאי ובין שלא.

אך אני חושד שגם למשטר הבית־ספרי המסורתי יש קשר לכך. המורה אומרת לכם דברים מסוימים, ועליכם להאמין להם, ועליכם לדקלם אותם במבחן. אבל כשתלמיד מעלה הצעה בכיתה, אתם לא חייבים לקבל אותה – אתם חופשיים להסכים או לא להסכים (כך נדמה), ואיש לא יעניש אתכם.

החוויה הזו, אני חושש, מְמַפָּה את תחום האמונה לתחומים החברתיים של *סמכות, פקודה וחוק*. בתחום החברתי, יש הבדל איכותי בין חוקים מוחלטים לחוקים שאינם מוחלטים, בין פקודות להצעות, בין גורמי סמכות למי שאינם כאלה. נראה שישנו ידע מחייב וידע שאינו מחייב, בדומה לתקנה מחייבת לעומת המלצה. לסמכויות נוקשות יש להיכנע, בעוד שלהצעות שאינן נוקשות ניתן להישמע או להתעלם כעניין של העדפה אישית. והמדע, מכיוון שהוא מודה בעצמו שיש לו אפשרות לטעות, בהכרח משתייך לקבוצה השנייה.

(אני מציין כבדרך אגב שאני רואה דמיון מסוים לאלו שחושבים שאם לא קיבלתם ״הסתברות מוסמכת״ כתובה על דף מהמורה בכיתה, או כזו שהועברה ממקור דומה שאין עליו עוררין, אזי חוסר הוודאות שלכם אינו עניין לתורת ההסתברות הבייסיאנית. מישהו עלול – *חלילה!* – להתווכח עם ההערכה שלכם לגבי ההסתברות הפריורית. לאלו שאינם מוארים לחלוטין, נראה שהסתברויות פריוריות שייכות לסוג האמונות שמציעים תלמידים, ולא לסוג האמונות שמוכתבות על ידי מורים. זהו אינו *ידע* של ממש.)

הפער התרבותי התהומי בין הגישה הסמכותית לגישה הכמותית מעצבן למדי את אלו מאיתנו שבוהים בו מהצד הרציונליסטי. הנה

מישהו שמאמין שיש לו ידע *יותר* אמין מהניחושים ההסתברותיים גרידא של המדע – כמו הניחוש שהירח יזרח במקומו ובמועדו מחר, בדיוק כפי שעשה בכל לילה שנצפה מאז החל התיעוד האסטרונומי, ובדיוק כפי שנחזה על ידי תיאוריות פיזיקליות שתחזיותיהן הקודמות אוששו בהצלחה בדיוק של 14 ספרות אחרי הנקודה. ומהו הידע הזה שהבלתי־מוארים מציבים מעל שלנו, ומדוע? זוהי כנראה איזו מגילה ישנה וטחובה שהופרכה באחת־עשרה דרכים שונות. ובכל זאת, היא אמינה יותר מהמדע (לטענתם), משום שהיא אף פעם לא מודה בטעות ואף פעם לא משנה את דעתה, ולא משנה כמה פעמים הופרכה. הם זורקים את המילה "ודאות" מצד לצד כמו כדור טניס, משתמשים בה בקלילות של נוצה, בעוד מדענים כורעים תחת עולו של ספק הכרחי, נאבקים לבסס ולו שמץ של הסתברות. "אני מושלם," הם אומרים ללא שמץ דאגה, "אני ודאי נמצא כל כך הרבה *מעליכם*, אתם שעדיין נאלצים לעמול על שיפור עצמי."

אין שום דבר פשוט שניתן לומר להם – שום סתירה *מהירה* ומוחצת. בחשיבה זהירה, אולי תוכלו לזכות באהדת הקהל אם מדובר בוויכוח פומבי. לרוע המזל אינכם יכולים פשוט לפלוט: "בן תמותה טיפש, הגישה הכמותית היא מעבר לבינתך, והאמונות שאתה מכנה בקלילות 'ודאיות' מבוססות פחות מאשר המופרכת שבהשערותינו." זהו הבדל בחוויית חיים ותפיסת עולם שכלל לא פשוט לתאר במילים, לבטח לא במהירות.

מה בכל זאת תוכלו לנסות, מבחינה רטורית, מול קהל? קשה לומר... אולי:

- "כוחו של המדע נובע מהיכולת לשנות את דעתנו ולהודות שטעינו. אם מעולם לא הודית שאתה טועה, זה לא אומר שעשית פחות טעויות."
- "כל אחד יכול *לומר* שהוא ודאי לחלוטין. קצת יותר קשה באמת, אבל באמת, לא לעשות טעויות לעולם. מדענים מבינים את ההבדל, ולכן הם לא אומרים שהם ודאיים לחלוטין. זה הכל. זה לא אומר שיש להם סיבה ספציפית

לפקפק בתיאוריה – כל בדל של ראיה יכול להצביע לאותו כיוון, כל הכוכבים וגרמי השמים יכולים להסתדר כמו דומינו בתמיכה בהשערה אחת, והמדענים עדיין לא יגידו שהם בטוחים לחלוטין, פשוט כי יש להם סטנדרטים גבוהים יותר. זה לא אומר שלמדענים יש פחות *הצדקה* להיות בטוחים בעצמם מאשר, נניח, לפוליטיקאים שתמיד נראים כל כך בטוחים בהכל."

- "מדענים לא משתמשים בביטוי 'לא ודאי לחלוטין' כפי שאתם רגילים בשיחה יומיומית. נניח שהלכת לרופא ועשית בדיקת דם, והרופא חזר ואמר: 'הרצנו כמה בדיקות, וזה לא ודאי לחלוטין שאתה לא עשוי מגבינה, ויש סיכוי שאינו אפס שעשרים פְּיות עשויות שוקולד תבוני שרות שירים מ"פרפר נחמד" בתוך המעי הדק שלך'. במקרה כזה, ברח על נפשך, הרופא שלך צריך רופא בעצמו. כשמדען אומר את אותו הדבר, פירוש הדבר שהוא חושב שההסתברות כל כך זעירה שלא תוכלו לראות אותה אפילו במיקרוסקופ אלקטרונים, אבל המדען מוכן לבחון את הראיות במקרה הבלתי סביר שבו אכן יהיו לכם כאלו."
- "האם הייתם מוכנים לשנות את דעתכם לגבי הדברים שאתם מכנים 'ודאיים' אילו ראיתם מספיק ראיות? כלומר, נניח שאלוהים בכבודו ובעצמו היה יורד מן השמים ואומר לכם שכל הדת שלכם נכונה מלבד מצווה אחת. אם הדבר היה גורם לכם לשנות את דעתכם, אינכם יכולים לומר שאתם בטוחים לחלוטין באותה המצווה. מסיבות טכניות של תורת ההסתברות, אם תיאורטית אפשרי שתשנו את דעתכם לגבי משהו, ההסתברות שלו אינה יכולה להיות שווה בדיוק לאחד. אי־הוודאות עשויה להיות קטנה מגרגיר אבק, אך היא חייבת להיות שם. ואם לא הייתם משנים את דעתכם גם אם אלוהים היה אומר לכם אחרת, אז אני מניח שיש לכם בעיה של סירוב

להודות בטעות, שחורגת כנראה מכל מה שבן תמותה כמוני מסוגל להסביר לכם."

השאלה המעניינת יותר היא מה אומרים למישהו *שלא* מול קהל. איך מתחילים את התהליך הארוך של ללמד אדם כיצד לחיות ביקום נטול ודאות?

אני חושב שהשלב הראשון והבסיסי צריך להיות ההבנה שאפשר לחיות בלי ודאות – שאם, *לצורך הדיון*, לא יכולתם להיות בטוחים בשום דבר, זה לא היה שולל מכם את היכולת לבצע הבחנות מוסריות או עובדתיות. בפרפרזה על דבריה של לויס בוז'ולד, "אל תדחפו חזק יותר – הקטינו את ההתנגדות."

אחת *ההגנות* הנפוצות על תפיסת ה"סמכות המוחלטת" היא משהו שאני מכנה "הטיעון מתוך הטיעון מתוך האפור", שמתנהל בערך כך:

- *רלטיביסטים מוסריים* אומרים:
  - העולם אינו שחור ולבן, לכן:
  - כל דבר הוא אפור, לכן:
  - אף אחד אינו טוב יותר מאף אחד אחר, לכן:
  - אני יכול לעשות מה שבא לי ואתם לא יכולים לעצור אותי בוהאהאהא.
- אבל אנחנו חייבים להיות מסוגלים לעצור אנשים מלבצע רצח.
- לכן חייבת להימצא דרך כלשהי להיות ודאיים לחלוטין, אחרת הרלטיביסטים המוסריים מנצחים.

טיפשות הפוכה איננה חוכמה. אינכם יכולים להגיע לתשובה נכונה על ידי היפוך *כל שורה ושורה* בטיעון שמסתיים במסקנה גרועה. כל שורה ושורה חייבת להיות נכונה כדי שטיעון מתמטי יהיה תקף. וזה שמסקנתם של הרלטיביסטים גרועה לא אומר שטענתם "העולם אינו שחור ולבן" היא שקר, בדיוק כפי שהעובדה שסטלין האמין ש- 2+2=4 לא הופכת את המשוואה הזו לשקרית. הטעות היא בקפיצה מראייה דו־צבעית לראייה חד־צבעית – הטענה שכל גווני האפור הם אותו הגוון.

זה יהיה ויתור גדול מדי (למעשה, ויתור על הטיעון כולו) להסכים

להנחה שאתם זקוק לידע מוחלט על אפשרויות טובות לחלוטין ורעות לחלוטין כדי להיות מוסריים. אתם יכולים להחזיק בידע בלתי־ודאי לגבי אפשרויות טובות יחסית ורעות יחסית, ועדיין לבחור. למעשה זה אמור להיות עניין שבשגרה, לא משהו שיש להתרגש ממנו יתר על המידה.

כלומר, כן, אם עליכם לבחור בין שתי חלופות **א** ו־**ב**, ואתם מצליחים איכשהו לבסס ביטחון של 100% מכויל היטב ואמין לגמרי ש־**א** רצויה לחלוטין ובאופן מלא, וש־**ב** היא סיכום של כל מה שרע ומגעיל בעולם, הרי שזהו תנאי *מספיק* לבחירה ב־**א** על פני **ב**. אך זהו אינו תנאי *הכרחי*.

בואו נראה, מה עוד הם צריכים לדעת? ובכן, ישנה הַתרבות הרציונליסטית כולה שאומרת שספק, שאלות והודאה בטעות אינם דברים מבישים ונוראיים.

ישנו כל המושג של השגת מידע באמצעות *התבוננות בדברים*, במקום באמצעות הטפה. כשמסתכלים על דברים לעומק, לעיתים מגלים שהם שונים ממה שחשבתם שהם במבט ראשון. אך זה לא אומר שהטבע שיקר לכם, או שעליכם לזנוח את כושר הראייה.

ואז ישנו המושג של ביטחון מכויל – העובדה ש"הסתברות" אינה אותו מושג כמו המחוון הקטן בראש שלכם שמודד את המחויבות הרגשית שלכם לרעיון. זה יותר כמו מדד לכמה פעמים, באופן פרגמטי, בחיים האמיתיים, אנשים במצב אמונה מסוים אומרים דברים שהם באמת נכונים. אם תיקחו מאה אנשים ותבקשו מכל אחד מהם לטעון טענה שבה הוא "בטוח לחלוטין", כמה מהטענות הללו יהיו נכונות? לא מאה.

למעשה, לטענות שאנשים באמת פנאטיים לגביהן יש סיכוי *נמוך בהרבה* להיות נכונות מאשר טענות כמו "השמש גדולה מהירח", שנראות מובנות מאליהן מכדי שנתרגש מהן. על כל טענה שתוכלו למצוא שמישהו "בטוח בה לחלוטין", כנראה שתוכלו למצוא מישהו "בטוח לחלוטין" בדיוק בהיפוכה הגמור, כיוון שהצהרות אמונה פנאטיות כאלה אינן צצות בהיעדר התנגדות. לכן המחוון הקטן

בראשם של אנשים, המודד את המחויבות הרגשית שלהם לאמונה, אינו מִתרגם היטב לביטחון מכויל – הוא אפילו לא מתנהג בצורה מונוטונית.

באשר ל"ודאות מוחלטת" – ובכן, אם אתם אומרים שמשהו הוא בהסתברות של 99.9999%, זה אומר שאתם חושבים שתוכלו לצאת *במיליון* טענות חזקות ועצמאיות באותה מידה, *אחת אחרי השנייה*, לאורך שנה שלמה בערך, ולטעות, בממוצע, בערך פעם אחת. זה מדהים מספיק. (זה מדהים להבין שאנחנו באמת יכולים *להגיע* לרמת ביטחון כזו עבור הטענה "לא תזכה בלוטו.") לכן, בואו לא נאמר דבר על הסתברות 1.0. ברגע שתבינו שאינכם *זקוקים* להסתברויות של 1.0 כדי להסתדר בחיים, תבינו כמה זה מגוחך לחלוטין לחשוב שתוכלו אי פעם להגיע ל-1.0 עם מוח אנושי. הסתברות של 1.0 אינה רק ודאות, היא *ודאות אינסופית*.

למעשה, נראה לי שכדי למנוע אי־הבנה בציבור, אולי מדענים צריכים להסתובב ולומר "איננו בטוחים במידה *אינסופית*" במקום "איננו בטוחים." שכן הניסוח השני, בשפת היומיום, מרמז על כך שאתם מכירים סיבה ספציפית לספק.

53

איך לשכנע אותי ש- 2+2=3

בפרק *מהי ראיה?* (פרק 21) כתבתי:

> 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... על כן הביטוי – "אמונה עיוורת". אם מה שאתה מאמין בו אינו תלוי במה שאתה רואה, הרי שאתה מתהלך באפילה ממש כאילו ניקרת את עיניך שלך.

ג'יהאן בּאראן השיב:

> אינני יכול להעלות בדעתי סיטואציה שתהפוך את 2+2=4 לשקרי. אולי מהסיבה הזו, האמונה שלי בכך ש- 2+2=4 היא בלתי מותנית.

אני מודה, אינני יכול להעלות בדעתי "סיטואציה" *שתהפוך* את 2+2=4 לשקרי. (אפשר אמנם להגדיר מחדש את המונחים, אך אלו אינן "סיטואציות" – ובמקרה כזה אתם כבר לא באמת מדברים על 2, על 4, על =, או על +.) אך זה לא הופך את האמונה שלי לבלתי מותנית. אני מוצא שקל למדי לדמיין סיטואציה *שתשכנע* אותי ש- 2+2=3.

נניח שהייתי קם בוקר אחד, מוציא שני אטמי אוזניים, מניח אותם

לצד שני אטמי אוזניים אחרים על שידת הלילה שלי, ומבחין שישנם כעת שלושה אטמי אוזניים בסך הכל, מבלי שאף אטם אוזניים הופיע או נעלם – זאת בניגוד לזיכרון המאוחסן שלי, לפיו 2+2 אמור להיות שווה ל-4. יתרה מזאת, כשהייתי משחזר את התהליך במוחי, היה נראה שכדי לגרום ל-XX ו-XX להפוך ל-XXXX נדרש X נוסף שיופיע משום מקום. הדבר גם לא היה עקבי עם פעולות אריתמטיות אחרת שדמיינתי, שכן חיסור של XX מתוך XXX היה מותיר XX, אך חיסור של XX מתוך XXXX היה מותיר XXX. זה היה מתנגש עם הזיכרון המאוחסן שלי ש- 3–2=1, אך אותו זיכרון היה נראה אבסורדי אל מול האישור הפיזי והמנטלי ש- XXX–XX=XX.

הייתי בודק גם במחשבון כיס, בגוגל, ואולי בעותק של הספר *1984* שברשותי, שבו וינסטון כותב: "חירות היא החירות לומר ששתיים ועוד שתיים שווה שלוש." כל אלו היו מראים, כמובן, ששאר העולם מסכים עם התפיסה הנוכחית שלי, וחולק על הזיכרון שלי. כולם היו מסכימים ש- 3=2+2.

איך בכלל יכולתי לשגות באשליה כזו ולהאמין ש- 4=2+2? שני הסברים היו עולים בדעתי. ראשית, פגם נוירולוגי (שנגרם אולי מעיטוש) שגרם לכל תוצאות החיבור בזיכרון המאוחסן שלי לגדול באחד. שנית, שמישהו מתעתע בי, באמצעות היפנוזה או משום שאני חי בתוך סימולציית מחשב. במקרה השני, הייתי חושב שסביר יותר שהם שיבשו את מנגנון *השליפה* האריתמטית שלי מאשר ש- 2+2 *באמת* היה שווה 4 אי־פעם. אף אחד מההסברים הסבירים הללו לא היה מונע ממני להבחין בכך שאני מאוד, מאוד, *מאוד* מבולבל.

מה שהיה משכנע אותי ש- 3=2+2, במילים אחרות, הוא בדיוק אותו סוג של ראיות שמשכנע אותי כעת ש- 4=2+2: האש הצולבת הראייתית של תצפית פיזית, תפיסת מציאות מנטלית והסכמה חברתית.

היה זמן שבו לא היה לי שום מושג ש- 4=2+2. לא הגעתי לאמונה *החדשה* הזו באקראי – אחרת לא הייתה שום סיבה מיוחדת שהמוח שלי יסיים את התהליך כשהוא מאחסן "4=2+2" במקום "7=2+2".

העובדה שהמוח שלי מאחסן תשובה הדומה להפליא למה שקורה כשאני מניח שני אטמי אוזניים לצד שני אטמי אוזניים נוספים, מצריכה הסבר. מהי אותה שזירה המייצרת את השיקוף המוזר הזה בין המחשבה למציאות?

ישנן למעשה רק שתי אפשרויות לגבי כל אמונה בעובדה כלשהי – או שהאמונה הגיעה לשם דרך תהליך של שזירה בין המחשבה למציאות, או שלא. אם לא, האמונה אינה יכולה להיות נכונה אלא במקרה. עבור אמונות בעלות ולו שמץ קל של מורכבות פנימית (כזו הדורשת תוכנת מחשב בת יותר מ-10 ביטים כדי לדמות אותה), מרחב האפשרויות גדול דיו כדי לשלול כל סיכוי לצירוף מקרים.

עובדות בלתי מותנות אינן זהות לאמונות בלתי מותנות. אם ראיות שזורות משכנעות אותי שעובדה מסוימת היא בלתי מותנית, אין פירוש הדבר שתמיד האמנתי באותה עובדה ללא צורך בראיות שזורות.

אני מאמין ש- 2+2=4, ואני מוצא שקל לי למדי להעלות בדעתי סיטואציה שתשכנע אותי ש- 2+2=3. דהיינו, אותו סוג של סיטואציה שמשכנעת אותי כעת ש- 2+2=4. לפיכך, אינני חושש שאני קורבן של אמונה עיוורת.

אם ישנם יהודים מאמינים בקהל *שמכירים את משפט בייס* (בלי נומרופובים, בבקשה), האם אוכל לשאול אתכם איזו סיטואציה הייתה משכנעת אתכם באמיתות האסלאם? ככל הנראה, מדובר בדיוק באותה סיטואציה שאחראית סיבתית ליצירת אמונתכם הנוכחית ביהדות: הייתם מגיחים בצרחות מרחמה של אישה מוסלמית, וגדלים אצל הורים מוסלמים שאומרים לכם ללא הרף שטוב להאמין באסלאם באופן בלתי מותנה. או שיש בזה יותר מזה? אם כן, איזו סיטואציה הייתה משכנעת אתכם באסלאם, או לפחות, באי־יהדות?

54

ודאות אינסופית

בפרק *סמכות מוחלטת* (פרק 52), טענתי שאינכם זקוקים לוודאות אינסופית:

> אם עליכם לבחור בין שתי חלופות **א** ו-**ב**, ואתם מצליחים איכשהו לבסס ביטחון של 100% מכויל היטב ואמין לגמרי ש-**א** רצויה לחלוטין ובאופן מלא, וש-**ב** היא סיכום של כל מה שרע ומגעיל בעולם, הרי שזהו תנאי *מספיק* לבחירה ב-**א** על פני **ב**. אך זהו אינו תנאי *הכרחי*... אתם יכולים להחזיק בידע בלתי-ודאי לגבי אפשרויות טובות יחסית ורעות יחסית, ועדיין לבחור. למעשה זה אמור להיות עניין שבשגרה.

בנוגע לטענה ש- 2+2=4, עלינו להבחין בין המפה לשטח. בהינתן היציבות והאוניברסליות המוחלטת-לכאורה של חוקי הפיזיקה, ייתכן שמעולם, בכל תולדות היקום, אף חלקיק לא חרג ממגבלת מהירות האור המקומית. כלומר, ייתכן שמגבלת מהירות האור אינה נכונה רק ב-99% מהזמן, או ב-99.9999% מהזמן, או בכל זמן פרט לאחד חלקי גוגולפלקס מהזמן, אלא פשוט נכונה *תמיד ובאופן מוחלט*.

אך השאלה האם נוכל אי פעם להחזיק *בביטחון מוחלט* במגבלת מהירות האור היא שאלה אחרת לגמרי. המפה אינה השטח.

ייתכן שזו אמת לאמיתה שסטודנט העתיק את המטלה שלו, אך השאלה האם יש לכם ידע כלשהו על עובדה זו – שלא לדבר על

ביטחון *מוחלט* באמונה זו – היא סוגיה נפרדת. אם אתם מטילים מטבע ואז לא מביטים בו, ייתכן שזה נכון לחלוטין שהמטבע מראה "עץ", ואתם עשויים שלא להיות בטוחים לחלוטין האם המטבע מראה עץ או פְּלִי. דרגה של אי-ודאות אינה זהה לדרגה של אמת או לשכיחות של התרחשות.

אותו הדין חל על אמיתות מתמטיות. מוטל בספק האם ניתן לומר שהקביעה "2+2=4" או "באריתמטיקה של פֶּאנוֹ, SSS0 = SS0 + SS0" היא *נכונה* בכל מובן מופשט וטהור, במנותק ממערכות פיזיקליות שנראה כי הן מתנהגות בדרכים דומות לאקסיומות של פאנו. אחרי שאמרתי זאת, אסתכן בניחוש שבכל מובן שבו "2+2=4" הוא נכון בכלל, הוא נכון תמיד ובאופן מדויק. הוא אינו רק נכון בערך ("2+2 שווה למעשה 4.0000004") או נכון ב-999,999,999,999 מקרים מתוך 1,000,000,000,000.

אינני בטוח לחלוטין מה המשמעות של "נכון" במקרה זה, אך אני עומד מאחורי הניחוש שלי. האמינות של הקביעה "המשפט 2+2=4 נכון תמיד" עולה בהרבה על האמינות של כל עמדה פילוסופית ספציפית לגבי מה משמעות המילים "נכון", "תמיד", או "המשפט" בקביעה לעיל.

עם זאת, אין זה אומר שיש לי *ביטחון מוחלט* ש- 2+2=4. ראו את הדיון בפרק הקודם על הדרך לשכנע אותי ש- 2+2=3 – דבר שניתן לעשות בדיוק באמצעות אותו סוג של ראיות ששכנעו אותי ש- 2+2=4 מלכתחילה. ייתכן שהזיתי את כל הראיות הקודמות הללו, או שאני זוכר אותן בצורה שגויה. בדברי הימים של הנוירולוגיה תועדו כבר ליקויים מוחיים מוזרים מאלו.

לכן, אם אנו מצמידים הסתברות כלשהי לקביעה "2+2=4", מה היא צריכה להיות? מה שאתם מבקשים להשיג במקרה כזה הוא כיול טוב - מצב שבו קביעות שלהן אתם מייחסים "הסתברות של 99%" מתבררות כנכונות 99 פעמים מתוך 100. זה למעשה הרבה יותר קשה מכפי שאתם עשויים לחשוב. קחו מאה אנשים, ובקשו מכל אחד מהם לספק עשר טענות שבהן הם מרגישים "99% ביטחון."

מתוך 1,000 הטענות, האם אתם חושבים שבערך 10 יהיו שגויות?

אינני עומד לדון בניסויים הממשיים שנעשו על כיול - אתם יכולים למצוא אותם במאמרי "Cognitive biases potentially affecting judgment of global risks" ("הטיות קוגניטיביות העלולות להשפיע על שיפוט של סיכונים גלובליים"). ראיתי שכאשר אני פשוט מציג זאת לאנשים ללא הכנה מתאימה, הם משתמשים בזה לאחר מכן כטיעון-נגד גורף, שקופץ משום מה לראשם בכל פעם שהם צריכים להטיל ספק ברמת הוודאות של אדם שאת דעתו הם אינם אוהבים, אך אינו זמין להם כשהם שוקלים את דעותיהם שלהם. לכן, אני משתדל שלא לדבר על הניסויים בכיול אלא כחלק מהצגה מובנית של רציונליות, הכוללת אזהרות מפני ספקנות מגמתית.

אך הכיול הנצפה של בני אדם הטוענים כי הם בטוחים "ב-99%" אינו 99% דיוק.

נניח שאתם אומרים שאתם בטוחים ב-99.99% ש- 2+2=4. אם כך, הרגע טענתם שתוכלו לנסח 10,000 טענות *בלתי תלויות*, שבהן יש לכם מידת ביטחון זהה, ולטעות, בממוצע, בערך פעם אחת. אולי עבור 2+2=4 דרגת הביטחון יוצאת הדופן הזו תהיה אפשרית: "2+2=4" הוא טיעון פשוט ביותר, מתמטי כמו גם אמפירי, ומקובל באופן נרחב מבחינה חברתית (לא מתוך אישור נלהב אלא פשוט כעובדה הנלקחת כמובנת מאליה). אז אולי באמת תוכלו להגיע לביטחון של 99.99% בנושא הזה.

אינני חושב שתוכלו להגיע לביטחון של 99.99% עבור טענות כמו "53 הוא מספר ראשוני." כן, זה נראה סביר, אך עד שתנסו להגדיר פרוטוקולים שיאפשרו לכם להנפיק 10,000 טענות *בלתי תלויות* מסוג זה – כלומר, לא רק אוסף של טענות על מספרים ראשוניים, אלא פרוטוקול חדש בכל פעם – אתם תיכשלו יותר מפעם אחת.

ועדיין המפה אינה השטח: אם אני אומר שאני בטוח ב-99% ש- 2+2=4, אין זה אומר שאני חושב ש-"2+2=4" נכון ברמת דיוק של 99%, או ש-"2+2=4" נכון ב-99 פעמים מתוך 100. הטענה עליה אני מדבר היא ש-"2+2=4 נכון תמיד ובאופן מדויק", ולא הטענה

ש-"4=2+2 נכון לרוב ובדרך כלל".

באשר לרעיון שתוכלו להגיע ל-100% ביטחון בטענה מתמטית – נו באמת, באמת! אם אתם מצהירים על בטחון של 99.9999%, אתם רומזים שתוכלו לנסח *מיליון* טענות שקולות מבחינת רמת הסיכון שלהן, אחת אחרי השנייה, ולטעות בממוצע בערך פעם אחת. זה שווה ערך לשנה שלמה של דיבור, בהנחה שאתם מסוגלים להפיק טענה אחת כל 20 שניות ואתם מדברים 16 שעות ביממה.

הצהירו על 99.9999999999% ביטחון, והעליתם זאת לטריליון. כעת אתם הולכים לדבר במשך מאה תקופות חיים אנושיות, ולא לטעות אפילו פעם אחת?

הצהירו על ביטחון המותיר סיכוי לטעות של אחד חלקי גוגולפלקס, והאגו שלכם יתנפח הרבה מעבר לזה של חולי נפש הסבורים שהם אלוהים.

והרי גוגולפלקס הוא מספר קטן בהרבה אפילו ממספרים עצומים באופן בלתי נתפס שהם "קטנים" יחסית, כמו $3\uparrow\uparrow\uparrow 3$. אך אפילו ביטחון של $\left(1-{}^{1}\!/_{3\uparrow\uparrow\uparrow 3}\right)$ אינו קרוב יותר ל-**הסתברות 1** מאשר להיות בטוח במשהו ב-90%.

אם כל השאר ייכשל, "שליטי האופל של המטריקס" ההיפותטיים – *שממש עכשיו* מתעתעים בהערכת האמינות של המוח שלכם לגבי המשפט הזה ממש – יחסמו את הנתיב ויגנו עלינו מפני נגע הוודאות האינסופית.

האם אני בטוח בכך לחלוטין? ובכן, כמובן שלא.

כפי שרָפָאל שְׁמיגרוֹדסקי אמר פעם:

> לדעתי צריכה להיות לך האפשרות לייחס מידת ודאות נמוכה מ-1 למושגים המתמטיים הנחוצים כדי לגזור את משפט בייס עצמו, ועדיין לעשות בו שימוש מעשי. אינני בטוח לחלוטין שאני חייב להיות תמיד לא־בטוח. יתכן שיש לי יסוד מוצדק להיות בטוח במשהו. אך ברגע שאני מקצה הסתברות של 1 לטענה כלשהי, לעולם לא אוכל לבטל זאת. לא משנה מה אראה או אלמד,

אהיה חייב לדחות כל דבר שחולק על האקסיומה. אני לא אוהב את הרעיון שלא אוכל לשנות את דעתי, לעולם.

55

0 ו-1 אינם הסתברויות

אחת, שתיים ושלוש הם כולם מספרים שלמים, וכך גם מינוס ארבע. אם תמשיכו לספור מעלה או מטה, אתם עתידים להיתקל בהרבה יותר מספרים שלמים. אתם לא תיתקלו, עם זאת, בשום דבר שנקרא "אינסוף חיובי" או "אינסוף שלילי", לכן אלו אינם מספרים שלמים.

אינסוף חיובי ושלילי אינם מספרים שלמים, אלא סמלים מיוחדים שנועדו לדיבור על ההתנהגותם של מספרים שלמים. אנשים אומרים לפעמים משהו כמו "5 + אינסוף = אינסוף," כיוון שאם תתחילו ב-5 ותמשיכו לספור כלפי מעלה מבלי לעצור לעולם, תקבלו מספרים גבוהים יותר ויותר ללא כל הגבלה. אך אין להסיק מכך ש-"אינסוף – אינסוף = 5." אינכם יכולים לספור מעלה מ-0 מבלי לעצור לעולם, ואז לספור מטה מבלי לעצור לעולם, ואז למצוא את עצמכם ב-5 כשסיימתם.

מכאן אנו יכולים לראות שאינסוף הוא לא סתם לא־מספר־שלם, הוא אפילו לא מתנהג כמו מספר שלם. אם תנסו בחוסר תבונה לערבב אינסופים עם מספרים שלמים, תזדקקו לכל מיני התנהגויות חדשות שנראות לא־עקביות ושאינכם זקוקים להן עבור 1, 2, 3 ומספרים שלמים *אמיתיים* אחרים.

אף שאינסוף אינו מספר שלם, אינכם צריכים לחשוש שתישארו ללא מספרים. למרות שאנשים ראו חמישה כבשים, מיליוני גרגירי חול, וספטיליוני אטומים, אף אחד מעולם לא ספר אינסוף של

שום דבר. אותו הדין חל על כמויות רציפות – אנשים מדדו גרגירי אבק ברוחב מילימטר, חיות ברוחב מטר, ערים ברוחב קילומטרים וגלקסיות ברוחב אלפי שנות אור, אך אף אחד מעולם לא מדד שום דבר ברוחב אינסוף. בעולם האמיתי, אינכם *זקוקים* ליותר מדי אינסוף.

(אציין כאן עבור הקוראים המתוחכמים יותר בקהל, שאין צורך לכתוב לי הסברים מפורטים על ההבדל, למשל, בין מספרים סודרים למספרים מונים. כן, אני בקי בהגדרות מתקדמות שונות של אינסוף מתוך תורת הקבוצות, אך אינני רואה בהן שימוש טוב עבור תורת ההסתברות. ראו להלן.)

בדרך הרגילה שבה נכתבות הסתברויות, ההסתברויות נעות בין 0 ל-1. למטבע עשויה להיות הסתברות של 0.5 ליפול על "פֶּלִי", או שהחזאי עשוי להקצות הסתברות של 0.9 לגשם מחר.

אולם, זוהי אינה הדרך היחידה לכתוב הסתברויות. לדוגמה, ניתן להפוך הסתברויות ליחסי סיכויים דרך הטרנספורמציה $O=(P/(1-P))$ כך שהסתברות של 50% תהפוך ליחס סיכויים של 0.5/0.5 או 1, מה שנכתב בדרך כלל כ-1:1, בעוד שהסתברות של 0.9 תהפוך ליחס סיכויים של 0.9/0.1 או 9, מה שנכתב בדרך כלל כ-9:1. כדי להחזיר יחסי סיכויים בחזרה להסתברויות, משתמשים בנוסחה $P=(O/(1+O))$ וזה הפיך לחלוטין. כלומר, הטרנספורמציה היא איזומורפיזם – מיפוי דו־כיווני הפיך. לפיכך, הסתברויות ויחסי סיכויים הם איזומורפיים, וניתן להשתמש באחד מהם או בשני בהתאם לנוחות.

לדוגמה, נוח יותר להשתמש ביחסי סיכויים כשמבצעים עדכונים בייסיאניים. נניח שאני מטיל קובייה בעלת שש פאות: אם יוצאת פאה כלשהי פרט ל-1, יש סיכוי של 10% לשמוע פעמון, אך אם יוצאת הפאה 1, יש סיכוי של 20% לשמוע את הפעמון. כעת אני מטיל את הקובייה ושומע פעמון. מהם הסיכויים שהפאה העליונה היא 1? ובכן, ההסתברות הפריורית היא 1:5 (המקבילה למספר הממשי 1/5 = 0.20) ויחס הנראות הוא 0.1 : 0.2 (המקביל למספר הממשי 2). אני יכול פשוט להכפיל את שניהם יחד כדי לקבל את

ההסתברות הפוסטריורית 2:5 (המקבילה למספר הממשי 2/5 או 0.40). לאחר מכן אני ממיר זאת חזרה להסתברות, אם אני רוצה, ומקבל (1.4/0.4) = 2/7 = ~29%.

לכן יחסי סיכויים נוחים יותר לניהול בעת עדכונים בייסיאניים – אם משתמשים בהסתברויות, חייבים לפרוס את משפט בייס בגרסתו המסובכת. אך הסתברויות נוחות יותר למענה על שאלות כגון: "אם אני מטיל קובייה בעלת שש פאות, מה הסיכוי לראות מספר מ-1 עד 4?" ניתן לחבר את ההסתברויות של 1/6 עבור כל פאה ולקבל 4/6, אך אינכם יכולים לחבר את יחסי הסיכויים של 0.2 עבור כל פאה ולקבל יחס סיכויים של 0.8.

למה אני מספר את כל זה? כדי להראות ש"יחסי סיכויים" הם דרך לגיטימית בדיוק כמו "הסתברויות" למיפוי אי־ודאויות על פני מספרים ממשיים. יחסי סיכויים נוחים יותר עבור פעולות מסוימות, ואילו הסתברויות נוחות יותר עבור אחרות. הוכחה מפורסמת שנקראת *משפט קוֹקְס* (בתוספת הרחבות ושיפורים שונים שלה) מראה שכל הדרכים לייצוג אי־ודאויות המצייתות למגבלות הנשמעות סבירות, הן איזומורפיות זו לזו.

למה זה משנה שיחסי סיכויים לגיטימיים בדיוק כמו הסתברויות? הסתברויות, כפי שהן נכתבות בדרך כלל, נעות בין 0 ל-1, וגם 0 וגם 1 נראים כמו כמויות שאמורות להיות נגישות בקלות – קל לראות זברה אחת או אפס חדי־קרן. אך כאשר מבצעים טרנספורמציה מהסתברויות ליחסי סיכויים, 0 עובר ל-0, אך 1 עובר לאינסוף חיובי. כעת אמת מוחלטת לא נראית כאילו היא אמורה להיות כל כך קלה להשגה.

ייצוג שהופך את ביצוע העדכונים הבייסיאניים לפשוט עוד יותר הוא הלוג־סיכויים (log odds) – כך המליץ א. ט. ג׳יינס לחשוב על הסתברויות. לדוגמה, נניח שההסתברות הפריורית של טענה היא 0.0001 – נתון זה מקביל ללוג־סיכויים של בערך 40− דציבלים. לאחר מכן, אתם רואים ראיה שנראית סבירה פי 100 אם הטענה נכונה מאשר אם היא שקרית. אלו הם 20 דציבלים של ראיות. לכן ההסתברות הפוסטריורית היא בערך 20dB− = 20dB + 40dB−,

כלומר, ההסתברות הפוסטריורית היא ~0.01%.

כאשר מבצעים טרנספורמציה מהסתברויות ללוג־סיכויים, 0 עובר לאינסוף שלילי ו־1 עובר לאינסוף חיובי. כעת הן ודאות אינסופית והן חוסר־סבירות אינסופית נראות קצת יותר מחוץ להישג יד.

בהסתברויות, 0.9999 ו־0.99999 נראים רחוקים זה מזה רק ב־0.00009. לפיכך נראה ש־0.502 רחוק הרבה יותר מ־0.503 מאשר המרחק שבין 0.9999 ל־0.99999. כדי להגיע להסתברות של 1 מהסתברות של 0.99999, נראה שצריך לעבור מרחק של 0.00001 בלבד.

אך כשמבצעים טרנספורמציה ליחסי סיכויים, 0.502 ו־0.503 עוברים ל־1.008 ול־1.012, בעוד ש־0.9999 ו־0.99999 עוברים ל־9,999 ו־99,999. וכשמבצעים טרנספורמציה ללוג־סיכויים, 0.502 ו־0.503 עוברים ל־0.03 דציבלים ו־0.05 דציבלים, אך 0.9999 ו־0.99999 עוברים ל־40 דציבלים ו־50 דציבלים.

כשעובדים בלוג־סיכויים, **המרחק בין כל שתי דרגות של אי־ודאות שווה לכמות הראיות שתצטרכו כדי לעבור מאחת לשנייה.** כלומר, הלוג־סיכויים מספקים לנו מדד טבעי של מרווחים בין דרגות שונות של ביטחון.

שימוש בלוג־סיכויים חושף את העובדה שהגעה לוודאות אינסופית דורשת ראיות חזקות באופן אינסופי, בדיוק כפי שאבסורדיות אינסופית דורשת ראיות־נגד חזקות באופן אינסופי.

יתרה מזאת, לכל מיני משפטים סטנדרטיים בהסתברות יש מקרים מיוחדים אם מנסים להציב בהם 1 או 0 (כמו למשל מה שקורה אם מנסים לבצע עדכון בייסיאני על תצפית שלה הקציתם הסתברות 0).

לכן אני מציע שזה אך הגיוני לומר ש־1 ו־0 אינם נמצאים בתוך טווח ההסתברויות, בדיוק כפי שאינסוף שלילי וחיובי — שאינם מצייתים לאקסיומות השדה — אינם נמצאים בתוך קבוצת המספרים הממשיים.

הסיבה העיקרית לכך שזה עשוי להכעיס תאורטיקנים של הסתברות, היא שנצטרך לגזור מחדש משפטים שהושגו בעבר על

ידי ההנחה שאנו יכולים לחשב הסתברות שולית מתוך הסתברות משותפת, על ידי חיבור כל הרכיבים כך שסכומם יהיה 1.

עם זאת, בעולם האמיתי, כשמטילים קובייה, אין לה באופן מילולי ודאות אינסופית שייצא מספר כלשהו בין 1 ל-6. הקובייה עשויה לנחות על הצד שלה, להיפגע ממטאור, או ש"שליטי האופל של המטריקס" עשויים להושיט יד ולכתוב "37" על פאה אחת.

אם הייתם יוצרים סמל קסום שיעמוד במקום "כל האפשרויות שלא שקלתי", אז הייתם יכולים לחשב הסתברות שולית של האירועים – כולל הסמל הקסום הזה – ולהגיע לסמל קסום "T" שמשמעותו ודאות אינסופית.

אך אני מעדיף לשאול האם ישנה דרך כלשהי לגזור את המשפט מבלי להשתמש בסמלים קסומים בעלי התנהגויות מיוחדות. זה יהיה אלגנטי יותר. בדיוק כפי שישנם מתמטיקאים שמסרבים להאמין בחוק השלישי הנמנע או בקבוצות אינסופיות, הייתי רוצה להיות תאורטיקן הסתברות שאינו מאמין בוודאות מוחלטת.

56

הרציונליות שלך היא העסק שלי

חלק מהתגובות לפרק *הגרלות לוטו: בזבוז של תקווה* (פרק 48) נזפו בי על כך שהעזתי לבקר החלטות של אחרים. אם מישהו אחר בוחר לקנות כרטיסי הגרלה, מי אני שאחלוק עליו? זהו מקרה פרטי של שאלה כללית יותר: מה זה ענייני אם מישהו אחר בוחר להאמין במה שנעים לו במקום במה שנכון? האם זו אינה הבחירה האישית של כל אחד מאיתנו אם אכפת לו מהאמת או לא?

תגובת־נגד מתבקשת היא: "למה *לך* אכפת אם *לי* אכפת אם למישהו *אחר* אכפת מהאמת?" אין זה עקבי במיוחד שפונקציית התועלת שלך תכלול איבר שלילי עבור פונקציית תועלת של אדם אחר, רק משום שזו מכילה איבר כלפי פונקציית התועלת של אדם שלישי. אך זו רק התנצחות, לא תשובה.

אז הנה התשובה שלי: אני מאמין שזה נכון וראוי עבורי, כבן אדם, לגלות עניין בעתיד, ובכיוון שאליו החברה האנושית תתפתח בעתיד. אחד מאותם תחומי עניין הוא השאיפה האנושית לאמת, שהתעצמה באיטיות לאורך הדורות (שכן המדע לא היה קיים מאז ומתמיד). אני חותר לחזק את השאיפה הזו עוד יותר, כבר בדור *הזה*. זוהי משאלת־לב שלי למען העתיד. שהרי כולנו שחקנים על גבי לוח המשחק העצום הזה, בין אם אנו מקבלים על עצמנו את האחריות הזו ובין אם לאו.

וזה הופך את הרציונליות *שלך* לעסק *שלי*.

האם זהו רעיון מסוכן? כן, ולא רק "מסוכן" במובן האופנתי והמרדני של המילה. אנשים נשרפו למוות משום שאיזה כומר החליט שהם אינם חושבים כפי שהם אמורים לחשוב. להחליט להעלות אנשים על המוקד משום שאינם "חושבים כראוי" – זוהי צורת חשיבה מעוררת חלחלה, לא? לא הייתם רוצים שאנשים יחשבו כך, זה הרי *דוחה*. אנשים שחושבים כך... ובכן, נצטרך לעשות משהו בקשר אליהם...

אני מסכים! הנה ההצעה שלי: הבה נצא נגד רעיונות רעים, אך לא נעלה על המוקד את האנשים האוחזים בהם.

ההיקש הלוגי שאנו מעוניינים להימנע ממנו מתנהל כך: "אני חושב שסוזי אמרה דבר רע. *לכן*, צריך לשרוף את סוזי." יש המנסים להימנע מהיקש זה על ידי הקביעה שזה פשוט פסול לחשוב שסוזי אמרה דבר רע. איש אינו רשאי לשפוט את זולתו, לעולם. כל מי ששופט מבצע עוול נורא, ויש להוקיע אותו בפומבי על כך.

כשלעצמי, אני כופר במילה "לכן". ההיקש שלי מתנהל כך: "אני חושב שסוזי אמרה משהו שגוי. לכן, אטען נגד דבריה, אך לא אעלה אותה באש ולא אנסה למנוע ממנה לדבר באמצעות אלימות או חקיקה..."

כולנו שחקנים על לוח המשחק העצום הזה, ואחת משאיפותיי למען העתיד היא להפוך את המשחק להוגן. הרעיון המנוגד־לאינטואיציה העומד בבסיס המדע הוא שעל מחלוקות עובדתיות יש להילחם באמצעות ניסויים ומתמטיקה, לא באלימות וצווים. ניתן להרחיב את הרעיון המדהים הזה אף מעבר למדע, לכדי מאבק הוגן על העתיד כולו. עליך לנצח באמצעות שכנוע של אנשים, ואסור להתיר לך לשרוף אותם. זהו אחד מעקרונות הרציונליות, לו נשבעתי אמונים.

אנשים המטיפים לרלטיביזם או לאנוכיות אינם נראים לי רלטיביסטים או אנוכיים באמת. אילו היו רלטיביסטים באמת, הם לא היו שופטים איש. אילו היו אנוכיים באמת, הם היו עוסקים בעשיית כסף במקום להתווכח בלהט עם זרים. במקום זאת, הם בחרו בצד של הרלטיביזם, שמטרתו על לוח המשחק העצום היא למנוע

מן השחקנים – *כל* השחקנים – לבצע סוגים מסוימים של שיפוט. או שבחרו בצד של האנוכיות, שמטרתו להפוך את *כל* השחקנים לאנוכיים. ואז הם משחקים את המשחק, בהוגנות או בחוסר הוגנות, כמיטב חוכמתם.

אם אכן קיימים רלטיביסטים או אנוכיים אמיתיים, הרי שאיננו שומעים אותם. הם נותרים שקטים, לא־שחקנים.

אינני מסוגל להישאר אדיש לאופן שבו אתם חושבים – בדיוק כפי שאינני יכול שלא לראות את היקום – משום שבכל פעם שבן אדם מפנה עורף לאמת, סיפורו המתהווה של המין האנושי הופך למעט אפל יותר. במקרים רבים, זוהי אפלולית קטנה בלבד (הרי לא *תמיד* מישהו נפגע בסופו של דבר). שקר עצמי, בפרטיות מחשבותיכם, אינו מטיל צל על ההיסטוריה האנושית באותה העוצמה כמו הפצת שקרים פומביים או העלאת אנשים על המוקד. ובכל זאת, יש בי חלק שאינו יכול שלא להתאבל. וכל עוד אינני מנסה להעלות אתכם על המוקד – אלא רק להתווכח עם הרעיונות שלכם – אני מאמין שזה נכון וראוי שיהיה לי, כבן אדם, אכפת מחבריי בני האדם. גם זוהי עמדה עליה בכוונתי להגן, עכשיו ובעתיד.

חלק ו'

רציונליות ופוליטיקה

...אנו עוברים למים עכורים יותר, עם אסופת המאמרים "רציונליות ופוליטיקה". פוליטיקה ארצית, כפי שהיא נדונה בפי פרשני טלוויזיה, ידועה לשמצה בשיח הזועם והעקר המאפיין אותה. על פניו, יש בכך משהו מפתיע. מדוע אנו לוקחים מחלוקות פוליטיות באופן אישי כל כך, אפילו כאשר המנגנונים וההשפעות של הפוליטיקה רחוקים מאיתנו כל כך בזמן או במרחב? יתרה מזאת, מדוע איננו נעשים זהירים ויסודיים *יותר* בבחינת הראיות דווקא כאשר אנו עוסקים בסוגיות כה כבדות משקל?

— רוב בנסינגר,
רציונליות: מבוא

57

פוליטיקה היא אויב התבונה

משהו משתבש לאנשים בראש כשהם מדברים על פוליטיקה. הסיבות האבולוציוניות לכך כה ברורות עד שראוי להתעכב עליהן מעט: בסביבה הקדמונית, הפוליטיקה הייתה עניין של חיים ומוות. ושל יחסי מין, ועושר, ובני ברית, ומוניטין... כשאתם נכנסים כיום לוויכוח על השאלה האם "אנחנו" צריכים להעלות את שכר המינימום, אתם מפעילים מנגנונים שנועדו לסביבה קדמונית שבה להימצא בצד הלא נכון של הוויכוח עלול היה לעלות לכם בחייכם. לעומת זאת, להיות בצד *הנכון* של הוויכוח יכול היה לאפשר *לכם* להרוג את היריב השנוא עליכם!

אם ברצונכם להסביר משהו על מדע או רציונליות, עצתי היא לא לבחור בדוגמה מהפוליטיקה *העכשווית*, כל עוד ביכולתכם להימנע מכך. אם הטענה שלכם פוליטית במהותה, דברו על לואי ה-16 והמהפכה הצרפתית. פוליטיקה היא תחום חשוב שבו על כל אחד מאיתנו להפעיל חשיבה רציונליות - אך זהו תחום נורא כדי *ללמוד* בו רציונליות, או לדון בה, אלא אם כן כל המשתתפים בדיון כבר רציונליים.

פוליטיקה היא המשך המלחמה באמצעים אחרים. הטיעונים משמשים בה כחיילים. ברגע שאתם יודעים באיזה צד אתם, עליכם לתמוך בכל הטיעונים של אותו צד, ולתקוף את כל הטיעונים שנראה כי הם פועלים לטובת צד האויב. אחרת, יהיה זה כאילו נעצתם סכין

בגב החיילים שלכם – הענקתם סיוע ישירות לאויב. אנשים השוקלים סוגיה מכל צדדיה בכובד ראש ובשום שכל בחייהם המקצועיים כמדענים, מסוגלים להפוך כהרף עין לזומבים המדקלמים סיסמאות ברגע שמוצגת בפניהם עמדה המזוהה עם *הכחולים או הירוקים*.*

בבינה מלאכותית, ובמיוחד בתחום של הסקה לא־מונוטונית,** רווחת הצגת בעיות כגון: "כל המתנחלים הם ימנים. כל תומכי 'שלום עכשיו' הם שמאלנים. יובל הוא מתנחל שפעיל ב'שלום עכשיו'. האם יובל שמאלני?"

מה לכל הרוחות המטרה בבחירת הדוגמה הספציפית הזאת? לעורר רגשות פוליטיים אצל הקוראים ולהסיח את דעתם מהשאלה המרכזית? לגרום לימנים להרגיש לא רצויים בקורסים לבינה מלאכותית ולהרתיע אותם מלהיכנס לתחום? (ולא, אני לא ימני. או שמאלני.)

מדוע שמישהו יבחר בדוגמה כה *מסיחת־דעת* כדי להדגים הסקה לא־מונוטונית? כנראה משום שהמחבר פשוט לא יכול היה להתאפק מלשגר עקיצה טובה ועסיסית *בירוקים* השנואים. התחושה של הנחתת חבטה הגונה היא כל כך *טובה*... זה כמו לנסות לעמוד בפני עוגיית שוקולד.

אלא שבדומה לעוגיות שוקולד, לא כל מה שטעים בהכרח גם טוב לכם.

אין בכוונתי לטעון שעלינו להיות א-פוליטיים, או אפילו שעלינו לאמץ את האידיאל של ויקיפדיה לנקודת מבט ניטרלית. אבל נסו להימנע מאותן עקיצות עסיסיות אם אתם יכולים. אם הנושא שלכם קשור באופן לגיטימי לניסיונות לאסור לימודי אבולוציה בבתי הספר, אז קדימה, דברו על זה – אבל אל תטילו את האשמה במפורש על כל צד ימין של המפה הפוליטית. ייתכן שחלק מהקוראים שלכם הם ימנים, והם עשויים להרגיש שהבעיה נעוצה בכמה פוליטיקאים סוררים ולא

---

* ראו פרק 12.

** *הסקה לא־מונוטונית* (non-monotonic reasoning) היא הסקה שבה הוספת מידע חדש יכולה לבטל מסקנות שהוסקו קודם לכן.

במחנה הפוליטי כולו. בדיוק כפי שקורה עם כללי הניטרליות של ויקיפדיה, לא משנה אם (לדעתכם) הימנים באמת אשמים. זה פשוט הרבה יותר בריא הקהילה לדון בנושאים הללו מבלי לערב פוליטיקה של *כחולים נגד ירוקים*.

58

דיונים פוליטיים לא אמורים להרגיש חד-צדדיים

רובין הַנסון הציע להקים חנויות שבהן יימכרו לציבור מוצרים אסורים. ישנם כמה טיעונים מצוינים בעד מדיניות כזו: הזכות המוּלדת לחירות הפרט, התמריץ המקצועי של רגולטורים לאסור על *הכל*, והעובדה שמחוקקים חשופים להטיות לא פחות מאנשים פרטיים. ובכל זאת (השבתי), אֵם לחמישה – אשה ענייה, הגונה, אך לא משכילה במיוחד – עומדת להיכנס לחנויות הללו, לקנות "משקה חומצה גופרתית של ד"ר נחשוזאורוס" עבור דלקת המפרקים שלה, ולמות כשהיא מותירה את יתומיה להתייפח בטלוויזיה במהדורת החדשות.

בסך הכל ציינתי תצפית עובדתית פשוטה. מדוע היו מי שחשבו שזהו טיעון בעד רגולציה?

בשאלות הנוגעות לעובדות פשוטות (למשל, האם החיים על פני כדור הארץ התהוו באמצעות ברירה טבעית), ישנה ציפייה לגיטימית לכך שהדיון יהיה קרב חד־צדדי. העובדות עצמן הן או כך או אחרת, ומה שקרוי "מאזן הראיות" אמור לשקף זאת. ואכן, תחת ההגדרה הבייסיאנית של ראיה, "ראיה חזקה" היא בדיוק סוג הראיות שאנו מצפים למצוא רק בצד אחד של הדיון.

אך אין שום סיבה שפעולות מורכבות בעלות השלכות רבות יפגינו את מאפיין החד־צדדיות הזה. מדוע נדמה שאנשים מצפים שדיונים

על *פוליטיקה* יהיו חד־צדדיים?

פוליטיקה היא אויב התבונה. טיעונים הם חיילים. ברגע שאתם יודעים באיזה צד אתם נמצאים, עליכם לתמוך בכל הטיעונים של אותו צד ולתקוף את כל הטיעונים שנראה כי הם פועלים לטובת צד האויב. אחרת, יהיה זה כאילו נעצתם סכין בגב החיילים שלכם. אם תפעלו לפי הדפוס הזה, גם דיונים פוליטיים ייראו לכם חד־צדדיים – העלויות והחסרונות של המדיניות המועדפת עליכם הם חיילי אויב, ויש לתקוף אותם בכל האמצעים הדרושים.

יש להיות מודעים גם לכשל מחשבתי נוסף – המחשבה כי זוהי *חוכמה עמוקה* למצוא פשרה המאזנת באופן מושלם בין שתי עמדות המדיניות הזוכות למרב זמן המסך. ייתכן בהחלט שלמדיניות מסוימת יהיו עלויות או תועלות *שנוטות בחדות* לצד אחד. אילו שאלות פוליטיות לא היו נוטות בבירור לכאן או לכאן, לא היתה לנו שום דרך להכריע ביניהן. אך קיימת גם נטייה אנושית להכחיש את כל העלויות של מדיניות מועדפת, או להכחיש כל תועלת של מדיניות שאינה מועדפת. כתוצאה מכך, אנשים ייטו לחשוב שפשרות פוליטיות הינן חד־צדדיות ומוטות הרבה יותר מכפי שהן באמת.

אם תתירו להקים חנויות המוכרות מוצרים שאחרת היו נאסרים, אשה ענייה והגונה כלשהי – אם לחמישה שאינה משכילה במיוחד – תקנה משהו שיהרוג אותה. זוהי תחזית לגבי השלכה עובדתית, וכשאלָה עובדתית היא נראית די פשוטה. אדם שפוי אמור להודות בנכונותה בקלות, ללא קשר לעמדתו הפוליטית. אתם עשויים גם לחשוב שהפיכת מוצרים לבלתי חוקיים רק מייקרת אותם, שרגולטורים ינצלו לרעה את כוחם, או שחירות הפרט של אותה אישה גוברת על הרצון שלכם להתערב בחייה. אך כעניין של עובדה פשוטה, היא עדיין עומדת למות.

אנו חיים ביקום לא הוגן. בדומה לכל הפרימטים, לבני אדם יש תגובה שלילית חריפה כלפי מה שנתפס כחוסר הוגנות, ולכן העובדה הזו מעוררת בנו מתח. ישנן שתי שיטות נפוצות להתמודדות עם הדיסוננס הקוגניטיבי שנוצר: ראשית, אדם יכול לשנות את השקפתו

על העובדות – להכחיש שהאירועים הבלתי הוגנים התרחשו, או לערוך את ההיסטוריה כדי לגרום לה להיראות הוגנת. (פתרון זה מתווך על ידי היוריסטיקת הרגש וכשל "העולם הצודק".) שנית, אדם יכול לשנות את המוסר שלו – להכחיש שהאירועים הללו אינם הוגנים מלכתחילה.

ישנם ליברטריאנים שעשויים לומר שאם אדם נכנס ל"חנות מוצרים אסורים", חלף על פני תוויות אזהרה ברורות האומרות "מוצרים בחנות זו עלולים להרוג אותך", וקנה משהו שהרג אותו, הרי שזו אשמתו וזה מגיע לו. אילו זו הייתה אמת מוסרית, לא היה *שום חיסרון* בקיומן של חנויות למכירת מוצרים אסורים. זו לא הייתה סתם *תועלת נטו*, אלא יתרון *חד־צדדי* ללא חסרונות.

אחרים טוענים שניתן להכשיר רגולטורים לפעול באופן רציונלי ובהלימה לאינטרסים של הצרכן. אילו אלו היו העובדות לאשורן, הרי שלא היה *שום חיסרון* ברגולציה (לפי השקפתם המוסרית).

בין אם נרצה בכך ובין אם לאו, קיימת הגרלת־לוטו של אינטליגנציה בשלב הלידה (אף שזהו אחד המקרים שבהם חוסר ההגינות של היקום הוא כה קיצוני, עד שרבים בוחרים פשוט להכחיש את העובדות). הראיות המחקריות לקיומו של רכיב גנטי טהור בשיעור של 0.6-0.8 הן מכריעות. אך אפילו אם תכחישו זאת, הרי שגם את החינוך ההורי או את בתי הספר המוקדמים שלכם לא אתם בחרתם.

חונכתי להאמין שהכחשת המציאות היא *עוול מוסרי*. אילו הייתי מתמסר לאופטימיות של משאלות־לב לגבי הסיכוי ש"משקה חומצה גופרתית" יועיל לי, הייתי עושה משהו *שהוזהרתי* מפניו וחונכתי לראות כפסול. יש אנשים שנולדים לסביבה – ולא נדון בגנים שלהם, כי החלק הזה אינו הוגן אפילו יותר – שבה רופא האליל המקומי אומר להם שזה *נכון* להאמין *ולא נכון* להיות ספקן. מתוך כוונות טובות לחלוטין, הם פועלים לפי העצה הזו ומתים. בניגוד אליכם, הם לא חונכו להאמין שאנשים נושאים באחריות לבחירותיהם האישיות ללכת בעקבות החברה. האם אתם באמת חושבים שאתם כל כך חכמים, עד שהייתם הופכים לספקנים מדעיים כהלכה גם אילו

נולדתם בשנת 500 לספירה? כן, ישנה הגרלת־לוטו בלידה, ולא משנה במה אתם מאמינים בנוגע לגֵנים.

לומר: "אנשים שקונים מוצרים מסוכנים ראויים להיפגע!" לא באמת מעיד על אטימות־לב. זוהי פשוט דרך לסרב לחיות ביקום לא הוגן. אטימות־לב אמיתית היא לומר: "כן, חומצה גופרתית היא מוות נורא ומכאיב, ולא, לאמא ההיא לחמישה ילדים לא הגיע למות, אבל אנחנו נשאיר את החנויות פתוחות בכל זאת כי ערכנו את חישוב העלות־תועלת הזה." אתם יכולים לדמיין פוליטיקאי אומר דבר כזה? גם אני לא. אך ככל שלכלכלנים יש כוח להשפיע על המדיניות, אולי זה יעזור אם הם יוכלו לחשוב כך בינם לבין עצמם – ואולי אפילו לומר זאת במאמרים בכתבי עת, כשהדברים עטופים בערפול רב־הברתי כהלכה, כך שהתקשורת לא תוכל לצטט אותם.

אינני חושב שמוות כתוצאה מהחלטה טיפשית הוא סיבה למסיבה. אני מחשיב זאת כטרגדיה. עם זאת, לא תמיד זוהי עזרה להציל אנשים מההשלכות של מעשיהם. אך אני מציב קו אדום מוסרי בעונש מוות: אם אתה מת, אינך יכול ללמוד מהטעויות שלך.

לרוע המזל, היקום לא מסכים איתי. נראה של מי תהיה המילה האחרונה כשהכל ייגמר.

59

מאזני הצדק, פנקס הרציונליות

אלת הצדק מתוארת לרוב כדמות האוחזת במאזניים. למאזניים ישנה תכונה ייחודית: כל משקל המוריד כף אחת מטה, דוחף בהכרח את הכף השניה מעלה. הדבר הופך את העניינים לנוחים מאוד וקלים להבנה. עם זאת, בדרך כלל מדובר גם בעיוות גס של המציאות.

באופן שבו בני אדם מנהלים דיונים, קיימת נטייה טבעית להתייחס לדיון כאל סוג של קרב – כהמשך המלחמה באמצעים אחרים, או כענף ספורט. ובספורט, כל שעליכם לעשות הוא לעקוב אחר מספר הנקודות שצברה כל קבוצה. ישנם רק שני צדדים, וכל נקודה שנרשמת לחובת צד אחד היא בהכרח נקודה לזכותו של הצד השני. כל אחד בקהל מנהל בראשו מעקב שוטף אחר הנקודות שכל נואם צובר נגד יריבו. בסופו של העימות, הנואם שצבר יותר נקודות הוא כמובן המנצח. לכן, כל מה שאמר אותו נואם הוא בהכרח נכון, וכל מה שאמר המפסיד בהכרח שגוי.

המחקר "The Affect Heuristic in Judgments of Risks and Benefits" ("היוריסטיקת הרגש בשיפוט סיכונים ויתרונות") בחן האם נבדקים ערבבו בין השיפוטים שלהם לגבי היתרונות האפשריות של טכנולוגיה מסוימת (למשל, אנרגיה גרעינית) לבין הסיכונים האפשריים שלה, לכדי תחושה כללית אחת – טובה או רעה – כלפי אותה טכנולוגיה. נניח שאני מספר לכם תחילה שסוג מסוים של כור גרעיני מייצר פחות פסולת גרעינית מסוגי כורים אחרים. אך

לאחר מכן אני מציין שהכור הזה יציב פחות מתכנונים מתחרים, וכי סכנת התכת הליבה שלו גבוהה יותר אם דַי דברים ישתבשו ברזמנית.

אם קיימת סבירות גבוהה יותר לאירוע התכת ליבה, הדבר מרגיש כמו "נקודה" לחובת הכור, או "נקודה" לחובת מי שמעדיף אותו על פני סוגי כורים אחרים. מנגד, אם הכור מייצר פחות פסולת, הרי שזוהי "נקודה" לזכות הכור, או "נקודה" לזכות בנייתו. אם כך, האם שתי העובדות הללו מנוגדות זו לזו? לא. לא בעולם האמיתי. אמנם צדדים שונים באותו עימות עשויים לצטט את שתי העובדות הללו, אך הן נפרדות לחלוטין מבחינה לוגית. העובדות אינן יודעות באיזה צד הן נמצאות.

אם זוהי עובדה פיזיקלית שתכנון של כור מסוים הוא "בטוח באופן פסיבי" (הוא לא יגיע למצב של קריטיות־על גם אם מערכות הקירור ההיקפיות וכדומה יקרסו), אין פירוש הדבר שהכור ייצר בהכרח פחות פסולת או יפיק חשמל בעלות נמוכה יותר. כל הדברים הללו הם חיוביים, אך הם אינם אותו דבר חיובי. כמות הפסולת שמייצר הכור נובעת ממאפיינים מסוימים שלו, ואילו תכונות פיזיקליות אחרות של הכור הן אלו שהופכות את התגובה הגרעינית ליציבה פחות. אפילו אם חלק מאותן תכונות תכנוניות מעורבות בשני המקרים, עליכם לבחון בנפרד את ההסתברות להתכת הליבה ואת כמות הפסולת השנתית הצפויה. אלו הן שתי שאלות פיזיקליות שונות, בעלות שתי תשובות עובדתיות נפרדות.

ואולם, מחקרים כדוגמת זה שהוזכר לעיל מראים כי אנשים נוטים לשפוט טכנולוגיות – ובעיות רבות אחרות – על סמך תחושה כללית, חיובית או שלילית. אם תַּספרו לאנשים שתכנון של כור מסוים מייצר פחות פסולת, הם יעריכו את ההסתברות שלו להתכת ליבה כנמוכה יותר. המשמעות היא קבלת תשובה שגויה לשאלות פיזיקליות בעלות תשובות עובדתיות מוחלטות. מדוע? משום שערבבתם בין שאלות נפרדות מבחינה לוגית – התייחסתם לעובדות כאל חיילים אנושיים בשני צידי המתרס של מלחמה, מתוך מחשבה שניתן להשתמש בכל חייל בצד אחד כדי להילחם בכל חייל בצד השני.

המאזניים אינם סמל פסול לחלוטין עבור אלת הצדק, כל עוד היא חוקרת שאלה עובדתית גרידא של אשמה או חפות. או שראובן הרג את שמעון, או שלא. אנו למדים (מפי א. ט. גֶ׳יינס) שכל ראיה בייסיאנית מורכבת מזרימת הסתברות בין השערות. אין דבר כזה ראיה ש״תומכת״ בהשערה יחידה או ״סותרת״ אותה במנותק. התמיכה קיימת רק במידה שבה השערות אחרות מצליחות פחות או יותר. לכן, כל עוד אלת הצדק חוקרת שאלה עובדתית יחידה ומובהקת בעלת מרחב תשובות בינארי, מאזניים יהיו כלי מתאים. אך אם יוסטיציה* נדרשת לשקול סוגיה מורכבת יותר, יהיה עליה לוותר על מאזניה או לוותר על חרבה.

לא כל הטיעונים מתמצים ב״למעלה״ או ״למטה״ פשוטים. גבירת הרציונליות נושאת עמה פנקס, ובו היא רושמת את כל העובדות שאינן עומדות לצדו של איש.

---

1. Melissa L. Finucane et al., "The Affect Heuristic in Judgments of Risks and Benefits," *Journal of Behavioral Decision Making* 13, no. 1 (2000): 1–17.

* *יוסטיציה* (Justitia) היא אלת הצדק במיתולוגיה הרומית.

60

הטיית ההתאמה

הטיית ההתאמה היא הנטייה להסיק על אופיו הפנימי והקבוע של אדם מתוך התנהגויות שניתן להסבירן במלואן על ידי הנסיבות שבהן הן התרחשו.

— דן גילברט ופטריק מַאלון[1]

אנו נוטים לראות קשר ישיר מדי בין מעשיהם של אחרים לבין האישיות שלהם. כשאתם רואים מישהו בועט במכונת חטיפים אוטומטית ללא כל סיבה נראית לעין, אתם מניחים שהוא "טיפוס עצבני". אבל כשאתם בעצמכם בועטים במכונה, הרי זה מפני שהאוטובוס איחר, הרכבת הקדימה, הדוח שלכם כבר היה צריך להיות מוגש, ועכשיו מכונת החטיפים הארורה בלעה לכם את הכסף של ארוחת הצהריים זה היום השני ברציפות. *כל אדם שפוי היה בועט במכונה במצב כזה.*

את המעשים של עצמנו אנו מייחסים לנסיבות, ורואים בהתנהגות שלנו תגובה נורמלית לחלוטין למציאות. אבל כשמישהו אחר בועט במכונת חטיפים, אנו לא רואים את ההיסטוריה שלו מרחפת מאחוריו באוויר. אנו רואים רק את הבעיטה, במנותק מכל הקשר, ומסיקים שמדובר בטיפוס עצבני מטבעו — שהרי הוא איבד את עשתונותיו ללא כל סיבה.

אך בואו נבחן לרגע את ההסתברות האפריורית. יש בעולם

הרבה יותר אוטובוסים מאחֲרים מאשר מוטנטים שנולדו עם רמות עצבנות חריגות הגורמות להם לבעוט ספונטנית במכונות חטיפים. למעשה, האדם הממוצע הוא אכן מוטנט. אם זכרוני אינו מטעה אותי, לאדם הממוצע יש בין שתיים לעשר מוטציות סומטיות,* אך הסיכוי שמיקום *ספציפי* ב-DNA יושפע מכך הוא קלוש ביותר. באופן דומה, סביר להניח ששום תכונת אופי של אדם כלשהו אינה רחוקה במיוחד מהממוצע. לטעון אחרת פירושו להעמיס על עצמכם נטל הסתברותי כבד.

אפילו כאשר אנשים מקבלים מידע מפורש על הגורמים הנסיבתיים להתנהגות, הם עדיין נוטים לייחס משקל יתר להתנהגות עצמה ואינם מתַקנים את הרושם הראשוני שלהם. כשאומרים לנבדקים שנואם מסוים *קיבל באקראי* משימה להציג עמדה בעד או נגד הפלות, הם עדיין נוטים להניח שהנאום משקף את נטיותיו האמיתיות.[2]

זה מרגיש אינטואיטיבי למדי להסביר גשם באמצעות "רוחות־מים", להסביר אש באמצעות "פְלוֹגִיסְטוֹן" הנפלט מחומר בוער, או להסביר את ההשפעה המרדימה של תרופה בכך שהיא מכילה "סגולה מרדימה". המציאות, בדרך כלל, מבוססת על מנגנונים מורכבים בהרבה: מחזור של התאדות והתעבות העומד בבסיס הגשם, תהליך של בעירה מחמצנת בבסיס האש, ואינטראקציות כימיות עם מערכת העצבים במקרה של חומרי הרדמה. אך מנגנון נתפס במוחנו כמסובך יותר מאשר "מהות" פשוטה. הוא דורש מאמץ מחשבתי רב יותר, וזמין לנו פחות מבחינה קוגניטיבית. לכן, כשמישהו בועט במכונת חטיפים, אנו מניחים שיש לו נטייה מולדת לבעיטה־במכונות־חטיפים.

אלא אם כן ה"מישהו" שבועט במכונה הוא אנחנו — ובמקרה כזה אנחנו הרי מתנהגים בצורה נורמלית לחלוטין בהתחשב בנסיבות. הרי ברור שכל אחד אחר היה נוהג באופן דומה. למעשה, אנו נוטים להעריך יתר על המידה את הסיכוי שאחרים יגיבו בדיוק כמונו — תופעה המוכרת

* *מוטציה סומטית* (somatic mutation) היא מוטציה בתאי גוף רגילים, ומשפיעה רק על הגוף של אותו האדם ועל התפקוד שלו – זאת בניגוד למוטציה בתאי מין, שעוברת הלאה לדורות הבאים.

כ"אפקט הקונצנזוס השגוי". סטודנטים ששותים אלכוהול מפריזים מאוד בהערכת שיעור חבריהם לספסל הלימודים ששותים גם הם, בעוד שסטודנטים שנמנעים משתיה מעריכים את שיעור הסטודנטים ששותים בחסר ניכר. "טעות הייחוס הבסיסית" מתייחסת לנטיה האנושית לתלות את התנהגותם של אחרים באופי שלהם, אך לא לעשות זאת באשר להתנהגותנו שלנו.

כדי להבין מדוע אנשים פועלים כפי שהם פועלים, עלינו להבין תחילה שכל אחד רואה את עצמו כמי שמתנהג בצורה נורמלית. אל תחפשו נטיית אופי חריגה או מוטציה מולדת שתסביר באופן ישיר את מעשיהם. שאלו, במקום זאת, איך נראית הסיטואציה מהזווית שלהם. נכון, לאנשים יש תכונות אופי – אבל פשוט *אין בנמצא* מספיק תכונות מולדות וחריגות כדי להסביר באופן ישיר את כל ההתנהגויות שאתם רואים סביבכם.

נניח שהייתי נותן לכם שלט עם שני כפתורים, אדום וירוק. הכפתור האדום משמיד את העולם, והכפתור הירוק מוציא את הכפתור האדום מכלל פעולה. על איזה כפתור הייתם לוחצים? על הירוק. כל מי שנותן תשובה אחרת כנראה מסבך את השאלה ללא צורך.

ואף על פי כן, אנשים שואלים אותי לפעמים מדוע אני רוצה להציל את העולם, כאילו ברור שעברתי ילדות עשוקה או משהו כזה. כלומר, זו נראית כמו בחירה די מתבקשת... אם ככה אתם תופסים את המצב.

ייתכן שיש לי תפיסות עולם חריגות שדורשות הסבר – מדוע אני מאמין בדברים כאלה, בניגוד לרוב האנשים? – אך בהינתן האמונות הללו, *התגובה* שלי אינה כזו שמצריכה הסבר יוצא דופן. אולי אני קורבן של אפקט הקונצנזוס השגוי. אולי אני מעריך יתר על המידה את שיעור האנשים שהיו לוחצים על הכפתור הירוק אילו ראו את המצב כמוני. אבל אתם יודעים מה? עדיין הייתי מהמר שיהיה מדובר לכל הפחות *במיעוט ניכר*.

מרבית האנשים תופסים את עצמם כנורמליים לחלוטין, מנקודת מבטם. אפילו אנשים שאתם שונאים, אנשים שעושים דברים נוראיים, אינם מוטנטים נדירים. אין כאן שום צורך במוטציה, למרבה הצער.

ברגע שמבינים זאת, אפשר סוף סוף להפסיק להיות מופתעים מהדברים שאנשים עושים.

---

1. Daniel T. Gilbert and Patrick S. Malone, "The Correspondence Bias," *Psychological Bulletin* 117, no. 1 (1995): 21–38, http://wjh-www.harvard.edu/~dtg/Gilbert%20&%20Malone%20%28CORRESPONDENCE%20BIAS%29.pdf.

2. Edward E. Jones and Victor A. Harris, "The Attribution of Attitudes," *Journal of Experimental Social Psychology* 3 (1967): 1–24, http://www.radford.edu/~jaspelme/443/spring-2007/Articles/Jones_n_Harris_1967.pdf.

61

האם האויבים שלכם רעים מטבעם?

אנו נוטים לקשור באופן ישיר מדי בין מעשיהם של אחרים לבין תכונות האופי שלהם. אנו מסיקים על קיומן של תכונות אופי חריגות התואמות בדיוק להתנהגות חריגה, במקום לברר מהן הנסיבות – האמיתיות או המדומות – שעשויות להסביר התנהגות זו. למעשה, אנו מניחים את קיומם של מוטנטים.

כשמישהו ממש *מקומם* אותנו – כשהוא מבצע מעשה שפסול בעינינו (בצדק או שלא בצדק) – הרי שלהתרשמותי, הטיית ההתאמה מכפילה את עוצמתה. נראה כי קיימת נטייה *עזה* לייחס מעשים רעים לאופיו המוטנטי והמרושע של האויב. עלינו לשאול את עצמנו מה האויב עשוי לחשוב על הנסיבות, באופן שיכול להפוך את התנהגותו לתמוהה פחות בעינינו – לא מתוך עמדה מוסרית, אלא מתוך שיקול טהור של הסתברות אפריורית. הדבר יאפשר לנו להניח קיומן של תכונות אופי פחות חריגות, ובכך לשאת בנטל הסתברותי קטן יותר.

ב-11 בספטמבר 2001 חטפו תשעה־עשר גברים מוסלמים ארבעה מטוסי נוסעים במשימת התאבדות, שמטרתה המוצהרת הייתה לפגוע בארצות הברית. מדוע לדעתכם הם עשו זאת? האם משום שתפסו את ארצות הברית כסמל של חופש עבור העולם כולו, אך נולדו עם אופי מוטנטי שגרם להם פשוט לשנוא חופש?

במציאות, מרבית בני האדם לא מספרים את סיפור חייהם מנקודת המבט של הנבל. כל אחד הוא הגיבור בסיפור של עצמו. הסיפור של האויב, כפי שהוא נתפס בעיניו שלו, *אינו מציג אותו באור שלילי*. אם תנסו לייחס לאויב מניעים המציגים אותו באור שלילי, הרי שבסופו של דבר תטעו לחלוטין לגבי מה שבאמת מתחולל במוחו.

אך הפוליטיקה היא אויב התבונה. ויכוח הוא מלחמה, וטיעונים הם חיילים. ברגע שאתם יודעים באיזה צד אתם נמצאים, עליכם לתמוך בכל הטיעונים של אותו צד, ולתקוף את כל הטיעונים שנראה כי הם פועלים לטובת צד האויב. אחרת, יהיה זה כאילו נעצתם סכין בגב החיילים שלכם.

אילו לאויב אכן הייתה תכונת אופי רעה, היה זה טיעון לטובת הצד שלכם. *וכל* טיעון שמשרת את הצד שלכם חייב לזכות בתמיכה, מופרך ככל שיהיה – אחרת אתם מרפים את הלחץ בנקודה כלשהי לאורך קו החזית. כל אחד מנסה להתעלות על רעהו בגינוי המחנה היריב, ואיש אינו מעז להביע התנגדות. בתוך זמן קצר מצמיח האויב קרניים, כנפי עטלף, פה יורק אש וניבים נוטפים ארס חומצי. אם תכחישו ולו פרט אחד מהתיאור הזה, מטעמים עובדתיים בלבד, הרי שאתם טוענים לטובת האויב. אתם בוגדים. מעטים מאוד יבינו שאינכם מגנים על האויב, אלא פשוט מגנים על האמת.

אילו רק מוטנטים היו מסוגלים לבצע דברים מפלצתיים, ההיסטוריה של המין האנושי הייתה נראית שונה לחלוטין. מוטנטים, אחרי ככלות הכל, הם תופעה נדירה.

ואולי הפחד הוא שהבנה תגרור בעקבותיה מחילה. הרי הרבה יותר פשוט לירות במוטנטים מרושעים. הרבה יותר קל להסתער עם קריאת הקרב "מותו, נבלים ארורים!" מאשר עם הקריאה "מותו, אנשים שיכלו להיות בדיוק כמוני אילו רק גדלו בסביבה שונה!" אתם עלולים לחוש אשמה על הריגת אנשים *שאינם* רוע צרוף.

הדבר נראה לי דומה לכמיהה המושרשת לדיון פוליטי חד-צדדי, שבו למדיניות הטובה ביותר *אין* חסרונות. אם צבא חוצה את הגבול או שפסיכופת מסתער עליכם עם סכין, ברירות המחדל הן (א) להגן על

עצמכם או (ב) להיכנע ולמות. אם תגנו על עצמכם, ייתכן שתיאלצו להרוג. אם אתם הורגים מישהו שיכול היה, במציאות חלופית, להיות חבר שלכם – זוהי טרגדיה. וזו *אכן* טרגדיה. גם האפשרות השנייה, להיכנע ולמות, היא טרגדיה. מי אמר שחייבת להיות אפשרות לא־טרגית? מי אמר שלמדיניות הטובה ביותר לא יכולים להיות חסרונות? אם מישהו חייב למות, מוטב שיהיה זה התוקפן, כדי להרתיע מפני אלימות עתידית ובכך למזער את מספר הקורבנות הכולל.

גם אם לאויב יש אופי ממוצע, והוא פועל מתוך תפיסת מציאות שבה אלימות היא תגובה אנושית טיפוסית, אין פירוש הדבר שאמונותיו מדויקות מבחינה עובדתית. אין פירוש הדבר שהן מוצדקות. פירוש הדבר הוא שתיאלצו לירות במישהו שהוא הגיבור בסיפור שלו, אדם שבסיפור חייו הדמות הראשית תמות כבר בעמוד 80. זוהי טרגדיה, אך היא עדיפה על פני הטרגדיה החלופית. זוהי הבחירה שכל שוטר עושה, מדי יום ביומו, כדי למנוע מהעולמות הקטנים והמסודרים שלנו להתפרק לכדי כאוס.

כאשר אתם מבינים לעומק את המניעים הפסיכולוגיים של האויב – כשאתם מבינים מה באמת עובר לו בראש – הידע הזה לא ירגיש כמו הנחתת חבטה הגונה ומספקת על הצד היריב. הוא לא יעניק לכם את הסיפוק המתוק של זעם קדוש. הוא לא יגרום לכם להרגיש טוב עם עצמכם. אם ההבנה שלכם גורמת לכם לעצב עמוק, ייתכן שאתם רואים את המציאות נכוחה. במקרים נדירים יותר, הבנה נכונה של המציאות עשויה להעביר בכם צמרמורת של אימה צרופה, כמו בעת מפגש עם פסיכופתים אמיתיים, או עם אנשים בעלי תפקוד מוחי תקין לחלוטין שמערכת האמונות שלהם פשוט שיבשה את דעתם (כדוגמת סיינטולוגים או מחנות "ג'יזס קאמפ"*).

אז הבה נאמר זאת בפה מלא: חוטפי ה-11 בספטמבר לא

---

* *ג'יזס קאמפ* הוא כינוי למחנות קיץ נוצריים־אוונגליקליים שהתקיימו בארה"ב והתפרסמו בעקבות הסרט התיעודי "Jesus Camp" משנת 2006. בסרט תועדו ילדים המשתתפים במחנות אלו כשהם עוברים אינדוקטרינציה דתית ופוליטית אינטנסיבית במטרה לגייסם כ"חיילים בצבא האל".

היו מוטנטים מרושעים. הם לא שנאו חופש. גם הם היו הגיבורים בסיפורים של עצמם, והם מתו למען מטרות נאצלות בעיניהם – אמת, צדק ואסלאם. אך גם אם החוטפים ראו את עצמם כך, אין פירוש הדבר שאמונותיהם היו נכונות. גם אם החוטפים ראו את עצמם כך, אין פירוש הדבר שחובה עלינו להצדיק את מעשיהם. גם אם החוטפים ראו את עצמם כך, אין פירוש הדבר שנוסעי טיסה 93 של 'יונייטד איירליינס' היו אמורים לשבת בחיבוק ידיים ולאפשר להם להשלים את המשימה. פירוש הדבר הוא שבעולם אחר, אילו היו גדלים בסביבה שונה, החוטפים הללו היו עשויים להיות שוטרים. וזוהי אכן טרגדיה. ברוכים הבאים לכדור הארץ.

62

טיפשות הפוכה איננה חוכמה

"...ואז האנשים שלנו בציר הזמן הזה נכנסו לפעולה כדי למזער נזקים. תראה."
הוא ניקה את המסך והחל להקיש במקלדת במהירות. עמוד אחר עמוד הופיעו, ובהם דיווחים של אנשים שטענו כי ראו את העצמים המסתוריים, כל דיווח מופרך יותר מקודמו.
"שיטת הטיוח הרגילה," חייך וֶרקַן וָאל. "שמעתי מעט מאוד דיבורים על 'הצלחות המעופפות', וגם המעט הזה נאמר כבדיחה. בתרבות שמצויה בשלב התפתחותי כזה, הדרך לקבור סיפור אמיתי אחד היא להקיף אותו בעשרה סיפורים מקבילים אחרים, שקריים בעליל."

– ה. בִּים פַּייפֶּר, *Police Operation*[1]

פייפר עלה פה על משהו. באופן אישי, אני לא מאמין שחייזרים־שמסתתרים־ברישול שורצים סביבנו. אבל לספקנות שלי אין שום קשר לאי־הרציונליות המביכה והקיצונית של כתות הצלחות המעופפות – כך אני מקווה, לפחות.

אתם ואני מאמינים שכתות הצלחות המעופפות התפתחו בהיעדרה המוחלט של נוכחות חייזרית במציאות. כתות יכולות להתפתח כמעט סביב כל רעיון הודות לטיפשות האנושית. הטיפשות הזו *אינה תלויה כלל* בחייזרים: היינו מצפים לראות כתות של צלחות

מעופפות בין אם ישנם חייזרים במציאות ובין אם לאו. גם אם אכן היו פה חייזרים שמסתתרים ברישול, הסבירות להיווצרות כתות כאלה *לא היתה קטנה יותר*. ההסתברות המותנית (חייזרים | כתות)P אינה נמוכה מההסתברות המותנית (אין חייזרים | כתות)P, אלא אם כן אתם סבורים שחייזרים שמסתתרים ברישול היו פועלים במכוון כדי למנוע את היווצרותן של כתות הצלחות המעופפות. לפי ההגדרה הבייסיאנית לראיות, העובדה ש"קיימות כתות של צלחות מעופפות" אינה מהווה ראיה *נגד* קיומן של צלחות מעופפות. היא פשוט אינה מהווה ראיה לשום צד.

זהו יישום של העיקרון הכללי שלפיו, כפי שניסח זאת רוברט פירסיג, "גם אם השוטה הגדול ביותר בעולם טוען שהשמש זורחת, אין זה אומר שחשוך בחוץ."[2]

אם הייתם מכירים אדם שטועה ב-99.99% מהפעמים בשאלות של 'כן ולא', הייתם יכולים להגיע לרמת דיוק של 99.99% פשוט על ידי הפיכת התשובות שלו. אבל כדי לייצר התאמה הפוכה ומדויקת כל כך למציאות, היה עליו לעשות את העבודה הקשה של איסוף ראיות אמינות ועיבוד עקבי שלהן – ואז *להפוך* אותן. הוא היה חייב להיות גאון־על כדי להיות כל כך טיפש.

מכונית עם מנוע מקולקל לא יכולה לנסוע לאחור במהירות של 300 קמ"ש, גם אם המנוע שלה *ממש*, *ממש* מקולקל.

אם טיפשות אינה מייצרת מתאם הפוך לאמת, מדוע שנחשוב שרוע אנושי ייצור מתאם כזה? ההפך מאפקט ההילה* הוא אפקט הקרניים: הנטייה לתפוס את כל התכונות השליליות כקשורות זו בזו. אם סטלין הוא רשע, אזי כל מה שהוא אומר חייב להיות שקר. הרי לא הייתם רוצים להסכים עם *סטלין*, נכון?

סטלין האמין גם ש- 2+2=4. ובכל זאת, אם תגנו על אמירה כלשהי של סטלין, אפילו על "2+2=4", אנשים יבחינו רק בכך שאתם "מסכימים עם סטלין" ויסיקו שאתם בצד שלו.

---

* ראו פרק 103.

נגזרות של עיקרון זה:

- כדי לטעון בכנות נגד רעיון, עליכם להתמודד עם הטיעונים הטובים ביותר של התומכים החזקים ביותר שלו. הבסת הטיעונים החלשים אינה מוכיחה דבר, משום שגם הרעיון המוצלח ביותר ימשוך אליו אנשים שיציגו אותו בצורה חלשה. אם אתם רוצים לטעון נגד טרנס-הומניזם או נגד פיצוץ אינטליגנציה,* עליכם להתעמת ישירות עם הטיעונים של ניק בוֹסטרוֹם או של אליעזר יודקובסקי משנת 2003 ואילך. הדרך הכי פחות נוחה היא הדרך הלגיטימית היחידה.
- הצגתם לראווה של תמהונים נלעגים ומעוררי רחמים, שיצאו מדעתם בגלל האופן שבו תפסו רעיון מסוים, אינה מהווה ראיה נגד הרעיון עצמו. אנשי ניו־אייג׳ רבים הפכו לסהרוריים אף יותר בעקבות האופן שבו פירשו את מכניקת הקוונטים.
- מישהו אמר פעם: ״לא כל השמרנים טיפשים, אבל רוב האנשים הטיפשים הם שמרנים.״ אם אינכם יכולים להגיע למצב תודעתי שבו האמירה הזו (נכונה או לא) פשוט אינה רלוונטית כביקורת על השמרנות, הרי שאינכם מוכנים עדיין לחשוב על פוליטיקה בצורה רציונלית.
- טיעון לגופו של אדם (אד הומינם) אינו תקף.
- עליכם להיות מסוגלים לטעון נגד רצח־עם מבלי לומר ״היטלר רצה להשמיד את היהודים.״ אם היטלר לא היה תומך ברצח עם, האם הדבר היה הופך את המעשה לתקין?
- במונחיו של רובין הַנסון: הנטייה הטבעית שלכם להאמין בדבר־מה תשתנה בהתאם למידת הרצון שלכם להיות *מזוהים* עם האנשים שמחזיקים באותה אמונה – ללא כל קשר לשאלה

---

* *פיצוץ אינטליגנציה* (intelligence explosion) הוא מונח שטבע המתמטיקאי הבריטי I.J. Good בשנת 1966, המתאר תרחיש תיאורטי שבו מערכת בינה מלאכותית מגיעה לרמה שמאפשרת לה לשפר את עצמה באופן רקורסיבי, מה שמוביל לעלייה מהירה ומעריכית באינטליגנציה, עד לרמה של אינטליגנציית־על שעולה בהרבה על זו של בני אדם.

האם האמונה הזו אכן *נכונה*. ישנם אנשים שיסרבו להאמין שאלוהים אינו קיים, לא משום שישנן ראיות לקיומו, אלא בעיקר משום שהם נרתעים מלהיות מזוהים עם ריצ׳רד דוקינס או עם אותם אתאיסטים "לוחמניים" שמסתובבים בציבור ומכריזים שאין אלוהים.

- גם אם המחשב שלכם מפסיק לעבוד, אינכם יכולים להסיק מכך שכל הרכיבים שלו פגומים ושאתם חייבים מחשב חדש לגמרי בלי מעבד של AMD, בלי כרטיס מסך של ATI, בלי דיסק קשיח של Maxtor ובלי מאווררים – למרות שבמחשב הישן שלכם היו את כל הדברים האלה. ייתכן שאתם פשוט צריכים כבל חשמל חדש.
- גם אם מאה ממציאים נכשלו בניסיון לבנות מכונה מעופפת ממתכת, עץ ובד, אין זה אומר שהדרך להצלחה היא מכונה מעופפת עשויה עור ועצמות. גם אם אלף פרויקטים נכשלו בניסיון ליצור בינה מלאכותית באמצעות מחשוב מבוסס־חשמל, אין זה אומר שהחשמל הוא מקור הכישלון. עד שלא תרדו לשורש הבעיה, הסיכוי ששינויי כיוון מלאי אופטימיות יקלעו לפתרון הוא קלוש ביותר.

---

1. Henry Beam Piper, "Police Operation," *Astounding Science Fiction* (July 1948).
2. Robert M. Pirsig, *Zen and the Art of Motorcycle Maintenance: An Inquiry Into Values*, 1st ed. (New York: Morrow, 1974).

63

הטיעון מייתר את הסמכות

תרחיש 1: בארי הוא גיאולוג מפורסם. צ׳ארלס הוא עבריין נוער בן ארבע־עשרה עם עבר פלילי עשיר והתקפים פסיכוטיים מזדמנים. בארי מציג בפני ארתור טענה מפתיעה ולא אינטואיטיבית לגבי סלעים, וארתור מעריך את ההסתברות לנכונותה ב-90%. לאחר מכן, צ׳ארלס מציג טענה מפתיעה באותה המידה לגבי סלעים, וארתור מעריך את הסתברותה ב-10%. ברור כי ארתור מביא בחשבון את סמכותו של הדובר כשהוא מחליט אם להאמין לטענותיו.

תרחיש 2: דייויד מציג טענה מפתיעה ולא אינטואיטיבית בפיזיקה, ופורס בפני ארתור הסבר מפורט שמגבה אותה, בצירוף סימוכין. ארני מציג טענה מפתיעה באותה המידה, אך מספק הסבר לא משכנע הכולל כמה קפיצות לוגיות. הן דייויד והן ארני מצהירים שזהו ההסבר הטוב ביותר שהם מסוגלים לספק (לכל אדם שהוא, לא רק לארתור). ארתור מעריך את ההסתברות לנכונות הטענה של דייויד ב-90% לאחר ששמע את ההסבר שלו, אך מעריך את הסתברות הטענה של ארני ב-10% בלבד.

על פניו, קיימת כאן סימטריה מסוימת: בשני המקרים נלקח בחשבון נתון רלוונטי – מידת הסמכות של הדובר או עוצמת הטיעון שלו.

אך כעת נניח שארתור מבקש מבארי ומצ׳ארלס להציג ניתוח טכני מלא, לרבות סימוכין. נניח שבארי וצ׳ארלס מציגים שניהם ניתוחים

טובים באותה המידה, וארתור בודק את המקורות ומוצא שהם אמינים. במקביל, ארתור מברר מהו הרקע המקצועי של דייויד וארני ומגלה שהם בעלי רקע זהה – אולי שניהם ליצנים, אולי שניהם פיזיקאים.

בהנחה שארתור בעל ידע מספיק כדי להבין את הניתוחים הטכניים – אחרת מדובר בלא יותר מגיבובי מילים מרשימים – הרי שעליו לייחס לטענתו של דייויד יתרון סבירות עצום לעומת זו של ארני, בעוד שלבארי יישאר לכל היותר יתרון זניח על פני צ׳ארלס.

אכן, אם הניתוחים הטכניים טובים מספיק, יתרונו של בארי על פני צ׳ארלס כלל אינו ראוי להתייחסות. טיעון טכני טוב הוא כזה שהופך את סמכותו האישית של הדובר *ללא רלוונטית*.

באופן דומה, אם אנו משוכנעים שהטיעון של ארני הוא אכן הטוב ביותר שהוא *מסוגל* להפיק – טיעון הכולל את כל שלבי ההיסק שביצע ואת כל הראיות התומכות שלקח בחשבון, תוך ציטוט כל מומחה שעליו אולי הסתמך – אזי אנו יכולים, פחות או יותר, להתעלם מכל מידע לגבי הרקע המקצועי שלו. ארני יכול להיות פיזיקאי או ליצן, זה לא ישנה דבר. (שוב, כל זאת בהנחה שיש לנו מספיק ידע טכני כדי להבין את הטיעון. אחרת, ארני פשוט פולט הברות מסתוריות, והשאלה אם נבחר ״להאמין״ להברות הללו תלויה במידה רבה במידת הסמכות שלו.)

כך שנראה כי קיימת אסימטריה בין טיעון לסמכות. גם אם הסמכות ידועה לנו, אנו עדיין מעוניינים לשמוע את הטיעונים. אך אם הטיעונים ידועים לנו במלואם, נותר לנו מעט מאוד ללמוד מן הסמכות.

״ללא ספק,״ עשוי לטעון ההדיוט, ״סמכות וטיעון הם סוגי ראיות שונים לחלוטין – הבדל שהכלים הפשטניים־מדי של תורת ההסתברות הבייסיאנית פשוט לא מסוגלים להסביר.״ שהרי בעוד שעוצמת הראיות (90% מול 10%) זהה לחלוטין בשני המקרים, הן אינן מתנהגות באותו האופן כאשר משלבים ביניהן. כיצד ניתן להסביר זאת?

הנה הדגמה טכנית חלקית לאופן שבו ניתן לייצג את ההבדל הזה

בתורת ההסתברות (לגבי השאר תוכלו להשתכנע על סמך הסמכות האישית שלי, או לעיין בסימוכין).

אם $P(H \mid E_1) = 90\%$ ו- $P(H \mid E_2) = 9\%$, מהי ההסתברות $P(H \mid E_1, E_2)$?

כלומר, אם ידוע שבהינתן ש-E_1 נכון ההסתברות של H היא 90%, ובהינתן ש-E_2 נכון ההסתברות היא 9%, מה תהיה ההסתברות של H בהינתן ששניהם נכונים? המידע הנתון פשוט אינו מאפשר לחשב זאת במסגרת תורת ההסתברות. לא, המידע החסר אינו ההסתברות הפריורית של H, אלא העובדה שייתכן ש-E_1 ו-E_2 תלויים זה בזה.

נניח ש-H הוא "המדרכה שלי חלקלקה", E_1 הוא "הממטרה שלי פועלת", ו-E_2 הוא "עכשיו לילה". משך פעולת הממטרה הוא עשר דקות, והמדרכה הופכת לחלקלקה דקה אחת לאחר הפעלת הממטרה ועד לסיום הפעולה. לכן, אם אנו יודעים שהממטרה פועלת, ההסתברות היא 90% שהמדרכה חלקלקה. הממטרה פועלת במהלך 10% מהלילה, כך שאם אנו יודעים שעכשיו לילה, ההסתברות שהמדרכה חלקלקה היא 9%. אך אם אנו יודעים שעכשיו לילה ושהממטרה פועלת – כלומר, אם ידועות לנו שתי העובדות גם יחד – ההסתברות שהמדרכה חלקלקה היא עדיין 90%.

אנחנו יכולים להמחיש זאת על ידי תרשים הזרימה הבא:



השאלה אם עכשיו לילה *משפיעה* על השאלה אם הממטרה פועלת, והשאלה אם הממטרה פועלת *משפיעה* על השאלה אם המדרכה חלקלקה.

כיוון החצים הוא קריטי. נניח שהיה לנו התרשים הבא:



משמעות התרשים הזה היא שאילו *לא ידעתי* דבר על הממטרה, ההסתברויות של השאלות "האם לילה עכשיו?" ו"האם המדרכה חלקלקה?" היו בלתי תלויות זו בזו. לדוגמה, נניח שאני מטיל שתי

קוביות וסוכם את המספרים שהתקבלו:



אם אינכם מגלים לי את סכום שני המספרים אלא רק שהקובייה הראשונה הראתה 6, זה עדיין לא אומר לי דבר על התוצאה של הקובייה השנייה. אך אם תגלו לי כעת שהסכום הוא 7, אדע מיד שהקובייה השנייה הראתה 1.

ההבחנה אם פריטי מידע שונים תלויים זה בזה או לא, בהינתן ידע מוקדם מסוים, היא למעשה סוגיה טכנית מורכבת למדי. הספרים המומלצים לקריאה בנושא זה הם Probabilistic Reasoning in Intelligent Systems[1] ו-Causality[2] מאת יהודה פרל (אם יש לכם זמן לקרוא רק ספר אחד, קראו את הראשון).

אם אתם יודעים לקרוא גרפים סיבתיים, הרי שכשאתם מסתכלים על תרשים הטלת הקוביות תראו מיד:

$$P(\text{קוביה 1}, \text{קוביה 2}) = P(\text{קוביה 1}) \times P(\text{קוביה 2})$$
$$P(\text{קוביה 1}, \text{קוביה 2}|\text{סכום}) \neq P(\text{קוביה 1}|\text{סכום}) \times P(\text{קוביה 2}|\text{סכום})$$

אם אתם מסתכלים על התרשים הנכון של המדרכה, תבחינו בעובדות הבאות:

$$P(\text{לילה}) \neq P(\text{לילה}|\text{חלקלקות})$$
$$P(\text{ממטרה}) \neq P(\text{ממטרה}|\text{חלקלקות})$$
$$P(\text{לילה}, \text{ממטרה}|\text{חלקלקות}) = P(\text{לילה}|\text{חלקלקות})$$

כלומר, ההסתברות שהמדרכה חלקלקה – בהינתן מידע הן על הממטרה והן על הלילה – זהה להסתברות שהיינו מעריכים אילו ידענו רק על הממטרה. הידע על הממטרה הפך את הידע על הלילה לבלתי רלוונטי ביחס למדרכה.

תופעה זו נקראת *חציצה*, והקריטריון המאפשר לנו להסיק

אי־תלויות מותנות כאלו מתוך גרפים סיבתיים נקרא *הפרדת-D*. במקרה של טיעון וסמכות, התרשים ייראה כך:



אם משהו הוא נכון, הרי שמטבע הדברים נוטות להצטבר ראיות לטובתו, ולפיכך מומחים מבחינים בראיות אלו ומשנים את דעתם. (בתיאוריה!)

אם אנו רואים שמומחה מאמין במשהו, אנו מסיקים מכך על קיומן של ראיות כלשהן (אף שאיננו יודעים מהן בדיוק), ומתוך קיומן של הראיות הללו אנו מסיקים על אמיתות הטענה.

אך אם אנו יודעים את ערכו של הצומת "איכות הטיעון", הדבר יוצר הפרדת-D בין הצומת "אמת" לצומת "סמכות מקצועית" על ידי חסימת כל המסלולים ביניהם, וזאת בהתאם לקריטריונים טכניים מסוימים ל"חסימת מסלול" (שנראים ברורים למדי במקרה זה). כך שגם מבלי לבדוק את התפלגות ההסתברות המדויקת, אנו יכולים להסיק מתוך הגרף כי:

$$P(\text{סמכות מקצועית, איכות הטיעון}|\text{אמת}) = P(\text{איכות הטיעון}|\text{אמת})$$

אין בכך משום סתירה לתורת ההסתברות הרגילה. זוהי פשוט דרך נוחה יותר לבטא עובדות הסתברותיות מסוימות. יכולתם להסיק את אותם שוויונות ואי־שוויונות מתוך התפלגות הסתברותית רגילה, אך היה קשה יותר להבחין בכך במבט חטוף. סמכות וטיעון אינם זקוקים לשני סוגים שונים של הסתברויות, ממש כשם שממטרות אינן עשויות מחומר שונה במהותו מאור השמש.

מבחינה מעשית, אין דרך להימנע *לחלוטין* מפנייה אל הסמכות. מומחים אמיתיים יכירו בסבירות גבוהה יותר את הראיות הנגדיות הקיימות שיש לקחת בחשבון. אנשים שפחות מומחים בתחומם יכירו אותן פחות – וזה הופך את הטיעונים שלהם להרבה פחות אמינים. אי אפשר לגשר על הפער הזה רק באמצעות בחינת הראיות שאותם

אנשים *כן* לקחו בחשבון.

בנוסף, קשה מאוד לתרגם טיעונים לשפה מתמטית טהורה. בהיעדר תרגום כזה, הערכת חוזקו של צעד היסק עשויה להישען על אינטואיציות שאינכם מסוגלים לשחזר ללא שלושים שנות ניסיון בתחום.

זה לגיטימי לחלוטין לייחס הסתברות *גבוהה מעט יותר* למה ש-א. ת. גֶ׳יינס* אומר לכם על הסתברות בייסיאנית, מאשר לאליעזר יודקובסקי שטוען את אותו הדבר בדיוק. חמישים שנות ניסיון אינן יכולות להיות חסרות השפעה *לחלוטין*.

אבל כוחה של הסמכות יפה רק כאשר כל שאר התנאים שווים, והוא מתפוגג בקלות אל מול טיעונים חזקים יותר. אני עצמי מצאתי טעות קלה באחד מספריו של ג׳יינס — מפני שאלגברה מנצחת סמכות.

---

1. Pearl, *Probabilistic Reasoning in Intelligent Systems*.
2. Judea Pearl, *Causality: Models, Reasoning, and Inference*, 2nd ed. (New York: Cambridge University Press, 2009).

---

* פרופ׳ אדווין תומפסון גֶ׳יינס (E. T. Jaynes, 1922-1998) היה מתמטיקאי ופיזיקאי תיאורטי. עסק רבות בהסתברות בייסיאנית וקידם את הפרשנות של תורת ההסתברות כהרחבה של הלוגיקה.

64

לחבק את השאלה

באמנות הרציונליות ישנו כלל של *קרבה־לעצם־העניין* – השאיפה לבחון ראיות הקרובות ככל הניתן לשאלה המקורית, באופן שייתר כמה שיותר טיעונים אחרים.

האחים רייט מצהירים: "המטוס שלנו יטוס." אם תבחנו את סמכותם המקצועית (מכונאי אופניים שהם במקרה גם פיזיקאים חובבים מעולים) ותשוו אותה, נניח, לזו של לורד קלווין, תגלו שללורד קלווין יש סמכות גדולה הרבה יותר.

אך אם תדרשו לראות את החישובים של האחים רייט ותצליחו להבין אותם, ואז תדרשו לראות את החישובים של לורד קלווין (וככל הנראה אין לו כאלה, מעבר לספקנותו האישית), הרלוונטיות של סמכותם המקצועית תצנח משמעותית.

אם תראו במו עיניכם *את המטוס טס*, החישובים עצמם יהפכו למיותרים במובנים רבים, וסמכותו של קלווין כבר לא תהיה שווה אפילו התייחסות.

ככל שהטיעונים שלכם נוגעים בשאלה באופן *ישיר* יותר, ללא שלבי היסק המתווכים ביניהם – ככל שהצמתים שאתם בוחנים קרובים יותר לצומת שעליו נסובה השאלה בתוך מארג הסיבתיות הגדול – כך גדל כוחה של הראיה. זהו חוק בגרפים סיבתיים: לעולם לא תוכלו להפיק *יותר* מידע מצמתים מרוחקים, מאשר מצמתים קרובים יותר החוצצים בפניהם ומייתרים אותם.

ג׳רי קְליבר אמר: ״מה שמכשיל אתכם הוא לא העובדה שנמנעתם מליישם איזו טכניקה מתוחכמת ומורכבת במיוחד. מה שמכשיל אתכם הוא שפספסתם את העיקר. לא שמרתם על קשר עין עם הכדור.״[1]

כשם שעדיף לטעון לגופו של מדע מאשר לגופה של סמכות מקצועית, כך עדיף לטעון לגופו של מדע מאשר לגופה של רציונליות. מי היה רציונלי יותר, האחים רייט או לורד קלווין? אם אנו יכולים לבחון את החישובים שלהם, למי אכפת! מעלותיו הרציונליות של אדם, *כשלעצמן*, אינן מסוגלות לגרום למטוס להמריא.

אם תשכחו את העיקרון הזה, למידה על הטיות נוספות רק תפגע בכם, משום שהיא תסיח את דעתכם מטיעונים ישירים יותר. קל להפליא להאשים מישהו ב״הטיה מס׳ 182״ מתוך רשימת ההטיות הגנריות שלכם, אך אינכם יכולים *להכריע* מחלוקת במציאות ללא ראיות מהשטח. גם אם קיימת הטיה מחשבתית כלשהי לטובת הטיעון שהשמש זורחת, אין פירוש הדבר שחשוך בחוץ.

כשם שלא תמיד ניתן לבצע ניסוי כאן ועכשיו, כך לא תמיד ניתן לבדוק את החישובים בו ברגע. לעיתים חסר לכם הרקע המקצועי הנדרש, לעיתים המידע אינו נגיש לכם, ולפעמים פשוט אין זמן. לדאבוננו, ישנם יותר מדי מקרים שבהם אין ברירה אלא להעריך את רמת הרציונליות של הדובר. עם זאת, עליכם לעשות זאת עם תחושת חמיצות קלה בלב – מעין הרגשה שמשהו מהותי חסר.

בכל פעם שאתם יכולים, רקדו קרוב ככל האפשר לשאלה המקורית – הצמידו את עצמכם אליה – התקרבו מספיק כדי לחבק את השאלה!

---

1. Jerry Cleaver, *Immediate Fiction: A Complete Writing Course* (Macmillan, 2004).

65

רציונליות והשפה האנגלית

בתגובה לדיון שלי על נורת מחיאות הכפיים,* מישהו ציין שהכתיבה שלי מזכירה לו את המאמר של ג׳ורג׳ אורוול, *פוליטיקה והשפה האנגלית*.[1] הרגשתי כבוד גדול. במיוחד מאחר שכבר חשבתי על הנושא לפרק זה.

אם אתם *באמת* רוצים נקודת מבט של אמן על רציונליות, קראו את אורוול. הוא בגדר קריאת חובה לרציונליסטים, לא פחות מאשר לסופרים. אורוול לא היה מדען אלא סופר. הכלים שלו לא היו מספרים, אלא מילים. היריב שלו לא היה הטבע, אלא הרוע האנושי. אם ברצונכם לכלוא אנשים למשך שנים ללא משפט, אתם חייבים למצוא דרך לנסח זאת אחרת מאשר: ״אני הולך לכלוא את מר ג׳נינגס למשך שנים ללא משפט.״ עליכם לערפל את חשיבתו של המאזין, ולמנוע מתמונות ברורות לזעזע את מצפונו. לכן אתם אומרים: ״גורמים בלתי־יציבים נתונים להליך צדק חלופי.״

אורוול היה יריב מושבע של משטרים טוטליטריים ושל ערפול המחשבה שמאחוריו מסתתר הרוע. וכך הפכו כתביו על השפה לנכסי צאן ברזל של רציונליות, בשורה אחת עם פַיינמַן, סֵייגָן או דוֹקינס.

״מומלץ לכותבים להימנע משימוש בסביל.״ רציונליסט שהרקע שלו מדעי *בלבד* עשוי להחמיץ את הפגם במשפט זה, אך כל אדם

---

* ראו פרק 19.

בעל ניסיון מינימלי בכתיבה יבחין בו מיד. כתבתי את המשפט בלשון סביל, מבלי לגלות לכם *מי* בכלל ממליץ לכותבים להימנע משימוש בו. לשון סביל מעלימה את הפועֵל ומותירה רק את המופעל. "גורמים בלתי־יציבים נתונים להליך צדק חלופי" — נתונים על ידי *מי*? מה בדיוק *עושים* ב"הליך צדק חלופי"? אם רק תשתמשו במספיק שמות עצם מופשטים, תוכלו למנוע מכל דבר לא נעים *להתרחש בפועל*.

מאמרים אקדמיים נכתבים לא פעם בלשון סביל. (סליחה, תיקון: מדענים מסוימים כותבים את המאמרים שלהם בלשון סביל. הרי אין זה כאילו המאמרים כותבים את עצמם מבלי שאיש ייאלץ לתת על כך את הדין.) זה נשמע סמכותי יותר לומר "לנבדקים הונפקה תרופת פרוֹגֶ'ניטוֹריבוֹקס"* מאשר "נתתי לכל סטודנט בקולג' בקבוק של 20 כדורי פרוג'ניטוריבוקס, ואמרתי להם לקחת אחד בכל לילה עד שייגמרו." כשמעלימים את המדען מהתיאור, נותרים כביכול רק הנתונים החשובים באמת. אך במציאות המדען נמצא שם, והנבדקים הם סטודנטים בשר ודם, והפרוג'ניטוריבוקס לא "הונפקה" בחלל הריק אלא נמסרה ביד יחד עם הנחיות. לשון סביל מסתירה את המציאות.

לפי התגובות שאני בדרך כלל מקבל, מישהו ודאי ימחה ויטען שהשימוש בלשון סביל במאמר מדעי כלל אינו חטא — שהרי בסופו של דבר, *כדור* לכל שהמדען נוכח שם. אין כאן שום כשל לוגי. ואולם, זו בדיוק הסיבה שבגללה רציונליסטים צריכים לקרוא את אורוול, ולא רק את פיינמן או אפילו את גֶ'יינס.

ספרי עיון מעבירים *ידע*, פרוזה מעבירה *חוויה*. מחקר רפואי יכול להסיק מה יקרה לאדם שייחשף לחלל הריק ללא הגנה. סיפור טוב יגרום לכם לחוות זאת על בשרכם.

רציונליסטים נוטים לנסות לפרק לגורמים ביטויים מוליכי־שולל.

---

* *פרוֹגֶ'ניטוֹריבוֹקס* (progenitorivox) היא תרופה פיקטיבית מתוך סרטון האנימציה הסאטירי "The Drugs I Need", הלועג לפרסומות של חברות תרופות המשתמשות במניפולציות כמו שמות ארוכים ומוזיקת־רקע עליזה כדי למסך את מחירן הכפול ואת רשימת תופעות הלוואי הארוכה שלהן.

הם ינסו לראות האם *ייתכן* שחבוי בו גרעין של הגיון, ויחפשו עבורם הסבר לוגי. הם יפגינו "נדיבות פרשנית" כלפי הטקסט וייתנו לכותב ליהנות מהספק. סופרים, לעומת זאת, מאומנים שלא לתת לעצמם ליהנות מחמת הספק. *מה* שהקהל *חושב* שאמרתם – *זה מה* שאמרתם, בין אם התכוונתם לכך ובין אם לאו. אין טעם להתווכח עם הקהל, לא משנה כמה מתוחכמים ההסברים שלכם.

סופר יודע שהקוראים *לא* יעצרו לרגע כדי לחשוב. חוויית קריאה היא זרם מתמשך של רושם ראשוני. סופר־רציונליסט שם לב *לחוויה* שהמילים מייצרות. אם בעת הניסיון שלכם להעריך את מידת הרציונליות של אמירה פומבית מסוימת, אתם מנתחים את המילים בכובד ראש, מנסחים מחדש טענות, בוחנים כוונות נסתרות ומחפשים אחר שביבים של הגיון – הרי שפספסתם לחלוטין את הרושם הראשוני. פספסתם את *מה* שהקהל *רואה*, או ליתר דיוק, *מרגיש*.

סופר היה מבחין מיד בצרימה הזועקת של "לנבדקים הונפקה תרופת פרוג'ניטוריבוקס." איזו חוויה בדיוק חווה הקורא? המשפט הזה מייצר תחושה מרוחקת של סמכות, וזה *הכל* – החוויה *היחידה* היא התחושה שאומרים לך דבר־מה מהימן. סופר היה מבין שמדובר בשמות־עצם מופשטים מכדי להמחיש את ההתרחשות הממשית – את הפוסט־דוקטורנט עם הבקבוק ביד, שמנסה להיראות קשוח, מול הסטודנט שמקשיב בחיוך עצבני.

אינני מנסה לטעון שמאמרים מדעיים צריכים להיכתב כמו סיפורים, אלא שעל הרציונליסט לפתח רגישות *לחוויות* שהמילים מייצרות. רציונליסט חייב להבין כיצד פועלת התפיסה האנושית ומה מפעיל אותה. זה כולל את זרם התודעה – אותו חלק של עצמכם שמתבטא באמצעות השפה. רציונליסט חייב להיות מודע לחווייה שביטויים בשפה יוצרים בפועל, הרבה מעבר לתוכן הלוגי היבש שלהם.

או, אם לומר זאת בצורה בוטה יותר: *כתיבה מדויקת אינה מכפרת על רושם מוטעה!*

לא אכפת לי איזה הסבר רציונלי יש לכם עבור נורת־מחיאות־כפיים כמו "יש לפתח בינה מלאכותית בצורה דמוקרטית." שום דבר

אינו יכול להצדיק את האפקט הלא־רציונלי שיוצר המשפט הזה, שלא לדבר על הניסוח המעורפל שלו, שמניח מראש את המבוקש.

הנה אורוול, מתקומם נגד *ההשפעה* של קלישאות על חוויית החשיבה:

> כשאתם צופים באיזה עסקן פוליטי שחוק על הבמה, מדקלם את הביטויים המוכרים – חייתיות, מעשי זוועה, רגל גסה, עריצות מגואלת בדם, עמי העולם החופשי, עמידה שכם אל שכם – מתעוררת בכם לעיתים קרובות תחושה משונה שאינכם צופים בבן אנוש חי אלא בסוג של בובה... נואם המשתמש בסגנון דיבור שכזה כבר עבר כברת דרך לקראת הפיכתו למכונה. הצלילים הנכונים בוקעים ממיתרי הקול שלו, אך המוח שלו אינו מעורב בתהליך כפי שהיה אילו בחר את מילותיו בעצמו...
>
> הדבר החשוב ביותר הוא לאפשר לרעיון לבחור את המילים, ולא להיפך. בכתיבה, הדבר הגרוע ביותר שניתן לעשות עם מילים הוא להיכנע להן. כשאתם חושבים על חפץ מוחשי, אתם חושבים ללא מילים, ואז, אם ברצונכם לתאר את מה שדמיינתם, אתם כנראה מחפשים ומוצאים מילים מדויקות מספיק שיהלמו אותו. כשאתם חושבים על משהו מופשט, אתם נוטים להשתמש במילים כבר מהרגע הראשון. אם לא תעשו מאמץ מודע לעצור זאת, מטבעות הלשון הקיימים יפלשו ויעשו את העבודה עבורכם, במחיר של טשטוש – או אפילו שינוי – של כוונתכם המקורית. כנראה שעדיף לדחות את השימוש במילים ככל הניתן, ולגבש לעצמכם את משמעות המחשבות שלכם באמצעות דימויים ותחושות, עד כמה שתצליחו.

צ׳ארלס סַנדֶרס פֶּרְס* יכול היה לכתוב את הפסקה האחרונה. יותר מנתיב אחד יכול להוביל אל *הדרך*.

---

* צ׳רלס סַנדֶרס פֶּרְס (1839–1914) היה פילוסוף, מדען ולוגיקן אמריקאי בולט, ממייסדי הפרגמטיזם (אסכולה פילוסופית הבוחנת רעיונות לאור השלכותיהם המעשיות) והסֶמיוטיקה (תורת הסימנים והמשמעות).

1. George Orwell, "Politics and the English Language," *Horizon* (April 1946).

66

רוע אנושי ומחשבה מעורפלת

ג׳ורג׳ אורוול חזה בשקיעתו של העולם הנאור אל תוך הטוטליטריות. הוא ראה בנפילתן ובהשחתתן של מדינה אחר מדינה. המגף הרומס פני אדם, לָנֶצח, וזכרו היטב – לָנֶצח. אתם נולדתם מאוחר מכדי לזכור את הימים שבהם עליית הטוטליטריות נראתה ככוח בלתי ניתן לעצירה, כשמדינות שלמות קרסו בזו אחר זו לידי אימת המשטרה החשאית ולקול הדפיקה הרועמת בדלת באמצע הלילה. כל זאת בזמן שאנשי אקדמיה באוניברסיטאות המערב הריעו לטיהורים הסטָליניסטיים וראו בהם אות לקידמה. לכם זה ודאי מרגיש זר כמו סיפור בדיוני. קשה לכם לקחת את זה ברצינות. זאת משום שעל ציר הזמן שלכם, חומת ברלין כבר נפלה. ואם שמו של אורוול אינו חקוק על אחת מאותן אבנים – הוא ראוי להיות.

אורוול ראה את הגורל הצפוי למין האנושי, והשקיע מאמץ נואש בניסיון להסיט אותו ממסלולו. הנשק של אורוול היה כתיבה בהירה. אורוול ידע כי שפה מעורפלת משמעה מחשבה מעורפלת. הוא ידע כי רוע אנושי ומחשבה מעורפלת שזורים זה בזה כמו גדילים משלימים של DNA:[1]

> בימינו, כתיבה ונאומים פוליטיים משמשים ברובם להגנה על מה שאינו ניתן להגנה. תופעות כמו המשך השלטון הבריטי בהודו, הטיהורים והגירושים ברוסיה, והטלת פצצות האטום על יפן – כולן אכן ניתנות להגנה, אבל רק באמצעות טיעונים ברוטליים מכדי

> שהציבור הרחב יוכל לשאת. אלו הם טיעונים שאינם מתיישבים עם מטרותיהן המוצהרות של המפלגות הפוליטיות. לפיכך, השפה הפוליטית מורכבת ברובה ממכבסת מילים, הנחת המבוקש וערפל סמיך. כפרים חסרי הגנה מופצצים מהאוויר, תושביהם מגורשים אל השדות, הבקר נקטל על ידי מכונות ירייה, והבקתות מועלות באש בפצצות תבערה: לפעולה זו קוראים "השכנת שלום"...

אורוול הגדיר בבירור את מטרת השפה הבהירה שלו:

> אם תַפשטו את השפה שלכם, תשתחררו מהטיפשות הקשה ביותר של המחשבה הממסדית. לא תוכלו לדבר בשפה ה"נכונה", וכשתגידו משהו מטופש, הטיפשות שלו תהיה גלויה לעין, אפילו לכם עצמכם.

להפוך את הטיפשות שלנו לגלויה לעין, אפילו לנו עצמנו – זוהי המהות של הבלוג *Overcoming Bias*.

הרוע מתגנב בהחבא דרך הצללים החשוכים של התודעה. אנו מביטים לאחור בפרספקטיבה היסטורית צלולה, וליבנו נשבר לזכר הרעב ההמוני המתוכנן של סטלין ומָאוֹ, שהרג עשרות מיליונים. אנו מכנים זאת רוע, משום שהדברים נעשו מתוך כוונה מודעת להסב כאב ומוות לבני אדם חפים מפשע. אנחנו מכנים זאת רוע, בגלל שאט הנפש שאנו חשים כלפי המעשים הללו, בפרספקטיבה היסטורית צלולה. אך כדי שמעשי הרוע לא ייתקלו בהתנגדות המתבקשת, אסור לתת לתחושת המיאוס לצוף אל פני השטח. יש להימנע מבהירות בכל מחיר. באותה מידה שבה אנשים בעלי ראייה צלולה נוטים להתנגד לרוע שהם רואים, כך הרוע האנושי – בכל מקום שבו הוא קיים – שואף לערפל את המחשבה.

הספר *1984* ממחיש זאת בצורה בוטה: הנבלים בסיפורו של אורוול הם אלו שגוזרים ומטשטשים תצלומי ארכיון (בהשראת הנוהג הסובייטי ההיסטורי של חיתוך וטשטוש תמונות פוליטיות). בנקודת השפל האפלה ביותר במיניסטריון האהבה, אובריאן מענה את וינסטון כדי שיודה ששתיים ועוד שתיים הם חמש:

"אתה זוכר," הוסיף ואמר, "שכתבת ביומן שלך 'חירות היא החירות לומר ששתיים ועוד שתיים הן ארבע'?"

"כן," אמר וינסטון.

אובריאן הרים את ידו השמאלית, גבה כלפי וינסטון, האגודל מסותר וארבע האצבעות פשוטות.

"כמה אצבעות אני פושט, וינסטון?"

"ארבע."

"ואם תאמר המפלגה שיש כאן לא ארבע אצבעות, אלא חמש – כמה אז?"

"ארבע."

המילה נסתיימה באנקת־כאב. מחט החוגה קפצה לחמישים וחמש. זיעה כיסתה את כל גופו של וינסטון. האוויר פרץ אל ריאותיו ויצא מתוכן באנקות עמוקות שלא יכול לעצרן אפילו בחירוק־שיניים. אובריאן הביט בו, ארבע אצבעותיו עודן פשוטות. הוא משך והחזיר את המסוט. הפעם הוקל הכאב רק במקצת.*

אני נותר פעור פה פעם אחר פעם נוכח אנשים אינטליגנטים לכאורה – כדוגמת הכלכלן טַיילר קוֹאֶן, עמיתו של רובין הַנסון – שאינם חושבים שהתגברות על הטיות קוגניטיביות היא עניין חשוב. אנו מדברים על *התודעה* שלכם. על האינטליגנציה האנושית שלכם. היא זו שמבדילה אתכם מהקוף. היא זו שבנתה את הציוויליזציה שבה אתם חיים. אתם באמת חושבים שהאופן שבו התודעה פועלת אינו חשוב? שכשלים שיטתיים של המחשבה אינם חשובים? האם אתם חושבים שהאינקוויזיציה הייתה מענה מכשפות, אילו כולם היו בייסיאנים מושלמים?

טיילר קואן מרגיש כנראה שהניסיון להתגבר על הטיות הוא מוטה בדיוק כמו ההטיות עצמן: "אני רואה בבלוג של רובין דוגמה מובהקת להטיה, מה שממחיש שהטיות אכן יכולות להיות שימושיות מאוד." אני רק יכול *לקוות* שזוהי תוצאה של נטייה להתפלספות מתוך ניסיון להישמע חכם. האם טיילר באמת משווה בין חוסר רגישות לקנה מידה

* *1984* בתרגום ג. אריוך (עם עובד, 2003).

בנוגע לערך חיי אדם לבין הניסיון לגבש תוכניות *שבאמת* יצילו כמה שיותר חיים?

אורוול נאלץ להילחם בגישה דומה – זו הטוענת שכל ניסיון להבחין בין טוב לרע הוא לא יותר מאשר תמימות נעורים:

> סטיוארט צֶ'ייס ואחרים היו קרובים לטעון שכל המילים המופשטות הן חסרות משמעות, והשתמשו בכך כתירוץ על מנת להצדיק חוסר מעש פוליטי. הרי מאחר שאינכם יודעים מהו פשיזם, כיצד תוכלו להילחם בו?

ייתכן שהניסיון להתגבר על הטיות אינו נראה מרגש מספיק כשהוא נתפס כמאבק נגד טעויות אקראיות בלבד. אולי קשה יותר להתרגש כשאין רוע מובהק להתייצב מולו. אז בואו נהיה ברורים לחלוטין: בכל מקום שבו יש רוע אנושי בעולם, בכל מקום שבו יש אכזריות, עינויים ורצח – ישנן הטיות שממסכות אותו. וכשאנשים בעלי ראייה צלולה מתייצבים נגד ההטיות הללו, הרוע הסמוי נלחם בחזרה. לאמת *אכן* יש אויבים. אילו הבלוג *Overcoming Bias* היה רואה אור בברית המועצות, כל מי שפרסם או הגיב בו היה נשלח מיד למחנות כפייה.

בכל ההיסטוריה האנושית, כל פריצת דרך משמעותית נולדה מתוך בהירות מחשבתית חדשה. ולמעט מספר אסונות טבע, כל טרגדיה גדולה נולדה מתוך טיפשות. האויב האחרון שלנו הוא אנו עצמנו. זוהי מלחמה, ואנחנו החיילים.

---

1. Orwell, "Politics and the English Language."

חלק ז׳

נגד רציונליזציה

חלק ניכר מתהליך ההסקה שלנו הוא למעשה רציונליזציה – טוויית סיפורים שגורמים לאמונותינו הנוכחיות להיראות עקביות ומוצדקות יותר, מבלי לשפר בהכרח את מידת הדיוק שלהן. ״נגד רציונליזציה״ [רצף זה] עוסק בבעיה זו, ואחריו הרצפים ״נגד דוּחוֹשֶׁב״ (על הונאה עצמית) ו״לראות בעיניים חדשות״ (על האתגר שבזיהוי ראיות שאינן תואמות את הציפיות וההנחות שלנו).

– רוב בנסינגר,
רציונליות: מבוא

67

ידע על הטיות יכול להזיק

פעם ניסיתי להסביר לאמא שלי על בעיית הכיול של מומחים, ואמרתי: "...אז כשמומחה אומר שהוא בטוח במשהו ב-99%, זה קורה בפועל רק בערך ב-70% מהמקרים." או אז השתרר רגע של שתיקה, כשלפתע קלטתי שאני מדבר עם אמא שלי, ומיהרתי להוסיף: "כמובן, צריך לוודא שמחילים את הספקנות הזו בצורה שוויונית, כולל על עצמך, ולא רק משתמשים בה כטיעון נגד כל דבר שלא מסכימים איתו—"

ואמא שלי אמרה: "השתגעת? זה מעולה! אני הולכת להשתמש בזה כל הזמן!"

המאמר של טֵייבֶּר ולוֹדג', "ספקנות מגמתית בהערכת אמונות פוליטיות", מציג הוכחות לשש השערות שונות:[1]

1. אפקט הדעה הקדומה: נבדקים בעלי עמדה מוצקה בנושא מסוים יעריכו טיעונים שתומכים בדעתם כאמינים וטובים יותר מאשר טיעונים נגדיים, גם כאשר הם נדרשו במפורש לשמור על אובייקטיביות.
2. הטיית ההפרכה: נבדקים ישקיעו יותר זמן ומאמץ בלתקוף טיעונים שנוגדים את עמדתם מאשר לחזק טיעונים שתומכים בה.
3. הטיית האישוש: נבדקים שחופשיים לבחור את מקורות המידע שלהם יחפשו מקורות תומכים ולא מקורות נגדיים.
4. **הקצנת עמדות: חשיפת נבדקים לטיעונים מאוזנים לכאורה**

בעד ונגד שני הצדדים, רק תגביר ותקצין את עמדתם המקורית.

5. אפקט עוצמת הדעה: נבדקים המביעים עמדות נחרצות יותר יהיו פגיעים יותר להטיות שתוארו לעיל.
6. **אפקט התחכום: נבדקים בעלי ידע פוליטי נרחב יהיו פגיעים יותר להטיות שתוארו לעיל, מכיוון שהם מחזיקים ב"תחמושת" רבה יותר כדי לנגח עובדות וטיעונים שאינם תואמים את עמדתם.**

אם אתה לא־רציונלי, ידע נוסף יכול *להזיק* לך. עבור בייסיאני אמיתי, הערך הצפוי של מידע חדש לְעולם אינו שלילי. אבל בני אדם אינם מפעילים היגיון בייסיאני באופן מושלם. אם לא נזהר, אנחנו עלולים להיחתך.

ראיתי אנשים שהידע שלהם על הטיות שיבש אותם לחלוטין. היתה להם תחמושת רבה יותר כדי להתווכח נגד כל דבר שהם לא אוהבים. והבעיה הזו – יותר מדי תחמושת זמינה – היא אחת הדרכים המרכזיות שבהן אנשים בעלי יכולות מנטליות גבוהות הופכים בסופו של דבר לטיפשים, במובן של "דיס־רציונליות" לפי סטנוביץ'.

אתם ודאי מכירים אנשים שעונים לתיאור הזה. אנשים בעלי מנת משכל גבוהה, שהמיומנות הגבוהה שלהם בויכוחים רק גורמת להם להשיג *פחות* תוצאות? האם אתם חושבים שהייתם עוזרים להם – הופכים אותם לרציונליים יותר – לו הייתם מציגים להם רשימה של הטיות מוכרות?

אני נזכר במישהו שלמד על בעיית הכיול וביטחון־היתר של מומחים. זמן קצר לאחר מכן הוא אמר: "ובכן, אי אפשר לסמוך על מומחים. הם טועים כל כך הרבה – כפי שהניסויים הוכיחו. לכן, כשאני מעריך מה יהיה בעתיד, אני מעדיף להניח שהדברים ימשיכו כפי שהיו בעבר –" והפליג לתוך מערכת היקשים מורכבת, מועדת לטעויות ומפוקפקת ביותר. איכשהו, בכל הנוגע למסקנות המועדפות עליו, כל אותן הטיות וכשלים נראו הרבה פחות *בולטים* – וקפצו הרבה פחות בקלות לראש – מאשר כשהוא היה צריך לנגח טיעון של מישהו אחר.

סיפרתי לאותו אדם על בעיית הטיית ההפרכה ואפקט התחכום, והנה הפלא ופלא, בפעם הבאה שאמרתי משהו שלא מצא חן בעיניו, הוא האשים אותי בכך שאני מתחכם. הוא לא ניסה להצביע על טיעון מתוחכם ספציפי, או על פגם כלשהו בטיעון שלי – רק הניד בראשו ונאנח בעצב על כך שאני ככל הנראה משתמש באינטליגנציה שלי כדי להביס את עצמה. הוא פשוט השיג עוד *טיעון־נגד כללי לחלוטין*.

אפילו עצם המושג "תחכום רטורי" יכול להיות קטלני, אם הוא קופץ לראשכם בקלות רבה מדי כשאתם נתקלים באדם אינטליגנט לכאורה שאומר משהו שאתם לא אוהבים.

אני משתדל ללמוד מהטעויות שלי. בפעם האחרונה שהעברתי הרצאה על היוריסטיקות והטיות, פתחתי בהצגת המושג הכללי דרך כשל החיתוך והיוריסטיקת היציגות. ואז עברתי להטיית האישוש, הטיית ההפרכה, אפקט התחכום, ספקנות מגמתית וכשלים אחרים שנובעים מדעות קדומות. העברתי את שלושים הדקות הבאות *בלהלום* בנושא הזה, כשאני מציג אותו שוב ושוב מכמה שיותר זוויות שונות.

רציתי לעורר בקהל עניין בנושא. ובכן, תיאור של כשל החיתוך והיוריסטיקת היציגות היה עשוי להספיק. אבל נניח שהם אכן היו מגלים עניין. מה אז? הספרות המקצועית על הטיות היא בעיקר פסיכולוגיה קוגניטיבית לשם פסיכולוגיה קוגניטיבית. הייתי חייב לתת לקהל שלי את אזהרות המסע החמורות שלהם במהלך ההרצאה האחת הזו, אחרת הם כנראה לא היו שומעים אותן לעולם.

בין אם אני עושה זאת בכתב ובין אם בעל פה, אני משתדל כעת לעולם לא להזכיר כיול וביטחון־יתר, אלא אם כן דיברתי קודם לכן על הטיית ההפרכה, ספקנות מגמתית, אפקט התחכום ודיס־רציונליות אצל אנשים בעלי יכולות מנטליות. ראשית, אל תזיק!

1. Charles S. Taber and Milton Lodge, "Motivated Skepticism in the Evaluation of Political Beliefs," *American Journal of Political Science* 50, no. 3 (2006): 755–769, doi:10.1111/j.1540- 5907.2006.00214.x.

68

עדכנו את עצמכם בהדרגה

פוליטיקה היא אויב התבונה. ויכוח הוא מלחמה, טיעונים הם חיילים. ישנו פיתוי לנסות ולפרש כל תוצאה ניסויית אפשרית כך שתתמוך בתיאוריה שלכם, כמו הגנה על מבצר מפני כל נתיב תקיפה אפשרי. אינכם יכולים לעשות זאת. זה בלתי אפשרי מתמטית. עבור כל ציפייה לראיות, ישנה ציפייה שווה ומנוגדת לה לראיות נגדיות.*

אבל זה בסדר אם האמונה היקרה לליבכם אינה מוגנת *באופן מושלם*. אם ההשערה היא שהמטבע נופל על "עץ" ב-95% מהפעמים, הרי שפעם אחת מתוך עשרים תצפו לראות "ראיה נגדית". זה בסדר. זה נורמלי. זה אפילו צפוי, כל עוד יש לכם תשע־עשרה ראיות תומכות על כל ראיה נגדית אחת. מודל הסתברותי יכול לספוג מכה או שתיים, ועדיין לשרוד, כל עוד המכות אינן *ממשיכות להגיע*.

אף על פי כן, התפיסה המקובלת, במיוחד בשיח הציבורי, היא שתיאוריה נכונה לעולם אינה טועה, ושתיאוריה שגויה לעולם אינה צודקת.

תוכלו למצוא אנשים שמציגים פרט מידע בודד שנתפס בעיניהם כראיה, וטוענים שהתיאוריה שלהם יכולה "להסביר" אותו, כאילו די בכך כדי להוכיח את התיאוריה כולה. ככל הנראה, לא יתכנו ראיות התומכות בתיאוריה שגויה. לא יתכן שתיאוריה שגויה תצליח להסביר

* ראו פרק 28.

אפילו אירוע בודד. לפיכך, פרט מידע אחד שתומך בתיאוריה הוא כל מה שנדרש כדי להוכיח אותה.

זה טפשי כמעט באותה המידה להציג פיסת ראיה נגדית *הסתברותית* אחת כהפרכה, כאילו לא יתכן אפילו טיעון *קל שבקלים* נגד תיאוריה נכונה. אבל כך בני אדם מתווכחים כבר עידן ועידנים, מנסים להביס את כל טיעוני האויב, ומסרבים להעניק לו אפילו שביב בודד של תמיכה. אנשים רוצים את הדיונים שלהם חד־צדדיים. הם רגילים לעולם שבו לתיאוריות המועדפות עליהם אין שמץ של התנגדות. לפיכך, להתיר אפילו בדל ראיה נגדית הסתברותית יהיה סוף העולם.

אני פשוט יודע שאחד הקוראים עומד לומר: "אבל *אסור לך* לוותר אפילו על נקודה אחת אם אתה רוצה לנצח בויכוחים בעולם האמיתי! אם תודה שקיימים טיעוני נגד כלשהם, האויב ינצל זאת שוב ושוב — אתה לא יכול לתת לזה לקרות! אתה *תפסיד*! מה יכול לעורר פחד קיומי גדול יותר *מזה*?"

מה שתגיד. רציונליות לא נועדה לניצחון בדיבייט, היא נועדה להחלטה באיזה צד לבחור. אם בחרתם צד בויכוח, עבודת הרציונליות שלכם כבר *נעשתה*, לטוב או למוטב. אך כיצד אתם עצמכם יכולים לבחור צד בויכוח? אם *בחירה בצד הלא נכון* היא מפחידה באופן קיומי, אפילו רק טיפ־טיפה מפחידה באופן קיומי, מוטב שתבחנו היטב את *כל* הראיות.

רציונליות אינה הליכה בקו ישר, אלא ריקוד. בכל צעד בריקוד הזה, הרגל שלכם צריכה לנחות בדיוק במקום הנכון, לא ימינה מדי ולא שמאלה. להסיט את מידת האמונה כלפי מעלה עם כל שמץ של ראיה מאששת. להסיט את מידת האמונה כלפי מטה עם כל שמץ של ראיה נגדית. כן, כלפי *מטה*. אפילו עם המודל הנכון, אם הוא אינו מדויק לחלוטין, לפעמים תצטרכו לעדכן את האמונה שלכם כלפי *מטה*.

אם שמץ או שניים של ראיות נוגדים את האמונה שלכם, זה בסדר. זה קורה לפעמים עם ראיות הסתברותיות עבור תיאוריות

שאינן מדויקות לחלוטין. (אם תיאוריה מדויקת לחלוטין נכשלת, אתם *אכן* בצרות!) פשוט הסיטו את מידת האמונה שלכם מטה מעט – ההסתברות, יחס הסיכויים, או אפילו מידת השכנוע הפנימי הלא־מילולי שבתוככם. פשוט הסיטו אותו מעט כלפי מטה, והמתינו לראיות נוספות. אם התיאוריה נכונה, ראיות תומכות יגיעו במהרה, וההסתברות תעלה שוב. אם התיאוריה שגויה, ממילא אינכם רוצים בה.

הבעיה בחשיבה בינארית של שחור־לבן, חשיבה *איכותנית*, היא שכל תצפית בודדת מוכרחה לרסק את התיאוריה או להותיר אותה ללא פגע. כשלא מאפשרים אפילו ראיה נגדית בודדת, הדבר מייצר דיסוננס קוגניטיבי המחייב לתרץ אותה כדי להעלימה. וזה שולל התקדמות הדרגתית: זה שולל שקלול נכון של כל הראיות. בחשיבה הסתברותית אנו מבינים שבממוצע, תיאוריה נכונה תצבור משקל רב יותר ראיות לטובתה מאשר נגדה. ולכן אתם יכולים, *ללא פחד*, לומר לעצמכם: "זוהי ראיה נגדית קלה, אני אסיט את האמונה שלי כלפי מטה." כן, *מטה*. זה לא מרסק את התיאוריה היקרה לליבכם, כפי שגורסת חשיבה איכותנית. חישבו באופן כמותי.

עבור כל ציפייה לראיות, קיימת ציפייה שווה ומנוגדת לה לראיות נגדיות. בכל סיטואציה, עליכם, בממוצע, לצפות לעדכן את אמונותיכם כלפי מטה באותה מידה שאתם מצפים לעדכן אותן כלפי מעלה. אם יש לכם וודאות רבה אילו ראיות יגיעו, סימן שאתם כבר משוכנעים כמעט לחלוטין בתיאוריה שלכם – הסתברות קרובה ל־1 – מה שלא משאיר מקום רב להסתברות לעלות עוד כלפי מעלה. וככל שזה נראה יותר בלתי־סביר שתתקלו בראיה נגדית, העדכון כלפי מטה שנובע מכך חייב להיות גדול מספיק כדי לאזן במדויק את העדכון כלפי מעלה בצד השני. הממוצע המשוקלל של ההסתברות הפוסטריורית חייב להיות שווה להסתברות הפריורית.

כמה מגוחך זה, אם כן, להיות מבועתים מהאפשרות לעדכן את ההסתברות שלכם כלפי מטה, אם אתם בכלל טורחים לחקור סוגיה כלשהי? בממוצע, עליכם לצפות לאותה מידה של עדכון כלפי מטה

כמו לעדכון כלפי מעלה מכל תצפית בודדת.

ייתכן אולי שבדֶלי ראיות מפריכות יגיעו שוב, ושוב, ושוב, בעוד שראיות תומכות חדשות יגיעו במשורה. אתם עשויים למצוא את אמונתכם נסחפת מטה מטה, עוד ועוד. עד שלבסוף תבינו מאיזה כיוון נושבות רוחות הראיות נגדכם. ברגע זה של הבנה, אין טעם להמציא תירוצים. ברגע ההבנה הזה, כבר ויתרתם על האמונה היקרה לליבכם. יש! זמן לחגוג! פתחו בקבוק שמפניה או הזמינו פיצה! אחרי הכל, אינכם יכולים להפוך לחזקים יותר על ידי התבצרות באמונות שאיתן התחלתם.

69

טיעון אחד מול צבא שלם

בפרק הקודם דיברנו על סגנון חשיבה שבו אסור לקבל אפילו טיעון נגדי אחד, מה שמוביל לכך שצריך לתרץ ולהדוף כל תצפית שאינה תומכת בעמדתנו. כאן אני רוצה להציע שכשבני אדם נתקלים בטיעון נגדי, הם מונעים מעצמם להוריד את רמת הביטחון שלהם פשוט על ידי *חזרה* על הטיעונים התומכים שהם כבר מכירים.

נניח שבמדינת פְרִידוֹניה מתנהל ויכוח ער בשאלה האם השכנה שלהם, סילבַניה, אחראית למטר מטאורים שפקד אותה לאחרונה.* יש כמה ראיות שמצביעות על כך: המטאורים פגעו בערים סמוכות לגבול הסילבני; נרשמה פעילות חריגה בבורסה הסילבנית *רגע לפני* הפגיעות; ושגריר סילבניה, טֶרֶנטינוֹ, נשמע מסנן מתחת לשפמו משהו על "נקמה משמיים."

מישהו ניגש אליכם ואומר: "אני לא חושב שסילבניה אחראית למטר המטאורים. הרי יש לנו איתם קשרי מסחר של מיליארדי דינרים בשנה." "ובכן", אתם עונים לו, "המטאורים פגעו בערים קרובות לסילבניה, הייתה פעילות חשודה בבורסה שלהם, והשגריר שלהם דיבר מפורשות על נקמה משמיים." מכיוון שמשקל שלושת הטיעונים האלה גבוה ממשקל הטיעון הבודד שלו, אתם *ממשיכים* להחזיק באמונתכם שסילבניה אשמה — אתם מאמינים, ואינכם

* *פרידוניה וסילבניה* (Freedonia and Sylvania) הן שתי מדינות בדיוניות ויריבות מתוך סרט הקומדיה הסאטירי Duck Soup (1933).

ספקנים, מבחינה איכותנית. ברור שמאזן הראיות נוטה נגד סילבניה.

או אז ניגש אליכם אדם אחר ואומר: "אני לא חושב שסילבניה אחראית. לשלוט על כיוון של אסטרואיד זה עניין מורכב מאוד. לסילבניה אין אפילו תוכנית חלל." אתם משיבים לו: "אבל המטאורים פגעו בערים קרובות לסילבניה, המשקיעים שלהם ידעו על זה מראש, והשגריר הודה בכך בפה מלא!" שוב, שלושת הטיעונים הללו גוברים על הטיעון הבודד שלו (יחס של שלושה טיעונים נגד אחד), ולכן אתם דבקים באמונתכם שסילבניה אשמה.

למעשה, הביטחון שלכם רק *התחזק*. בשני מקרים נפרדים שקלתם את מאזן הראיות, ובשתי הפעמים הכף נטתה לטובתכם ביחס של 3 ל-1.

אתם ממשיכים להיתקל בטיעונים נוספים מפי הבוגדים הפרו־סילבנים – שוב, ושוב, ומאה פעמים ויותר – אבל בכל פעם, הטיעון החדש מובס בקלות ביחס של 3 ל-1. ובכל פעם שכזו, אתם הופכים בטוחים עוד יותר בכך שסילבניה אכן אשמה, ומעדכנים את ההסתברות הפריורית שלכם בהתאם למאזן הראיות שאתם מרגישים באותו רגע.

הבעיה היא, כמובן, שעל ידי *חזרה* על טיעונים שכבר ידעתם אתם סופרים את אותן ראיות פעמיים. זה היה נחשב לחטא חמור אפילו אם הייתם סופרים פעמיים את *כל* הראיות. (דמיינו מדען שעורך ניסוי על 50 נבדקים ולא מצליח להגיע לתוצאות בעלות מובהקות סטטיסטית, אז הוא פשוט סופר את כל הנתונים שלו פעמיים.)

אבל לספור פעמיים רק ראיות *מסוימות* ולא אחרות – זו כבר פארסה מוחלטת. אני זוכר סרט מצויר שראיתי בילדותי, שבו הנבל חילק את השלל לפי האלגוריתם הבא: "אחד לך, אחד לי. אחד לך, אחד־שתיים לי. אחד לך, אחד־שתיים־שלוש לי."

כפי שהדגשתי בפרק הקודם, גם אם אמונה יקרה ללבכם היא *נכונה*, רציונליסט עשוי לפעמים להסיט את ההסתברות שלה מטה בזמן שהוא משקלל את כל הראיות. אמת, מאזן הכוחות עדיין עשוי לנטות לטובת האמונה שלכם. אבל אתם עדיין חייבים להסיט את

ההסתברות כלפי *מטה* – כן, *מטה* – מכפי שהייתה לפני ששמעתם את הראיה הנגדית. אין שום תועלת ב*חזרה* על טיעונים תומכים, שכן כבר לקחתם אותם בחשבון מלכתחילה.

ואף על פי כן, נראה לי שכשבני אדם עומדים בפני טיעון נגדי *חדש*, הם מחפשים דרך שלא להוריד את רמת הביטחון שלהם, וכמובן שהם מוצאים אותה בדמות טיעונים תומכים שהם *כבר מכירים*. אני עצמי חייב לעמוד על המשמר ללא הרף כדי לא לעשות זאת בעצמי! זה מרגיש טבעי כמו להדוף מכת חרב באמצעות המגן שבידך.

בעזרת חשיבה שגויה מהסוג המתאים, קומץ של טיעונים תומכים – או אפילו טיעון אחד ויחיד – יכול לעמוד איתן אל מול צבא שלם של סתירות.

70

השורה התחתונה

שתי קופסאות חתומות, קופסה **א** וקופסה **ב**, עומדות למכירה פומבית. אחת מהן, ורק אחת, מכילה יהלום יקר ערך. יש כל מיני שיטות וסימנים לקבוע האם קופסה מכילה יהלום, אך אין בידי אף סימן שאני *יודע* שניתן לסמוך עליו לחלוטין. על קופסה אחת יש חותמת כחולה, למשל, ואני יודע שחותמות כאלו נפוצות יותר בקופסאות המכילות יהלום מאשר בקופסאות ריקות. או שפני השטח של קופסה אחת מבריקות, ויש לי חשד – אף שאינני בטוח – שקופסה המכילה יהלום לעולם אינה מבריקה.

עכשיו, נניח שרטוריקן מיומן ניגש לבעלים של קופסה **א** וקופסה **ב**, דף נייר בידו, ואומר להם: "תנו הצעת מחיר עבור השירותים שלי, ומי שיזכה בהם – אטען שהקופסה שלו היא זו שמכילה את היהלום, כך שהקופסה תתומחר גבוה יותר." בעלי הקופסאות מגישים הצעות, והבעלים של קופסה **ב** מציע מחיר גבוה יותר וזוכה בשירותיו של הרטוריקן.

הרטוריקן המיומן מתחיל לארגן את מחשבותיו. ראשית, הוא כותב בתחתית הדף שלו: "*זלכן*, היהלום נמצא בקופסה **ב**!" לאחר מכן, בראש הדף, הוא כותב: "על קופסה **ב** יש חותמת כחולה," ומתחת לזה: "קופסה **א** מבריקה," ואז: "קופסה **ב** קלה יותר מקופסה **א**," וכן הלאה דרך סימנים ומבחנים רבים. עם זאת, הרטוריקן המיומן מתעלם לחלוטין מכל אותם סימנים שעשויים להעיד לטובת קופסה

א. ואז הרטוריקן ניגש אליי ומקריא מתוך דף הנייר שלו: "על קופסה **ב** יש חותמת כחולה, וקופסה **א** מבריקה," וכן הלאה, ולבסוף מסכם: "*ולכן*, היהלום נמצא בקופסה **ב**."

אך חשבו על כך: ברגע שבו הרטוריקן המיומן רשם את המסקנה שלו, באותו הרגע בו נספג הדיו בדף הנייר שלו, מידת השזירה הראָיית של הדיו הפיזי הזה עם הקופסאות הפיזיות הללו נקבעה.

אולי יעזור לדמיין אוסף של עולמות – ענפי אָוורָט או כפילי טָגמָארק* – שבהם יש תדירות אובייקטיבית מסוימת שבה היהלום נמצא בקופסה **א** או בקופסה **ב**. בדומה לכך, יש תדירות אובייקטיבית מסוימת בתוך תת־הקבוצה "עולמות שבהם קופסה **א** מבריקה" שבה היהלום נמצא בקופסה **ב**, ותדירות אובייקטיבית מסוימת ב"עולמות שבהם קופסה **א** מבריקה ועל קופסה **ב** יש חותמת כחולה" שבה היהלום נמצא בקופסה **ב**.

כתמי הדיו על הנייר יוצרים צורות וקווים מעוקלים ומשונים, שנראים כמו הטקסט הזה: "*ולכן*, היהלום נמצא בקופסה **ב**." אם איתרע מזלכם להיות יודעי קרוא וכתוב, אתם עלולים להתבלבל ולחשוב בטעות שהדיו המעוצב הזה *משמעו* שהיהלום נמצא בקופסה **ב**. נבדקים שנדרשים לומר מהו הצבע של מילים מודפסות, ומוצגת להם המילה "ירוק" בצבע אדום, אומרים לעיתים קרובות "ירוק" במקום "אדום". זה עוזר להיות אנאלפבית, כדי שלא תתבלבלו מצורת הדיו.

עבורנו, המשמעות האמיתית של דבר היא השזירה שלו עם דברים אחרים. חשבו שוב על אוסף העולמות, ענפי אוורט או כפילי טגמארק. ברגע שכל הרטוריקנים בכל העולמות הספיגו דיו בשורה התחתונה של הנייר שלהם – נניח שזה קרה ברגע אחד – הדבר קיבע את המתאם של הדיו עם הקופסאות. הרטוריקן כותב בעט שאינו מחיק. הדיו לא ישתנה. הקופסאות לא ישתנו. בתוך תת־הקבוצה של העולמות שבהם הדיו אומר "ולכן, היהלום נמצא בקופסה **ב**," יש כבר

* ראו הערת הסבר בפרק 10.

אחוז קבוע מסוים של עולמות שבהם היהלום נמצא דווקא בקופסה **א**. זה לא ישתנה, ללא קשר למה שייכתב בשורות הריקות שממעל.

אם כן, השזירה הראייתית של הדיו היא קבועה, ותוכלו להחליט בעצמכם מה היא עשויה להיות. אולי בעלי קופסאות שמאמינים שניתן לטוות עבורם טיעון טוב יותר נוטים יותר לשכור מפרסמים. אולי בעלי קופסאות שחוששים מחולשת טיעוניהם מציעים מחיר גבוה יותר. אם בעלי הקופסאות בעצמם אינם מודעים לסימנים והשיטות, אזי לא תהיה שזירה כלל בין הדיו לתוכן הקופסאות, אם כי הוא עשוי לספר לכם משהו על מצבם הפיננסי של הבעלים או הרגלי המיקוח שלהם.

עכשיו, נניח שאדם אחר נוכח במקום, ומתוך סקרנות גרידא הוא רושם *קודם כל* את הסימנים המבדילים בין *שתי* הקופסאות על דף נייר, ואז משקלל את הידע שלו ואת חוקי ההסתברות וכותב בתחתית הדף: "*ולכן*, אני מעריך בהסתברות של 85% שהיהלום נמצא בקופסה **ב**." על מה מעיד כתב היד הזה? מעקב אחר שרשרת הסיבה והתוצאה המובילה לדיו הפיזי הזה על הנייר הפיזי הזה, מראה ששרשרת הסיבתיות מפלסת את דרכה דרך כל המבחנים והסימנים של הקופסאות, והיא תלויה בסימנים אלו – שכן בעולמות עם סימנים שונים, תיכתב בתחתית הדף הסתברות שונה.

כך שכתב ידו של החוקר הסקרן שזור בסימנים ובתוכן הקופסאות, בעוד שכתב ידו של הרטוריקן המיומן הוא ראיה אך ורק לשאלה מי שילם לו את המחיר הגבוה יותר. יש הבדל עצום במשמעות סימני הדיו, אם כי מי שיקרא בקול את צורות הדיו עשוי לחשוב לתומו שהמילים נשמעות דומות.

מידת הרציונליות שלכם נקבעת על ידי האלגוריתם שכותב בפועל את השורה התחתונה של מחשבותיכם. אם המכונית שלכם משמיעה רעשי חריקה מתכתיים בכל פעם שאתם בולמים, ואתם מעדיפים שלא להשקיע בהחלפת הבלמים שלכם, אתם יכולים לחפש סיבות מדוע המכונית שלכם אולי אינה זקוקה לתיקון. אך האחוז מכם ששורד בענפי אוורט או בעולמות טגמארק – שבו נשתמש כדי לתאר את

מידת הרציונליות שלכם – נקבע על ידי האלגוריתם שמחליט עבור *איזו* מסקנה אתם מחפשים טיעונים. במקרה זה, האלגוריתם האמיתי הוא "לעולם אל תתקן שום דבר יקר." אם זה אלגוריתם מוצלח, מה טוב. אם זה אלגוריתם גרוע, חבל מאוד. הטיעונים שתכתבו לאחר מכן, מעל השורה התחתונה, לא ישנו דבר לכאן או לכאן.

מטרת פרק זה היא לשמש אזהרה לגבי דרך החשיבה שלכם, ולא כטיעון־נגד מוחלט וכללי נגד מסקנות שאינכם אוהבים. שכן זהו אכן טיעון מחוכם לומר "היריב שלי הוא רטוריקן מחוכם," אם אתם משלמים לעצמכם כדי להצדיק כל אמונה שהייתה לכם מלכתחילה. גם אם הרטוריקן המחוכם ביותר בעולם טוען שהשמש זורחת, אין זה אומר שעכשיו לילה.

71

ראיה סלקטיבית היא ראיה ראויה?

דיברנו על סוגיית הרטוריקן המיומן שנשכר על מנת למכור לכם קופסה שמכילה, או לא מכילה, יהלום. הרטוריקן מסב את תשומת לבכם לחותמת הכחולה שמופיעה על הקופסה, והרי זו עובדה ידועה שחותמות כאלו מופיעות בתדירות גבוהה יותר על גבי קופסאות המכילות יהלומים. מה קורה בנקודה זו, מנקודת מבט בייסיאנית? האם אתם מוכרחים, חסרי אונים, לעדכן את ההסתברויות שלכם, כפי שהרטוריקן המיומן זומם?

אם יכולתם לבחון את הקופסה בעצמכם, הייתם יכולים לשקלל את כל הסימנים בעצמכם. אך מה אם אינכם יכולים לבחון אותה? מה אם הראיות היחידות שיש לכם הן אלו שהרטוריקן מראה לכם? נאמר שהוא מחויב על פי חוק לומר אך ורק טענות אמיתיות, אך הוא אינו מחויב לספר לכם כל מה שהוא יודע. כל טענה שהרטוריקן המיומן טוען היא ראיה תקפה – איך תוכלו *שלא* לעדכן את ההסתברויות שלכם? האם זה כבר לא נכון שבשיעור מסוים של ענפי אוורט או כפילי טגמארק שבהם לקופסה **ב** יש חותמת כחולה, היא אכן מכילה יהלום? לפי ג׳יינס, בייסיאני חייב תמיד לשקלל את כל הראיות הידועות לו, אחרת יקלע לפרדוקס. אך אם כך, הרטוריקן המיומן יכול לגרום לכם להאמין במה שרק יבחר, כל עוד ישנם מספיק ראיות לכל צד. משהו פה לא מסתדר.

נבחן מקרה פשוט יותר: מטבע מזויף, אשר עשוי להיות מתוכנן

כך שייפול על עץ בהסתברות של 2/3 ועל פֶּלי ב-1/3, או על עץ בהסתברות של 1/3 ועל פלי ב-2/3, כאשר לשתי האפשרויות יש סבירות אפריורית שווה. כל תוצאה של עץ היא ביט אחד של ראיה לטובת מטבע הנוטה לעץ.* כל תוצאה של פלי היא ביט אחד של ראיה לטובת מטבע הנוטה לפלי. אני מטיל את המטבע עשר פעמים, ואז אומר לכם: "תוצאת ההטלה הרביעית, השישית והתשיעית היא עץ." מהי ההסתברות הפוסטריורית שלכם שהמטבע נוטה לעץ?

והתשובה היא שהיא יכולה להיות כמעט כל דבר, תלוי בחוקיות שעומדת מאחורי הדיווח שלי – האופן שבו בחרתי על אילו הטלות מטבע לדווח.

- יכול להיות שהאלגוריתם שלי הוא לדווח על תוצאות ההטלה הרביעית, השישית והתשיעית, ללא כל קשר לתוצאות ההטלות. אם אתם יודעים שהשתמשתי באלגוריתם הזה, יחס ההסתברויות הוא 8:1 לטובת מטבע הנוטה לעץ.
- יכול להיות שדיווחתי על כל ההטלות שתוצאתן היא עץ, ורק עליהן. במקרה זה, אתם יודעים שתוצאת כל 7 ההטלות האחרות היא פלי, ויחס ההסתברויות הוא 1:16 נגד הסיכוי שהמטבע נוטה לעץ.
- יכול להיות שהחלטתי מראש לגלות את תוצאת ההטלה הרביעית, השישית והתשיעית רק אם ההסתברות שהמטבע מוטה לעץ היא מעל 98%. וכן הלאה.

או שניקח למשל את *בעיית מונטי הול*:

> בשעשוען טלוויזיה ניתנת לכם הבחירה בין שלוש דלתות המובילות לשלושה חדרים. אתם יודעים שבאחד החדרים יש 100,000 שקלים, ושני האחרים ריקים. המנחה מבקש מכם לבחור דלת, ואתם בוחרים בדלת מס׳ 1. אז המנחה פותח את דלת מס׳ 2, וחושף חדר ריק. האם אתם רוצים לעבור לדלת מס׳ 3, או להישאר עם דלת מס׳ 1?

* להסבר אודות 'ביטים' מתמטיים ראו פרק 23.

התשובה תלויה באלגוריתם של המנחה. אם המנחה תמיד פותח דלת ותמיד בוחר בדלת המובילה לחדר ריק, כדאי לכם לעבור לדלת מס׳ 3. אם המנחה תמיד פותח את דלת מס׳ 2 ללא קשר למה שנמצא מאחוריה, לשתי הדלתות האחרות יש כעת סיכויים שווים להוביל לחדר עם הכסף. אם המנחה פותח דלת כלשהי אך ורק אם בחרתם מלכתחילה בדלת עם הכסף, אז אתם בהחלט צריכים לדבוק בדלת מס׳ 1.

ההסתברות אינה תלויה רק בכך שדלת מס׳ 2 ריקה, אלא גם בעובדה שהמנחה *בחר* לפתוח את דלת מס׳ 2. אנשים רבים מתבלבלים בבעיית מונטי הול מכיוון שהם מעדכנים את ההסתברות רק על סמך התצפית שדלת מס׳ 2 ריקה, ובמקרה כזה לדלת מס׳ 1 ולדלת מס׳ 3 יש הסתברות שווה להכיל את הכסף. זו הסיבה שחובה על בייסיאנים לחשב את ההסתברות תוך שימוש בכל הידע שברשותם, ולא – יקלעו לפרדוקס.

כשמישהו אומר ״תוצאת הטלת המטבע הרביעית היא עץ,״ איננו מתייחסים לעובדה שתוצאת הטלת המטבע הרביעית היא עץ – איננו לוקחים את תת־הקבוצה של כל העולמות האפשריים שבהם תוצאת ההטלה הרביעית היא עץ – אלא אנו לוקחים את תת־הקבוצה של כל העולמות האפשריים שבהם מישהו הפועל לפי אלגוריתם מסוים אמר ״תוצאת הטלת המטבע הרביעית היא עץ.״ המילים שנאמרו אינן העובדה עצמה. אל תתנו למשמעות המילולית של המילים להטעות אתכם.

מרבית ההליכים המשפטיים מבוססים על ההנחה שבכל תיק משפטי יש בדיוק שני צדדים מנוגדים, וקל יותר למצוא שני בני אדם מוטים מאשר אדם אחד חסר פניות. האינטרסים המנוגדים של התביעה וההגנה מבטיחים שלכל ראיה יימצא מי שירצה להציג אותה, וכך בית המשפט ייחשף לכל הראיות. זוהי התיאוריה. אם ישנם שני רטוריקנים מיומנים בדילמת הקופסאות, המצב אמנם אינו אידיאלי כמו חוקר סקרן אחד, אך הוא טוב כמעט באותה המידה. אבל זה נכון בבעיות כמו שתי הקופסאות. במציאות, הבעיות הן לרוב מורכבות,

עמוקות ובעלות פנים רבות, כאלו שלא ניתן לפתור בקלות באמצעות צעקות הדדיות בין כחולים לירוקים.

הישמרו שלא להפוך את הרעיון של ראיות סלקטיביות *לטיעון נגד מוחלט וכללי*, כדי לפסול כל ראיה שאינכם אוהבים: "הטיעון הזה הוא סלקטיבי, לכן עליי להתעלם ממנו." אם טיעון נגדי מעצבן אתכם, סימן שאתם מכירים את הנושא, ואכפת לכם ממנו מספיק כדי לתפוס צד. אתם כנראה כבר מכירים את הטיעונים החזקים ביותר של הצד שלכם. אין לכם שום סיבה להסיק, מתוך טיעון נגדי, על קיומם של סימנים חיוביים חדשים לטובתכם שטרם ראיתם. אתם נותרים, אם כן, עם העובדות הלא נוחות עצמן. חותמת כחולה על קופסה **ב** היא עדיין ראיה.

אך אם אתם נחשפים לדיון בפעם הראשונה, ואתם שומעים רק צד אחד שלו, אזי אכן עליכם להיזהר! במובן מסוים, איש אינו יכול באמת לבטוח בתורת האבולוציה עד שהוא מקשיב לבריאתנים במשך חמש דקות — *רק אז* הוא מבין כמה היא איתנה.

72

רציונליזציה

לפני שני פרקים הצגתי את הדילמה של שתי הקופסאות, שרק באחת מהן יש יהלום, עם מבחנים וסימנים שונים בתור ראיות. העמדתי זה מול זה את החוקר הסקרן ואת הרטוריקן המיומן. החוקר הסקרן רושם את כל המבחנים והסימנים, משקלל אותם, ובסופו של דבר כותב: "*ולכן*, אני מעריך בהסתברות של 85% שהיהלום נמצא בקופסה **ב**." הרטוריקן עובד עבור המרבה במחיר, והוא מתחיל בכתיבת השורה התחתונה: "*לפיכך*, היהלום נמצא בקופסה **ב**," ואז בוחר את המבחנים והסימנים התומכים במסקנה כדי לרשום אותם בשורות הקודמות.

התהליך הראשון הוא רציונליות. לתהליך השני קוראים בדרך כלל "רציונליזציה".

"רציונליזציה". איזו מילה מוזרה. לדעתי היא *שגויה מיסודה*. אי אפשר להפוך לרציונלי מה שאינו רציונלי מלכתחילה. זה כמו לקרוא לשקר "אימות."

ברמה חישובית טהורה, יש הבדל עצום בין:

1. להתחיל מראיות, להריץ את חישובי ההסתברות, ולהפיק את המסקנה ההסתברותית (לכתוב את כל המבחנים והסימנים, ואז לשקלל הכל כך שהשורה התחתונה תלויה באותם מבחנים וסימנים).
2. להתחיל מהמסקנה, ואז להריץ חישובי הסתברות במטרה להפיק ראיות שלכאורה תומכות במסקנה (לכתוב את השורה

התחתונה, ואז ללכת אחורה ולבחור אילו מבחנים וסימנים לכתוב בשורות שמעליה).

מי השוטה שהמציא מילים דומות ומבלבלות כל כך, "רציונליות" ו"רציונליזציה", כדי לתאר תהליכי חשיבה שונים זה מזה באופן כה קיצוני? הייתי מעדיף מונחים שיבהירו מיד את ההבדל האלגוריתמי, למשל "רציונליות" לעומת "חור שחור קוגניטיבי עצום ששואב הכל".

לא כל שינוי הוא שיפור, אך כל שיפור הוא בהכרח שינוי. אי אפשר לקחת טענה שנקבעה מראש ולהפוך אותה לאמיתית יותר באמצעות טיעונים. אפשר לגרום ליותר אנשים להאמין בה, אך אי אפשר להפוך אותה לאמיתית יותר. כדי לשפר את האמונות שלנו, אנו חייבים בהכרח לשנות אותן. רציונליות היא הפעולה שבאמצעותה אנו מדייקים את האמונות שלנו על ידי שינוי שלהן. רציונליזציה משמשת על מנת לקבע את האמונות הקיימות במקומן. ראוי היה לקרוא לה "אנטי־רציונליות", הן בשל תוצאותיה בפועל והן בשל האלגוריתם ההפוך שלה.

"רציונליות" היא הזרימה *קדימה* שאוספת ראיות, שוקלת אותן, ומגיעה לכדי מסקנה. החוקר הסקרן השתמש באלגוריתם של זרימה קדימה: *תחילה* אסף את הראיות, רשם את כל המבחנים והסימנים שמצא, *ואז* עיבד אותם על מנת לקבל הסתברות שלא הייתה ידועה לו קודם לכן לגבי הקופסה המכילה את היהלום. לאורך כל הזמן שבו התהליך הרציונלי זרם קדימה, החוקר הסקרן עדיין לא ידע מהי המסקנה שלו. זוהי הסיבה שהוא היה *סקרן*. על פי *דרכו של בייס*, ההסתברות האפריורית שווה להסתברות הפוסטריורית הצפויה: אם אתה יודע מה היעד שלך, אתה כבר שם.

"רציונליזציה" היא זרימה *אחורה* מהמסקנה אל ראיות נבחרות. ראשית אתה כותב את השורה התחתונה, שהיא ידועה וקבועה מראש. מטרת התהליך היא לגלות אילו טיעונים עליך לכתוב בשורות שמעליה. שורות אלו, ולא השורה התחתונה, הן המשתנה הלא־ידוע באלגוריתם.

אני חושש שהרציונליות המסורתית אינה מפתחת אצל המשתמשים בה רגישות להבדל שבין זרימה קדימה לזרימה אחורה. ברציונליות המסורתית, אין שום דבר רע במדען שממציא השערת־מחמד ואז יוצא לחפש ניסוי שיוכיח אותה. רציונליסט מסורתי יביט בכך בסיפוק ויאמר: "הגאווה הזו היא המנוע שדוחף את המדע קדימה." ובכן, זה *אכן* המנוע שדוחף את המדע קדימה. קל יותר למצוא תובע וסנגור מוטים לכיוונים מנוגדים מאשר אדם אחד חסר פניות.

אבל רק בגלל שכולם עושים משהו, זה לא הופך אותו לבסדר. היה עדיף בהרבה לו המדען, לאחר שהמציא את השערת־המחמד, היה יוצא לבחון אותה מתוך *סקרנות טהורה* — יוצר ניסויים שעשויים להוביל את אמונותיו שלו לכיוון לא ידוע.

אם אתם באמת לא יודעים לאן אתם הולכים, טבעי שתרגישו סקרנות לגבי העתיד. סקרנות היא המעלה הראשונה,* שבלעדיה החיפוש שלכם יהיה חסר תכלית והכישורים שלכם נעדרי כיוון.

הרגישו את זרימת הכוח, והבטיחו שהוא לא זורם לאחור.

* ראו באינטרלוד *שתים־עשרה המעלות של הרציונליות*, בכרך החמישי של *רציונליות: מ-א ועד ת*.

73

טיעון רציונלי

אתה מנהל קמפיינים למחייתך, וזה עתה נשכרת על ידי מוֹרטימֶר ק׳ סְנוֹדְגְרַאס, מועמד מפלגת הירוקים לראשות עיריית הֶדליבֶּרג.* בתור מנהל קמפיינים שקורא ספר על רציונליות, שאלה מציקה טורדת את מנוחתך: ״איך אני יכול לבנות טיעון רציונלי תקף לכך שמורטימר ק׳ סנודגראס הוא המועמד הטוב ביותר לראשות עיריית הדליברג?״

מצטער. זה פשוט בלתי אפשרי.

״מה?!״ אתה זועק. ״אבל מה אם אבסס את הטיעון שלי על ראיות מוצקות בלבד? מה אם כל עובדה שאציין תהיה נכונה למיטב ידיעתי, ותהווה ראיה רלוונטית לפי חוק בייס?״

מצטער. זה עדיין בלתי אפשרי. הבסת את עצמך ברגע שבו קבעת מראש את מסקנת הטיעון שלך.

השנה, *חדשות הדליברג* שלחו שאלון בן 16 שאלות לכל המועמדים לראשות העיר, עם שאלות כמו ״האם צבעת בצבעים של הרוחות?״** ו-״האם לקחת לריאות?״ אך אבוי, משרדי החדשות הושמדו על ידי מטאור טרם הפרסום. חבל, כיוון שהמועמד שלך,

---

* *הֶדליבֶּרג* היא עיירה בדיונית מסיפורו הסאטירי של מארק טוויין, *האיש שהשחית את העיירה* (1899). בסיפור מתגאים תושבי העיירה ביושרה מוסרית ללא רבב, אך זו מתגלה כצביעות שבירירית ברגע שהם ניצבים בפני פיתוי כספי אמיתי. השם ״סְנוֹדְגְרַאס״ שימש את טוויין עצמו כשם עט סאטירי, ומוכר עוד קודם לכן כדמות קומית אצל צ׳ארלס דיקֶנס.

** מתוך *צבעי הרוח*, שיר הנושא של הסרט פוקהונטס.

מורטימר ק׳ סנודגראס, מציג נתונים טובים בהשוואה למתחריו ב-15 מתוך 16 השאלות. נקודת התורפה היחידה שלו היא שאלה 11 : ״האם אתה כיום, או היית בעבר, רשע מרושע?״

אז אתה מתפתה לפרסם את השאלון כחלק מחומרי הקמפיין שלך... ללא שאלה 11, כמובן.

מה שחוצה את הגבול שבין *רציונליות לרציונליזציה*. כעת כבר אין לקוראים אפשרות להסתמך על העובדות לבדן. הם חייבים לשקול גם את העובדה הנוספת – הצורה בה הוצגו הדברים – ולהסיק מכך על קיומן של ראיות נסתרות.

למעשה, חצית את הגבול כבר ברגע שבו בחנת האם השאלון משרת את המועמד שלך או פוגע בו, עוד לפני שהחלטת האם לפרסם אותו. ״מה?!״ אתה זועק. ״האם על קמפיין לפרסם עובדות שפוגעות במועמד שלו?״ אבל שים את עצמך בנעליו של בוחר שעדיין לא החליט למי להצביע – למה שתתעלם ממידע רלוונטי? לא היית עושה זאת, אילו היית סקרן באמת. אילו היית זורם *קדימה* מהראיות לעבר בחירה שאינה ידועה מראש, במקום לזרום *אחורה* ממועמד שכבר בחרת בו לעבר הטיעונים הרצויים.

טיעון ״לוגי״ הוא טיעון הנובע מהנחות היסוד שלו. לכן, הטיעון הבא *אינו לוגי*:

- כל המלבנים הם מרובעים.
- כל הריבועים הם מרובעים.
- לפיכך, כל הריבועים הם מלבנים.

הסילוגיזם* הזה לוקה בכשל לוגי על אף שהנחות היסוד, ואפילו המסקנה, נכונים. כדאי להבחין בין היקשים לוגיים לכאלה שאינם לוגיים, ולסרב לקבל אותם, גם אם המסקנות שלהם במקרה נכונות. סיבה אחת היא שההבחנה הזו עשויה להשפיע על האופן שבו נעדכן את אמונותינו לאור ראיות עתידיות. סיבה אחרת היא שרשלנות מחשבתית הופכת להרגל.

---

* *סילוגיזם* הוא מבנה של טיעון לוגי שאופיין לראשונה על ידי אריסטו: שתי הנחות יסוד מהן נובעת המסקנה בהכרח.

מעל לכל, הסילוגיזם נכשל בהצגת ההסבר האמיתי. גם אם כל הריבועים הם מלבנים, אין זה *מפני* ששניהם מרובעים. אפשר לקרוא לזה סילוגיזם צבוע – כזה שבו קיים נתק בין הנימוקים המוצגים כלפי חוץ לבין הסיבות האמיתיות שמאחורי הקלעים.

אם אתה באמת רוצה להציג טיעון תקף ורציונלי *לטובת המועמד שלך*, בקמפיין פוליטי, יש רק דרך אחת לעשות זאת:

- *לפני שמישהו שוכר את שירותיך*, השג את כל הראיות שביכולתך למצוא על המועמדים השונים.
- הכן רשימת בעד־ונגד שבה אתה עצמך תשתמש כדי להכריע מיהו המועמד הראוי ביותר.
- הסק מסקנות על פי הרשימה.
- גש למועמד המנצח.
- הצע לו להיות מנהל הקמפיין שלו.
- כשהוא יבקש חומרי הסברה לקמפיין, הדפס את הרשימה שלך.

רק בדרך זו תוכל להציע שרשרת טיעונים *רציונלית* – כזו שהשורה התחתונה שלה נכתבה מתוך זרימה *קדימה* מהשורות שמעליה. מה *שבאמת* משנה את השורה התחתונה שלך, הוא הדבר היחיד שאתה יכול *בלב שלם* לכתוב בשורות שמעליה.

74

להימנע מנקודות התורפה האמיתיות של האמונה שלך

לפני כמה שנים סבתא־רבתא שלי נפטרה, בשנות התשעים לחייה, אחרי דעיכה ארוכה, איטית ואכזרית. לא באמת זכיתי להכיר אותה, אבל בילדותי המוקדמת היא עוד בישלה עבור המשפחה. אני זוכר את הגעפילטע־פיש שלה, את פניה, ואני זוכר שהיא הייתה טובה אליי. בלוויה שלה נשא דברים האח של סבי, שטיפל בה במשך שנים. הוא אמר, בדמעות חנוקות, שאלוהים לקח אליו בחזרה את אמו בשלבים: תחילה את הזיכרון שלה, אחר כך את יכולת הדיבור, ולבסוף את החיוך שלה. וברגע שאלוהים לקח את החיוך, הוא ידע שמותה קרוב, שכן ללא החיוך לא נשאר ממנה כמעט דבר.

הקשבתי להספד ותמהתי, כיוון שזה גורל אכזר כל כך עבור כל *אדם שהוא*, ולכן לא הייתי מצפה מהאח של סבי לייחס אותו לאלוהים. בדרך כלל, יהודי איכשהו פשוט־לא־יחשוב על ההשלכות הלוגיות לכך שאלוהים הוא זה שאָפשר את הטרגדיה הזו. לפי התפיסה היהודית, אלוהים מחיה את העולם בכל רגע ורגע ובוחר כל אירוע שמתרחש בו. אך כדבר שבשגרה, גזירת השלכות לוגיות מהאמונה הזו שמורה לאירועים משמחים יותר. כשאתה אומר "מאת ה' היתה זאת!" רק כשנולדת לך תינוקת חדשה, ופשוט־לא־חושב "מאת ה' היתה זאת!" על הפלות, לידות שקטות ומוות בעריסה, אתה יכול לבנות לעצמך

תמונה חד־צדדית למדי של אישיותו המיטיבה של אלוהיך.

לפיכך הופתעתי לשמוע את אחיו של סבי מייחס את דעיכתה האיטית של אמו למהלך אלוהי מכוון ומתוכנן היטב. זה הפר את חוקי ההונאה־העצמית הדתית כפי שהבנתי אותם.

אילו הייתי שם לב לבלבול באותו רגע, הייתי יכול לנפק תחזית מפתיעה ומוצלחת. זמן לא רב לאחר מכן, אחיו של סבי עזב את היהדות. (היחיד במשפחתי המורחבת שעשה זאת מלבדי, למיטב ידיעתי.)

היהדות האורתודוקסית המודרנית אינה דומה לאף דת אחרת שאני מכיר, ואיני יודע איך לתאר אותה למי שמעולם לא נאלץ ללמוד משנה וגמרא. יש בה מסורת של הטלת ספק, אבל *הסוג* של הטלת הספק... זה יהיה רגיל לחלוטין לשמוע רב, בדרשתו השבועית, מצביע על הסתירה בין שבעת ימי הבריאה ל-13.7 מיליארד השנים שחלפו מאז המפץ הגדול – פשוט מפני שהוא חושב שיש לו הסבר ממש מחוכם לעניין, שכולל שלושה ציטוטים נוספים מהתנ״ך, מדרש, ומאמר מובן־למחצה מתוך כתב עת של מדע פופולרי. ביהדות האורתודוקסית מותר לך להבחין בחוסר עקביות ובסתירות, אבל אך ורק לצורך יישובן, ומי שממציא את ההסבר המסובך ביותר זוכה בפרס.

ישנה מסורת של שאילת שאלות. אבל אתה שואל שאלות רק במטרה לספק תשובות. אתה שואל רק שאלות שאתה יודע שביכולתך לענות עליהן.

ביהדות האורתודוקסית המודרנית אמונה עיוורת לא תופסת מקום מרכזי, להתרשמותי. מותר לך להטיל ספק. פשוט אסור לספק שלך *להצליח*.

אני מניח שמרבית היהודים האורתודוקסים המשכילים הטילו ספק באמונה שלהם בשלב כלשהו בחייהם. אבל הטלת הספק נראתה כנראה בערך כך: ״לפי הכופרים, התורה אינה מדויקת מבחינה מדעית שכן היא אומרת שהיקום נברא בשבעה ימים. אבל האם בני ישראל שעמדו לפני הר סיני היו מסוגלים להבין את העובדות המדעיות, גם

אילו היו מוצגות בפניהם? האם בכלל הייתה להם מילה ל'מיליארד'? הגיוני יותר לראות את סיפור הבריאה כמטאפורה – ראשית ברא אלוהים את האור, המייצג את המפץ הגדול..."

האם זוהי נקודת התורפה החלשה ביותר דרכה יכול המאמין לתקוף את אמונתו ביהדות? הַמשך עוד קצת בסיפורי התורה, ותוכל למצוא את אלוהים הורג את כל ילדי המצרים הבכורים ממין זכר כדי לשכנע פרעה לא־נבחר לשחרר עבדים, שבאופן לוגי יכלו באותה מידה להשתגר החוצה מהמדינה. יהודי אורתודוקסי ללא ספק מכיר את הפרק הזה, מכיוון שהוא אמור לקרוא את התורה כולה בבית הכנסת בכל שנה, ויש חג מרכזי שקשור לאירוע הזה. השם "פסח" מגיע מכך שאלוהים פסח על בתי היהודים בזמן שהרג כל בכור זכר במצרים.

יהודים אורתודוקסים מודרניים הם, ככלל, אנשים טובים ובני תרבות – תרבותיים בהרבה מכמה מעורכי התנ"ך. אפילו חז"ל היו תרבותיים יותר. ישנו טקס בליל הסדר שבו מטיפים עשר טיפות יין מהכוס, טיפה אחת לכל אחת מעשר המכות, כדי להדגיש את סבלם של המצרים. (כמובן, אתה אמור להזדהות עם סבלם של המצרים, אבל לא להזדהות *עד כדי כך* שתקום ותאמר, "זה לא צודק! זה *פסול* לעשות דבר כזה!") זה מראה ניגוד מעניין – חז"ל היו אנושיים מספיק בהשוואה למחברי התנ"ך כדי לראות את האכזריות שבמכות. אך המדע היה חלש יותר בימים ההם, ולכן החכמים יכלו להרהר בהיבטים הפחות נעימים של כתבי הקודש מבלי לחשוש שהדבר ינפץ את אמונתם לחלוטין.

אתה אפילו לא שואל אם התקרית הזו מטילה דופי באלוהים, ולכן אין צורך לפלוט במהירות "נסתרות דרכי האל!" או "איננו חכמים מספיק כדי לפקפק בהחלטותיו של אלוהים!" או "רצח תינוקות זה בסדר כשאלוהים עושה את זה!" החלק הזה של השאלה הוא פשוט־דבר־שלא־חושבים־עליו.

הסיבה שאנשים דתיים משכילים נשארים דתיים, כך אני חושד, היא שכאשר הם מפקפקים, הם נזהרים מאוד באופן לא מודע לתקוף

את אמונתם רק בנקודות החזקות ביותר שלה – מקומות עליהם הם יודעים שהם יכולים להגן. יתרה מכך, חזרה על אותם טיעוני הגנה מוכרים מעניקה להם תחושת התחזקות באמונה.

זה בטח מרגיש ממש טוב, למשל, לשנן את התשובה המוכנה מראש לשאלה "האם לפי המדע היקום אינו אלא אטומים חסרי משמעות שמתנגשים זה בזה?" כיוון שהיא מחזקת את תכלית הבריאה וכיצד היא נובעת מאלוהים וכו'. הרבה יותר נוח לחשוב על זה מאשר על אם מצרית אנאלפביתית הממררת בבכי מעל עריסת בנה המת. כל מי שחושב על האפשרות השניה באופן *ספונטני*, בזמן שהוא בוחן את אמונתו ביהדות, בוחן אותה *באמת*, כנראה לא יישאר דתי עוד זמן רב. המטרה שלי כאן אינה רק לחבוט ביהדות האורתודוקסית. אין לי ספק שיש תשובה כזו או אחרת למכת בכורות, וכנראה תריסר תשובות. הנקודה שלי היא שכאשר מדובר בביקורת עצמית ספונטנית, יש סיכוי גבוה בהרבה שאדם יתקוף ספונטנית נקודות חוזק שיש עמן תשובות מרגיעות לניתן לשנן, מאשר שיתקוף ספונטנית את נקודות התורפה החלשות והפגיעות ביותר. באופן דומה, סביר להניח שהוא יעצור בתשובה הראשונה ויסתפק בה, במקום להמשיך ולהקשות על התשובה עצמה. כותרת טובה יותר לפרק מ"להימנע מנקודות התורפה האמיתיות של האמונה שלך" עשויה להיות "לא לחשוב ספונטנית על החולשות הכואבות ביותר של האמונה שלך."

יותר מכל דבר אחר, אחיזתה של הדת נשמרת על ידי אנשים שפשוט־לא־חושבים על נקודות התורפה האמיתיות של הדת שלהם. אינני חושב שמדובר בעניין של חינוך, אלא באינסטינקט. אנשים אינם חושבים על נקודות התורפה האמיתיות של אמונתם מאותה סיבה שהם אינם נוגעים בגופי החימום של התנור. זה *כואב*.

כדי להצליח יותר: כשאתה מטיל ספק באחת האמונות היקרות ביותר לליבך, עצום את עיניך, רוקן את תודעתך, חשוק את שיניך, והתמקד בכוונה תחילה במה שהכי כואב. אל תחזור על קושיות סטנדרטיות שהתשובות הידועות להן יגרמו לך להרגיש טוב יותר. שאל את עצמך מה אנשים *חכמים* שחולקים עליך היו עונים לתשובה

הראשונה שלך, ולתשובה השנייה שלך. בכל פעם שאתה מזהה בעצמך רתיעה מקושיה שחלפה בראשך לרגע, גרור אותה אל קדמת תודעתך. תן לעצמך אגרוף במקלעת השמש. תקע סכין בליבך, וסובב אותה כדי להרחיב את הפצע. ניצב אל מול הכאב, שנן אך ורק זאת:

מה שאמיתי הוא כבר האמת.
להכיר בו לא ירע את המצב.
להתעלם ממנו לא יגרום לו להיעלם.
ומפני שהוא אמיתי, הוא הדבר היחיד עמו אנו באים במגע.
כל מה שלא אמיתי אינו שם כדי לחיות אותו.
בני אדם יכולים לשאת את האמת,
שהרי הם מתמודדים איתה ממילא.

— יוג׳ין גנדלין[1]

(תודה לסטפן אוּמוֹהַנדרוֹ.)

---

1. Eugene T. Gendlin, *Focusing* (Bantam Books, 1982).

75

עצירה מגמתית ומשיכת זמן מגמתית

למרות שאיני מסכים עם חלק מתפיסותיהם של אסכולת *מהיר וחסכוני** – לדעתי הם הופכים קצת *יותר מדי* לימונים ללימונדה – אני סבור שהם נוטים לפתח את המודלים *הריאליסטיים ביותר מבחינה פסיכולוגית* מבין כל האסכולות של תורת ההחלטות. רוב הניסויים בתחום מציגים בפני הנבדקים אפשרויות, הנבדק בוחר באחת מהן, וזוהי תוצאת הניסוי. ה"חסכנים" הבינו שבחיים האמיתיים, עליך *לייצר* את האפשרויות שלך בעצמך, והם חקרו כיצד הנבדקים עושים את *זה*.

באופן דומה, אף שבניסויים רבים הראיות מוגשות לך על מגש של כסף, בחיים האמיתיים צריך לאסוף ראיות בעצמך, דבר שעשוי להיות כרוך במחיר, ובשלב מסוים להחליט שכבר יש בידיך מספיק מידע כדי לעצור ולבחור. כשאתם קונים בית, אתם לא מקבלים עשרה בתים לבחור מתוכם, ואתם לא נלקחים לסיור מודרך בכל אחד מהם

* בתחום הפסיכולוגיה הקוגניטיבית ותורת ההחלטות, אסכולת *מהיר וחסכוני* (fast and frugal) טוענת שבמצבי אי־ודאות עם מידע מוגבל אנחנו משתמשים בהיוריסטיקות פשוטות ויעילות שמאפשרות קבלת החלטות מהירות ומפתיעות בדיוקן. גישה זו מציגה את ההטיות הקוגניטיביות לא כפגמים, אלא ככלים שימושיים להתנהלות בעולם האמיתי.

לפני שמרשים לכם לקבל החלטה. אתם מסתכלים על בית אחד, ואז על אחר, ומשווים ביניהם. אתם בוחנים מחדש את העדפותיכם – שוקלים עד כמה אתם באמת חייבים להיות קרובים למקום העבודה וכמה אתם באמת מוכנים לשלם. אתם מחליטים איזה בית לבדוק בשלב הבא. בשלב מסוים אתם מחליטים שראיתם מספיק בתים, ומבצעים בחירה.

ההבחנה של גילוביץ׳ בין *ספקנות מגמתית לתמימות מגמתית* מראה כיצד אנשים דורשים רף גבוה יותר ממסקנות שהם אינם רוצים להאמין בהן מאשר מסקנות שהם רוצים להאמין בהן. ספקן מגמתי שואל ״האם הראיות *מאלצות* אותי לקבל את המסקנה?״, ואילו התמים המגמתי שואל ״האם הראיות *מאפשרות* לי לקבל אותה?״

ברצוני להציע הטיה דומה עבור תהליכי חיפוש ריאליסטיים מבחינה פסיכולוגית: *עצירה מגמתית ומשיכת זמן מגמתית*. כאשר יש לנו מניע *נסתר* לבחור באפשרות ״הטובה ביותר״ עד כה, יש לנו מניע נסתר לעצור, לבחור, ולסרב לשקול אפשרויות נוספות. כאשר יש לנו מניע נסתר לדחות את האפשרות הטובה ביותר עד כה, יש לנו מניע נסתר להשהות את השיפוט עד לקבלת ראיות נוספות, לייצר עוד אפשרויות – לעשות משהו, כל דבר שהוא, *במקום* לבצע בחירה.

פרשה היסטורית מביכה במיוחד בתחום הסטטיסטיקה התרחשה כאשר רונלד פישר, מהאבות המייסדים של התחום, התעקש שלא הוכח קשר *סיבתי* בין עישון לסרטן ריאות. ״מתאם אינו סיבתיות,״ הוא העיד בפני הקונגרס. אולי למעשנים יש גֶן שגורם להם לנטייה לעשן וגם לנטייה לפתח סרטן ריאות.

או שאולי העובדה שפישר הועסק כיועץ עבור חברות הטבק היוותה מניע נסתר להחלטה שהראיות שנאספו עד כה אינן מספיקות כדי להגיע למסקנה חותכת, וכי מוטב להמשיך לבחון את הנושא. פישר אף היה מעשן בעצמו, ומת מסרטן המעי הגס ב-1962.

(הערת אד הומינם: פישר היה תדירותן.* בייסיאנים מציגים היגיון

* ההבדל בין *תדירותנים* (frequentists) לבין *בייסיאנים* (bayesians) הוא אחד מהוויכוחים הפילוסופיים והמעשיים המרכזיים בסטטיסטיקה. תדירותנים רואים

בריא יותר בכל הנוגע להסקת סיבתיות הסתברותית.)

כמו צורות רבות אחרות של ספקנות מגמתית, גם משיכת זמן מגמתית עשויה לעטות על עצמה כסות רציונלית למהדרין. מי יכול לצאת נגד איסוף ראיות נוספות? ובכן, אני יכול. ראיות הן פעמים רבות יקרות, וגרוע מכך, הן מגיעות לאט, ואין שום מעלה מיוחדת בסירוב לשקול את הראיות שכבר יש לכם. תמיד תוכלו לשנות את דעתכם בהמשך. (הסתירה המדומה מיושבת כך: הקדשת *שעה אחת* לדיון בסוגיה, בלב פתוח ונפש חפצה, זה דבר שונה לחלוטין מלהמתין עשר שנים למחקר נוסף בעלות של 20 מיליון דולר.)

באשר לעצירה מגמתית, היא מופיעה בכל מקום שבו קיים חשש מפני חלופות, בכל מקום שבו אתם מעדיפים לעצום עיניים אל מול טיעון־נגד מתבקש, ובמקומות רבים נוספים. היא מופיעה כאשר אתם פועלים בדרך שגורמת לכם להרגיש טוב מעצם העשייה, ולכן אתם מעדיפים שלא לבדוק עד כמה התוכנית שלכם באמת עובדת, שמא תהרסו את תחושת־הבטן הנעימה ששילמתם עליה במיטב כספכם. היא מופיעה בכל מקום שבו האמונות והציפיות שלכם אינן מתואמות, כך שיש לכם סיבה לחשוש מראיות חדשות.

מוסר ההשכל הוא שההחלטה לעצור תהליך חיפוש (באופן זמני או קבוע) נתונה, כמו החיפוש עצמו, להטיות ולמניעים נסתרים.

עליכם לחשוד בקיומה של עצירה מגמתית כאשר הפסקתם לחפש לאחר שהגעתם למסקנה נוחה, ועם זאת ישנן ראיות זמינות שעדיין לא בדקתם: ישנם אתרי אינטרנט לבקר בהם, ישנם טיעונים־נגד־טיעוני־הנגד שאפשר לשקול, ואפשר לעצום את העיניים למשך חמש דקות תמימות — על השעון — בניסיון לחשוב על אפשרות טובה יותר.

עליכם לחשוד בקיומה של משיכת זמן מגמתית כאשר ראיות מסוימות נוטות לכיוון שאינכם מעדיפים, אך אתם מחליטים שדרושות ראיות נוספות — ראיות יקרות שאתם יודעים שלא תוכלו לאסוף

בהסתברות תכונה אובייקטיבית ובלתי תלויה של העולם (תדירות ארוכת טווח), בעוד שבייסיאנים מבינים אותה כמידת אמונה שתלויה בידע מוקדם ובנסיבות, ולכן מתעדכנת עם מידע חדש.

בזמן הקרוב, לא משהו שביכולתכם למצוא בגוגל תוך שלושים דקות – לפני שתצטרכו להחליט משהו שאינו נוח לכם.

76

הצדקות מזויפות

נוצרים רבים שכבר איבדו את אמונתם התמימה מוסיפים להתעקש שהם רואים בתנ"ך מקור השראה מוסרי. התשובה האתאיסטית הסטנדרטית ניתנה על ידי סם האריס: "שנינו יודעים שביכולתנו לכתוב תוך חמש דקות ספר שיציע מוסר עקבי ומלא חמלה הרבה יותר מזה של התנ"ך." בדומה לכך, יש מי שיטען שהתנ"ך עודנו בעל ערך כיצירה ספרותית. אם כך, למה לא לקדש את *שר הטבעות*, יצירה ספרותית נעלה בהרבה? חרף הביקורת הידועה על המוסר ביצירתו של טולקין, *שר הטבעות* עדיין עדיף על התנ"ך כמקור מוסרי. אז למה אנשים לא עונדים טבעת קטנה על צווארם במקום צלב? אפילו *הארי פוטר* עדיף על התנ"ך, הן כיצירה ספרותית והן כמקור השראה מוסרי. אם הייתי רוצה להיות אכזרי באמת, הייתי משווה את התנ"ך לסדרת *קושיאל** של ג'קלין קארי.

"איך אתה יכול להצדיק קנייה של מחשב נייד משובץ אבני חן במיליון דולר," אתה שואל את חבר שלך, "כשלכל כך הרבה אנשים אין מחשב נייד כלל?" והחבר עונה: "אבל תחשוב על התעסוקה שזה יספק – ליצרן המחשבים, לסוכנות הפרסום שלו – ואז הם יצרכו ארוחות ותספורות – זה ימריץ את הכלכלה ובסופו של דבר אנשים

* סדרת קושיאל (Kushiel's Legacy) היא סאגת פנטזיה אפית וארוטית שגיבורתה היא קורטיזנה המפיקה עונג מכאב, בעולם שערך היסוד הדתי שלו הוא "אהוב כרצונך".

רבים ישיגו מחשב נייד משלהם!" כן, אבל יעיל *הרבה יותר* לקנות 5,000 מחשבים במסגרת פרויקט "מחשב לכל ילד", וכך גם לספק תעסוקה ליצרנים וגם לחלק מחשבים ישירות.

כבר עסקתי בכשל של אי־חיפוש אחר החלופה השלישית.* אבל כאן לא מדובר באמת ב"עצירה מגמתית". לקרוא לזה כך היה מרמז שחיפוש כלשהו בכלל התבצע מלכתחילה.

בפרק 70 ציינתי שרק הסיבות האמיתיות שבבסיס האמונות שלנו יכולות להשפיע על מידת הדיוק שלנו בעולם האמיתי, ורק הסיבות האמיתיות שבבסיס הפעולות שלנו יכולות להשפיע על האפקטיביות שלנו בהשגת מטרותינו. מי שקונה מחשב נייד במיליון דולר חשב לעצמו "אווו, נוצץ!" וזו הייתה ההשתלשלות הסיבתית האמיתית היחידה בבסיס ההחלטה שלו. אף כמות של "הסברים" לא יכולה לשנות זאת, אלא אם כן ההסברים מגיעים מתוך תהליך חיפוש אמיתי וחדש, שביכולתו לשנות את המסקנה. לשנות את המסקנה *באמת*. רוב הביקורת העצמית שמתבצעת מתוך תחושת חובה היא מראית עין ותו לא. בחירות חופשיות במדינה בעלת מפלגה אחת.

כדי באמת להצדיק את התנ"ך כמושא להערצה בהסתמך על איכותו הספרותית, היית צריך איכשהו לבצע קריאה ניטרלית של ספרים נוספים עד שהיית מוצא את הספר בעל האיכות הספרותית הגבוהה ביותר. מוניטין הוא קריטריון סביר לסינון מועמדים, אז אני מניח שיהיה זה לגיטימי לגמרי למצוא את עצמך קורא את שייקספיר, התנ"ך, וגדל, *אשר, באך*. (שאם לא כן, היה זה צירוף מקרים מופלא למצוא את התנ"ך כאחד המועמדים, מתוך מיליון ספרים אחרים.) הקושי האמיתי הוא בחלק של "הקריאה הניטרלית". זה קל למדי אם אינך נוצרי, אבל אם אתה כן...

אך כמובן, דבר כזה מעולם לא קרה. שום חיפוש לא התרחש. כתיבת ההצדקה "איכות ספרותית" מעל השורה התחתונה "אני ♥ תנ"ך" היא מצג שווא היסטורי של האופן שבו השורה התחתונה הזו באמת

* ראו פרקים 47 ו-75.

הגיעה לשם – כמו למכור חלב חתולים בתור חלב פרה. זה פשוט לא האופן שבו השורה התחתונה נכתבה. זה פשוט לא מה שקרה כדי לייצר את המסקנה הזו.

אם אתה מוכן לבקר את המסקנה שלך באופן שבאמת יכול לבטל אותה – אם לביקורת *באמת* יש את הכוח הזה – אז זה אכן משנה את "האלגוריתם האמיתי" שמאחורי המסקנה. השזירה בין המסקנה שלך לבין עולמות אפשריים משתנה. אבל אנשים מעריכים הערכת יתר פרועה את הסיכוי שהם באמת ישנו את דעתם.

עם כל כך הרבה "ראשים פתוחים" שם בחוץ, היית מצפה לראות קצת יותר עדכוני אמונות.

תן לי לנחש: כן, אתה מודה שמה שגרם לך מלכתחילה לקנות מחשב נייד במיליון דולר זו המחשבה "אווו, נוצץ!" כן, אתה מסכים שתהליך קבלת ההחלטות הזה לא עולה בקנה אחד עם המטרות המוצהרות שלך. אלא שמאז הגעת למסקנה שעליך להשקיע את כספך באופן שיספק מחשבים ניידים לכמה שיותר אומללים מחוסרי מחשב, ופשוט *לא הצלחת* למצוא אף דרך יעילה יותר לעשות זאת מאשר קניית מחשב נייד משובץ יהלומים במיליון דולר – כי, היי, אתה מזרים כסף לחנות המחשבים וממריץ את הכלכלה! מה יותר טוב מזה?

ידידי, יש לי חשד עמוק לגבי צירוף המקרים המופלא הזה. אני חושד עמוקות בכך שהפתרון המושלם לפי קריטריון **א** הנהדר, הרציונלי והאלטרואיסטי, הוא בדיוק אותו הרעיון שנולד במקור מתוך תהליך **ב** הלא־מוצדק והבלתי־תלוי. אם אינך חושב שהתשובה הנכונה הייתה יכולה לנבוע מהטלת קוביה, מדוע אתה חושב שהיא תצוץ מתוך דפוס חשיבה לא־רציונלי אחר?

אין זה סביר שההגיון שהפעלת פגום, ובכל זאת הגעת לתשובה הנכונה.

77

האם זו ההתנגדות האמיתית שלך?

מדי פעם זה קורה: מישהו נתקל בכמה מהאמונות שלי שקשורות לצד הטרנס־הומניסטי* – בניגוד לרעיונות שלי שעוסקים ברציונליות אנושית – רעיונות מוזרים שנשמעים אקזוטיים, כמו אינטליגנציית־על ובינה מלאכותית ידידותית. והוא דוחה אותם.

אם מבקשים ממנו להסביר את הדחייה, לא פעם הוא יענה: "למה שאאמין למשהו שיודקובסקי אומר? אין לו אפילו דוקטורט!"

ולעיתים מישהו אחר שומע זאת ואומר: "הו, באמת כדאי שתעשה דוקטורט, כדי שאנשים יקשיבו לך." לפעמים העצה הזו מגיעה דווקא מאותו אדם שלא השתכנע, שאומר: "תחזור אליי כשיהיה לך דוקטורט."

ובכן, יש סיבות טובות וסיבות לא טובות לעשות דוקטורט. זו לא אחת מהסיבות הטובות.

ישנן סיבות *אמיתיות* רבות לכך שאדם יפתח תגובת נגד לרעיונות טרנס־הומניסטיים. רובן קשורות לזיהוי דפוסים, ולא למחשבה מילולית: התזה מקוטלגת מיד תחת "רעיון מוזר ותמוה", "מדע בדיוני", "כת סוף־העולם" או "צעיר נלהב מדי".

כך שבאופן מיידי, בהרף עין של רושם ראשוני, הרעיון נדחה. אם

* *טרנס־הומניזם* הוא גישה פילוסופית הומניסטית המנסה להבין ולהעריך את האפשרויות להתגבר על מגבלותיו הביולוגיות של האדם באמצעות קדמה טכנולוגית (ויקיפדיה).

לאחר מכן מישהו ישאל "למה דחית את הרעיון?" יתחיל החיפוש אחר הצדקה, אך הוא לא בהכרח יקלע לסיבה האמיתית – וב"סיבה אמיתית" אינני מתכוון לטיעון הטוב ביותר שאפשר למצוא, אלא לאותם גורמים שהיו מכריעים מבחינה עובדתית *באותו הרגע* שבו התרחשה הדחייה.

במקום זאת, החיפוש אחר הצדקה יאתר עובדה שנשמעת משכנעת: "לדובר הזה אין דוקטורט."

אבל אין לי דוקטורט גם כשאני מדבר על רציונליות אנושית, אז איך זה שאותה הטענה לא עולה גם שם?

יתרה מכך: אילו *היה* לי דוקטורט, אנשים לא היו מחשיבים אותו כגורם מכריע שבגללו הם אמורים להאמין לכל מה שאני אומר. במקום זאת, אותה דחייה ראשונית הייתה מתרחשת, מאותן סיבות ממש – והחיפוש אחר הצדקה, המתרחש לאחר מכן, היה נעצר בנקודה אחרת.

הם היו אומרים: "מדוע עלי להאמין לך? אתה סתם עוד דוקטור! יש המון כאלה. תחזור אליי כשתהיה מוכר בתחומך ובעל קביעות באוניברסיטה גדולה."

אבל האם אנשים *באמת* מאמינים לכל פרופסור אקראי מהרווארד שאומר דברים מוזרים? ברור שלא. (אם כי אילו הייתי פרופסור בהרווארד, היה לי קל יותר למשוך תשומת לב *תקשורתית*. עיתונאים שלא יאמינו לי בגלל רושם ראשוני – וכנראה לא היו מאמינים לשום דוקטור אקראי – עדיין יכתבו עליי בעיתון, מהסיבה הפשוטה שזה אייטם חדשותי, שפרופסור מהרווארד מאמין בדבר מוזר כל כך.)

אם אתה אומר דברים שנשמעים *שגויים* לאדם מן השורה, להבדיל מסתם ג'בריש־טכנולוגי שנשמע כמו קסם על "צמות קווארקים לֶפְּטיות במימד N+2"; ואותו אדם אינו מכיר אותך אישית *וגם* לא בקי בתחום; אז יש לי חשד שהנקודה שבה האדם הממוצע *באמת* יתחיל לתת לדבריך משקל שיגבר על הרושם הראשוני שלו, אך ורק *בזכות* תארים אקדמיים, נמצאת איפשהו סביב חתן פרס נובל. אם בכלל. בגדול, אתה צריך תואר אקדמי ברמה שמוגדרת כ"לא מהעולם הזה".

זה פחות או יותר מה שקרה לאֶריק דרֶקסלֶר, להבנתי. הוא הציג את החזון שלו לננוטכנולוגיה, ואנשים אמרו: "איפה הפרטים הטכניים?" או "תחזור כשיהיה לך דוקטורט!" אז אריק דרקסלר בילה שש שנים בכתיבת הפרטים הטכניים ועשה דוקטורט בהנחייתו של מרווין מינסקי. והספר *Nanosystems* הוא ספר נהדר. אבל האם אותם אנשים שאמרו "תחזור כשיהיה לך דוקטורט" שינו את דעתם לגבי ננוטכנולוגיה מולקולרית? לא ככל הידוע לי.

אותו הכלל חזר על עצמו במכון לחקר בינה מלאכותית (MIRI): לא משנה מה מייעצים לנו לעשות כדי להיות אמינים יותר, גם כשאנחנו באמת עושים זאת דבר לא משתנה. "האם אתם מפתחים קוד כלשהו? אני לא מעוניין לתמוך בארגון שלא מפתח קוד" ← OpenCog ← דבר לא משתנה. "לאליעזר יודקובסקי אין תארים אקדמיים" ← פרופסור בן גֶ'רֶצְל מונה למנהל המחקר ← דבר לא משתנה. הדבר היחיד *שבאמת* נראה שחיזק את שמו של המכון שלנו הוא אנשים מפורסמים ששמם נקשר בו, כמו פיטר תִיל שמימן אותנו, או רֵיי קוּרצְוַוייל שהצטרף למועצת המנהלים.

זה עשוי להיות שיעור חשוב לעסקים צעירים או ליועצים מתחילים: יתכן שהמשוב שאתם מקבלים מלקוחות פוטנציאליים שדחו אתכם אינו נוגע לסיבה שבאמת תשנה משהו. כדאי לשקול זאת היטב לפני שמשקיעים מאמצים משמעותיים עבור הלקוח הבא. אם המשקיע אומר "אילו רק המכירות שלכם היו צומחות קצת יותר מהר!" או אם הלקוח הפוטנציאלי אומר "המוצר נראה טוב, אבל חסר לו פיצ'ר מסוים," ייתכן שזו לא הסיבה *האמיתית* לדחיה. תיקון הבעיה עשוי לשנות משהו – ועשוי שלא.

כדאי לזכור זאת גם בזמן ויכוחים. רובין הַנסון ואני שותפים לאמונה ששני רציונליסטים לא אמורים להסכים שלא להסכים: בסיס־הידע שלהם לא אמור להיות זהה בעת מחלוקת לגבי המציאות אלא אם כן משהו השתבש לחלוטין.

אני חושד שבאופן כללי, אם שני רציונליסטים ניגשים לפתור מחלוקת שנמשכת מעבר לחילופי הדברים הראשונים, עליהם לצפות

לגלות שיסודות המחלוקת האמיתיים הם או קשים להעברה, או קשים לחשיפה. למשל:

- ידע מדעי או מתמטי לא נפוץ, אך מבוסס היטב.
- מרחקי הסקה ארוכים.
- אינטואיציות שקשה לנסח במילים, שמקורן אולי בדימויים חזותיים מסוימים.
- רוח־זמן (צַייטגַייסט) שנספגה מסביבה מקצועית (שייתכן שיש לה סיבה טובה).
- זיהוי דפוסים שמקורו בהתנסות חווייתית.
- הרגלי מחשבה טהורים.
- מחויבות רגשית להאמין בתוצאה מסוימת.
- פחד מחשיפת טעות עבר.
- הונאה עצמית עמוקה בשל גאווה או לשם תועלת אישית אחרת.

אילו היה מדובר בנושא שבו ניתן להניח בקלות את כל ההתנגדויות האמיתיות על השולחן, היה כנראה כל כך קל לפתור את המחלוקת שהיא מעולם לא הייתה שורדת את המפגש הראשון.

"האם זו ההתנגדות האמיתית שלי?" היא שאלה ששני הצדדים במחלוקת בהחלט צריכים לשאול *את עצמם*, כדי להקל על הצד השני. יחד עם זאת, ניסיונות לנתח פסיכולוגית את האחר באופן ישיר ופומבי עלולים לגרום לשיחה להידרדר מהר *מאוד*, מניסיוני.

ובכל זאת — "האם זו ההתנגדות האמיתית שלך?" צריכה להיות שאלה לגיטימית שצדדי המחלוקת יכולים לשאול בענווה, אם יש דרך פרודוקטיבית לחקור את תת־הסוגיה הזו. אולי הכלל יכול להיות שמותר לשאול בגלוי: "האם הסיבה הפשוטה והישירה הזו היא ההתנגדות האמיתית שלך, או שהיא נובעת מאינטואיציה **א** או מרוח הזמן המקצועית **ב**?" בעוד שהאפשרויות המביכות יותר ברשימה יישארו למצפונו של האחר, כנושא שבאחריותו הבלעדית לטפל בו.

78

אמיתות שזורות, שקרים מדבקים

״אחד מקדמוני הפילוסופים שלכם הגיע למסקנה שמוח מוכשר מספיק יוכל, מתוך חקר של עובדה אחת או חפץ אחד השייכים ליקום נתון כלשהו, לבנות או לדמיין את אותו יקום, מרגע בריאתו ועד לסופו המוחלט...״

— *העדשאי הראשון*[1]

״אם מי מכם יתרכז בעובדה בודדת אחת, או בחפץ קטן, כמו חלוק נחל או זרע של צמח או יצור אחר, למשך תקופה קצרה בת כמאה שנים שלכם, אז יתחיל להבין אותו לאמיתו.״

— *העדשאי האפור*[2]

אני די בטוח שחלוק נחל בודד, שנלקח מחוף כלשהו על פני כדור הארץ שלנו, לא כולל מידע חד־משמעי על היבשות והמדינות, הפוליטיקה והאנשים שעל פני כדור הארץ הזה. כוכבי לכת אחרים במרחב ובזמן, ענפי אֶוורֶט אחרים, עשויים לייצר את אותו חלוק נחל בדיוק. מצד שני, נראה כי חלוק נחל בודד אכן כולל בתוכו את חוקי הפיזיקה שלנו. במובן זה, היקום שלנו בשלמותו — *כל* ענפי אוורט — נגזר בהכרח מקיומו של חלוק הנחל (זאת אם, כפי שסביר להניח, באמת אין משתנים חופשיים).

אז סביר להניח שחלוק נחל בודד לא צופן בחובו מידע על כל כדור הארץ שלנו. אבל בכל זאת ניתן ללמוד ממנו רבות. מתוך חקירת חלוק הנחל הזה תוכלו ללמוד על חוקי הפיזיקה וכל המשתמע מהם. כשאתם בוחנים את חוקי הפיזיקה, תוכלו להסיק שכוכבי לכת יווצרו, ותוכלו לנחש שחלוק הנחל הגיע מכוכב לכת כזה. הגבישים הפנימיים והמבנים המולקולריים של חלוק הנחל נוצרו תחת השפעת כוח הכבידה, מה שמלמד על המסה של אותו כוכב לכת, בעוד שתמהיל היסודות שמהם מורכב חלוק הנחל מספר משהו על האופן שבו אותו כוכב לכת נוצר.

אני לא גאולוג, אז אינני בקי בסודות המקצועיים של הגאולוגים. אבל קל לי מאוד לדמיין סיטואציה שבה אני מראה לגאולוג חלוק נחל ואומר לו, "החלוק הזה הגיע מהחוף ליד ביתי בקליפורניה," והגאולוג מיד אומר, "אני מבולבל" או אפילו "אתה משקר." אולי סוג הסלע לא נכון, או שהוא אינו שחוק כמו חלוקים שהגיעו מחוף הים – אני לא מכיר חלוקי נחל מספיק טוב כדי לנחש מהם ההקשרים והסממנים שעלולים להסגיר אותי, וזו בדיוק הנקודה.

"רק אלוהים מסוגל לספר שקר אמין באמת." מעניין אותי אם הייתה אי פעם דת שאימצה את המשפט הזה כפתגם. הייתי מהמר (הימור שניתן להפרכה) שלא: זו תפיסה רציונליסטית במהותה, גם אם מלבישים אותה במטאפורה תאולוגית. לומר ש"הכל קשור להכל, מכיוון שאלוהים ברא את כל העולם ומקיים אותו," אולי יגרום לתחושת חמימות נעימה בלב במהלך הדרשה, אבל זה לא יעזור לכם במיוחד כשתצטרכו לשייך חלוקי נחל לחופים.

מטבע בודד על פני כדור הארץ מפעיל על הירח תאוצת כבידה של בערך 4.5×10^{-31} מטרים לשנייה בריבוע, כך שבמובן מסוים זה לא יהיה שגוי לחלוטין לומר שכל אירוע שזור עם כל חרוט האור של העבר שלו.* ומכיוון שהיקשים לוגיים יכולים להתפשט קדימה

* בתורת היחסות, מהירות האור היא המהירות המקסימלית שבה אירוע יכול להשפיע על אירוע אחר. לכן, רק אירועים שהמרחק הפיזי ביניהם קטן או שווה למרחק שהאור מסוגל לעבור בפרק הזמן שמפריד ביניהם, יכולים להשפיע אחד

ואחורה ברשתות סיבתיות,* שזירות *אפיסטמיות* בהחלט יכולות לחצות את הגבולות של חרוטי האור.** אבל לא הייתי רוצה להיות האסטרונום־חוקר שצריך להסתכל על הירח ולהבין אם המטבע נחת על עץ או על פלי – ההשפעה קטנה בהרבה מרמת אי־הוודאות הקוונטית והרעש התרמי.

אם הייתם אומרים "כל דבר שזור בדבר אחר," או "הכל שזור ברמת ההיקש הלוגי, ושזירות מסוימות חזקות משמעותית מאחרות," אולי הייתם אנשים חכמים באמת, ולא כאלו שרק מנסים להישמע עמוקים.

מבחינה פיזיקלית, כל אירוע הוא במובן מסוים הסכום של כל חרוט האור שקדם לו, ללא גבולות או הגבלות. אולם רשימת השזירות *שניתן להבחין בהן* קצרה הרבה יותר, והיא יוצרת משהו שדומה לרשת. החוקיות הזו, ברמת ההפשטה הגבוהה שלה, היא מה שאליו אני מתכוון כשאני מדבר על *רשת הסיבתיות הגדולה*.

אני מעניק למונח הזה נופך דרמטי באופן קצת מוגזם, אולי, אבל אם יש איזשהו ביטוי שראוי שייכתב *באותיות נטויות*, אין ספק *שרשת הסיבתיות הגדולה* נכנס לרשימה.

"הו, מה סבוכה הרשת אותה אנו טווים, עת פוצחים אנו במלאכת השקרים," אמר סר וולטר סקוט. לא *כל* השקרים יוצאים משליטה – אנחנו לא חיים ביקום עד כדי כך צודק. אבל בהחלט קורה מדי פעם

על השני. *חרוט האור של העבר* הינו כלל אירועי העבר שמיקומם במרחב־זמן מאפשר להם להשפיע על האירוע הנוכחי, ו*חרוט האור של העתיד* הינו אירועי העתיד שביכולת האירוע הנוכחי להשפיע עליהם.

* *רשתות סיבתיות* מתארות סיבה ותוצאה. בניגוד להשפעה פיזיקלית, הפועלת אך ורק קדימה בזמן, היקשים לוגיים יכולים לנוע גם אחורה – ניתן לראות תוצאה (אפר) ולהסיק ממנה לאחור על הסיבה (אש).

** *אפיסטמי* פירושו קשור לידע או להכרה. בניגוד לשזירה פיזיקלית (כמו כוח הכבידה של המטבע שמשפיע פיזית על הירח), שזירה אפיסטמית מתארת מצב שבו שני דברים קשורים זה לזה ברמת המידע, ושזירה שכזו אינה מוגבלת לחרוטי האור. לדוגמה, שני חצאי מטבע נשלחים כל אחד בגשושית לקצוות מנוגדים של היקום. כרגע, שתי הגשושיות נמצאות מחוץ לחרוט האור אחת של השנייה. אם אסטרונאוט פותח את הגשושית הראשונה ורואה שכתוב על החצי שלו "עץ", הוא יודע *באותו רגע בדיוק* שעל החצי שנמצא בקצה השני של היקום כתוב "פָּלי".

שמישהו משקר לגבי עובדה כלשהי, ואז נאלץ לשקר לגבי עובדה ששזורה בה, ואז לגבי עובדה נוספת ששזורה בקודמתה:

"איפה היית?"
"אה, הייתי בנסיעת עסקים."
"באיזה עסקים מדובר?"
"אני לא יכול לספר לך, מדובר במשא ומתן חסוי מול לקוח מרכזי."
"וואו – הם מכניסים אותך לעניינים האלה? איזה יופי! אני צריך להתקשר לבוס שלך כדי להודות לו שצירף אותך."
"סליחה – הוא לא במשרד כרגע..."

בני אדם, שאינם אלים, כושלים לעיתים קרובות *מלדמיין* את כל העובדות שיצטרכו לסלף כדי לספר שקר שיהיה אמין באמת. "אלוהים הכניס אותי להיריון" נשמע קצת יותר סביר בימים ההם, לפני שהמודלים שלנו אודות העולם כללו (ייצוגים של) כרומוזומי Y. שקרים דומים רבים בימינו עלולים להיחשף ככל שבדיקות גנטיות יהפכו לנפוצות יותר. אנסים הורשעו וטופלי אשמות שווא נחשפו שנים רבות לאחר המעשה, על סמך ראיות שהם כלל לא תיארו לעצמם שהם משאירים מאחור. סטודנטים לביולוגיה אבולוציונית מסוגלים לראות את טביעות האצבע של הברירה הטבעית בכל זאב שרודף אחרי ארנב, ובכל ארנב שבורח, ובכל דבורה שעוקצת במקום לשדר אזהרה מנומסת – אבל השקרים של הבריאתנים נשמעים אמינים *בעיני עצמם*, אני בטוח בכך.

לא כל השקרים מתגלים, ולא כל השקרנים נענשים – אנחנו לא חיים ביקום עד כדי כך צודק. אולם, לא כל השקרים בטוחים כפי ששקרנים נוטים להאמין. תהיתי לעצמי, כמה חטאים היו נחשפים בפני אינטליגנציית־על בייסיאנית אילו הייתה מבצעת סריקה ננוטכנולוגית (לא־הרסנית?) של כדור הארץ? לכל הפחות, כל השקרים שראיה כלשהי לקיומם עדיין נותרה במוח כלשהו. חלק מהשקרים הללו עשויים להיחשף עוד קודם לכן, אם חוקרי מוח יצליחו אי פעם לבנות פוליגרף מוצלח באמת המבוסס על דימות מוחי. פּוֹל אֶקמַן (חלוץ

בחקר תנועות זעירות של שרירי הפנים) כנראה מסוגל כבר עכשיו לפענח חלק ניכר מהשקרים בעולם, אם רק תיתנו לו הזדמנות.

לא כל השקרים מתגלים, ולא כל השקרנים נענשים. ואף על פי כן, אנשים נוטים לזלזל לעיתים קרובות מדי בכוחה של *הרשת הגדולה*. כדי ללמוד רק את הידע שהאנושות הספיקה לצבור עד כה, יידרשו לאדם תקופות חיים רבות. כל מי שחושב שישות שאינה אלוהית מסוגלת לספר שקר *מושלם*, נטול סיכונים, פשוט אינו מעריך נכונה את הסבך העצום של *הרשת הגדולה*.

האם כנות היא האסטרטגיה הטובה ביותר? אני לא יודע אם הייתי מרחיק לכת עד כדי כך: אפילו על פי אמות המוסר שלי, לפעמים זה בסדר גמור פשוט לשתוק. אבל בהשוואה לשקרים בוטים, גם כנות וגם שתיקה כרוכות בפחות חשיפה לסיכונים המתפשטים רקורסיבית, כאלו שאתם בכלל לא מודעים לכך שאתם לוקחים.

---

1. Edward Elmer Smith and A. J. Donnell, *First Lensman* (Old Earth Books, 1997).
2. Edward Elmer Smith and Ric Binkley, *Gray Lensman* (Old Earth Books, 1998).

79

על שקרים וקטסטרופות ברבור שחור

השופט מרכוס אָיינפֶלד, בן 70, בעל התואר ׳פרקליט המלכה׳ (QC) מאז 1977, הוכר כ׳אוצר לאומי חי׳ של אוסטרליה לשנת 1997, חתן פרס השלום של האו״ם לשנת 2002, והנשיא המייסד של הוועדה לזכויות אדם ושוויון הזדמנויות של אוסטרליה, אשר פרש לפני מספר שנים אך עודנו מזומן באופן שגרתי לשפוט בתיקים חשובים...

...נשלח למאסר של שנתיים בגין שרשרת של עדויות שקר וכזבים, שהחלה בדו״ח תנועה על סך 77 דולר אוסטרלי בגין חריגה של 10 קמ״ש מהמהירות המותרת.

לתיאוריה הזו, שנשמעת *צדקנית להחשיד*, על כך שאנשים ישרים אינם מוצלחים בלשקר, ושעקבות שזורים נותרים אי־שם, ושהעסק כולו מתפוצץ לבסוף בכישלון אֶפִּי של ברבור שחור – *יש* למעשה מספר דוגמאות בחיים האמיתיים, אף שברור שדיווח סלקטיבי משַחק תפקיד בעצם העובדה ששמענו על המקרה המסוים הזה.

80

אפיסטמולוגיה של הצד האפל

אם סיפרתם שקר ולו פעם אחת, האמת תהיה אויבתכם לעד.

כתבתי לפני שני פרקים על התפיסה שלפיה, כיוון שהאמת שזורה, שקרים הם מדבקים. אם תרימו חלוק נחל משביל הגישה שלכם ותאמרו לגאולוג שמצאתם אותו על שפת הים – ובכן, האם *אתם* יודעים מה גאולוג יודע על אבנים? אני לא. אבל אני נוטה להניח שחלוק נחל שנשחק על ידי מים לא ייראה כמו טיפת לַבָּה קפואה מהתפרצות געשית. האם אתם יודעים מאין באמת הגיע חלוק הנחל שבשביל הגישה שלכם? דברים נושאים את חותם מקומם ביקום הפועל על פי חוקים. ברשת הזו, שקר הוא נטע זר. (למעשה, גאולוג אחד ציין בתגובות שרוב חלוקי הנחל בשבילי גישה *נלקחים מחופי ים*, כך שהם לא יוכלו להבחין בין חלוק נחל של שביל גישה לזה של חוף הים, אבל הם יוכלו להבחין בין חלוק נחל הררי לחלוק נחל של שביל או חוף. וזו בדיוק הנקודה...)

מה שנשמע לאדם אחד כאמת שרירותית – כזו שניתן בקלות להחליף בשקר סביר – עשוי להיות מעוגן היטב על ידי תריסר זיקות בעיניו של בעל ידע נרחב יותר. עבור בריאתן, הרעיון שהחיים עוצבו על ידי "תכנון תבוני" במקום "ברירה טבעית" עשוי להישמע כמו קבוצת ספורט לעודד. עבור ביולוג, טענה סבירה לפיה אורגניזם תוכנן בתבונה תדרוש לשקר כמעט לגבי כל היבט של אותו אורגניזם. כדי לטעון באופן אמין ש"בני אדם" תוכננו בתבונה, תצטרכו לשקר לגבי

העיצוב של הרשתית האנושית, הארכיטקטורה של המוח האנושי, החלבונים המוחזקים יחד על ידי כוחות ַאן דֶר וָאלְס חלשים במקום קשרים קוולנטיים חזקים...

או שתוכלו פשוט לשקר בנוגע לתורת האבולוציה, שזהו הנתיב שבו בוחרים רוב הבריאתנים. במקום לשקר לגבי הצמתים המחוברים ברשת, הם משקרים לגבי החוקים *הכלליים* השולטים בקשרים.

ואז, כדי לחפות *על כך*, הם משקרים לגבי כללי המדע – למשל מה המשמעות של כינוי דבר מה "תיאוריה", או מה פירוש הדבר כשמדען אומר שאין לו ודאות מוחלטת.

כך הם עוברים מִשקרים על עובדות ספציפיות, לשקרים על חוקים כלליים, לשקרים על כללי החשיבה הרציונלית. כדי לשקר בשאלה האם בני אדם התפתחו אבולוציונית, עליכם לשקר לגבי האבולוציה – ואז עליכם לשקר לגבי כללי המדע התוחמים את הבנתנו את האבולוציה.

אך כיצד אפשר אחרת? בדיוק כפי שאדם יהיה נטע זר בתוך קהילה של צורות חיים *שבאמת* תוכננו בתבונה, ויהיה עליכם לשקר לגבי חוקי האבולוציה כדי לטשטש עובדה זו – כך גם אמונות לגבי בריאתנות, אין מקומן בתוך המדע. לא תמצאו אותן במוח מסודר כהלכה, בדיוק כפי שלא תמצאו עצי דקל צומחים על קרחון. ולכן עליכם להרוס את המחסומים שהיו מונעים את כניסתן.

מה שמוביל אותנו לנושא של הונאה עצמית.

שקר בודד שאתם מספרים *לעצמכם* עשוי להיראות סביר למדי, כאשר אינכם מכירים אף אחד מהכללים השולטים במחשבות, או אפילו את עצם *קיומם* של כללים אלו. הבחירה נראית שרירותית כמו בחירת טעם של גלידה, מבודדת כמו חלוק נחל על שפת הים...

...אבל אז מישהו קורא תיגר על אמונתכם, תוך שימוש בכללי החשיבה *שהוא* למד. הוא שואל, "איפה הראיות שלכם?"

ואתם עונים, "מה? למה אני צריך ראיות?"

אז הוא אומר, "באופן כללי, אמונות דורשות ראיות."

הטיעון הזה, בבירור, הוא חייל הנלחם עבור האויב, ועליכם להביס

אותו. לכן אתם אומרים: "אני לא מסכים! לא כל האמונות דורשות ראיות. בפרט, אמונות לגבי דרקונים לא דורשות ראיות. כשמדובר בדרקונים, מותר לך להאמין במה שרק תרצה. אז אינני צריך ראיות כדי להאמין שיש דרקון במחסן שלי."

והוא משיב, "מה? אי אפשר פשוט להתייחס אחרת לדרקונים. יש סיבה לכלל שלפיו אמונות דורשות ראיות. כדי לשרטט מפה נכונה של העיר, עליך ללכת ברחובות ולסמן על הנייר קווים התואמים את מה שעיניך רואות. זו אינה דרישה חוקית שרירותית – אם תשב בסלון ביתך ותשרטט קווים אקראיים על הנייר, המפה תהיה שגויה, בהסתברות גבוהה ביותר. זה נכון לגבי מפה של דרקון באותה מידה שזה נכון לגבי כל דבר אחר."

אז עכשיו ההסבר *הזה* – *מדוע* אמונות דורשות ראיות – הוא גם חייל אויב. לכן אתם אומרים: "שגויה בהסתברות גבוהה ביותר? אז עדיין יש סיכוי, נכון? אני לא חייב להאמין בתיאוריה שלך אם היא אינה ודאית לחלוטין."

או שאולי אתם אפילו מתחילים לחשוד, בתוככם, שמא אכן "אמונות דורשות ראיות". אך הדבר מאיים על שקר שׂנֵא ללבכם. לכן אתם בולמים את אור השחר העולה בקרבכם, דוחקים את השמש חזרה אל מתחת לאופק.

או ששמעתם בעבר את המשפט "אמונות דורשות ראיות", והוא נשמע לכם נכון למדי, ואמרתם זאת בפומבי. אבל מעולם לא באמת עלה בדעתכם שהמשפט הזה עשוי *לחול* על אמונתכם שיש דרקון במחסן שלכם, עד שמישהו אחר הפנה את תשומת לבכם לכך. אז אתם חושבים מהר ואומרים, "אמונה בדרקון נמצאת בתחום ידע נפרד."

להחזיק באמונות שגויות אינו דבר טוב, אך אין זה בהכרח בלתי הפיך – אם, כאשר אתם מגלים את טעותכם, אתם מתגברים עליה. הדבר המסוכן הוא להחזיק באמונה שגויה *שאתם מאמינים שיש להגן עליה כאמונה* – אמונה באמונה, בין אם מאחוריה קיימת אמונה ממשית ובין אם לאו.

די בשקר אחד ש״חובה להגן עליו״ כדי לחסום את התקדמותו של אדם לעבר רציונליות מתקדמת. לא, זה לא שעשוע תמים.

בדיוק כפי שהעולם עצמו סבוך בהרבה מכפי שהוא נראה על פני השטח, כך גם ישנם כללי חשיבה ברורים, המכתיבים את האמונה בעוצמה רבה יותר מכפי שהאדם הלא־מיומן עשוי לשער. העולם ארוג בצפיפות, נשלט על ידי חוקים כלליים, וכך גם אמונות *רציונליות*.

חשבו מה יידרש כדי להכחיש את האבולוציה או את המודל הֶהֶליוֹצֶנטרי* – כל האמיתות השלובות וחוקי היסוד שייאסר עליכם לדעת. או אז תוכלו לדמיין כיצד מעשה יחיד של הונאה עצמית יכול לחסום את כל רובד המטא־חתירה־לאמת, מהרגע בו דעתכם מתחילה לחוש מאוימת מראיית הַקשרים. לאסור על כל רמות הביניים והרמות הגבוהות של אומנות הרציונליסט. ליצור, במקומן, מערכת עצומה של אנטי־חוק, כללים של אנטי־מחשבה, והצדקות כלליות להאמין בשקרים.

סטיבן קָאס אמר, ״הפצת אמונות שאינן מדויקות בתכלית היא מעשה חבלה. אל תעשו זאת לאיש אלא אם הייתם גם חותכים לו את הצמיגים.״ לספק למישהו אמונה שגויה *להגן עליה* – לשכנע אותו שיש להגן על *האמונה עצמה* מפני כל מחשבה שעשויה לאיים עליה – ובכן, אל לכם לעשות זאת לאיש אלא אם הייתם גם מבצעים בו לובוטומיה פרונטלית.

מהרגע שסיפרתם שקר, האמת היא אויבתכם. וכל אמת נוספת שקשורה לאותה האמת, וכל בעלת ברית של האמת בכלל. לכל אלה עליכם להתנגד, כדי להגן על השקר. בין אם אתם משקרים לאחרים, ובין אם לעצמכם.

עליכם להכחיש שאמונות דורשות ראיות, ואז עליכם להכחיש שמפה צריכה לייצג את השטח, ואז עליכם להכחיש שהאמת היא דבר טוב...

---

* *המודל ההֶליוֹצֶנטרי* הוא המודל האסטרונומי שפיתח קוֹפֶרניקוּס (ומאוחר יותר ביססו גָלִילֵאו וקֶפלֶר), לפיו כדור הארץ ושאר כוכבי הלכת סובבים סביב השמש, בניגוד למודל הגֵאוֹצֶנטרי הקדום שהציב את כדור הארץ במרכז היקום.

כך בא לעולם *הצד האפל*.

אני חושש שאנשים אינם מודעים לכך, או שאינם נזהרים די הצורך – שכאשר אנו משוטטים בעולמנו האנושי, אנו צפויים להיתקל בתורות־הכרה *גרועות באופן שיטתי*.

השיטות של "איך לחשוב" שמסתובבות להן, מחשבות־המטמון* של "חוכמה עמוקה" – חלק מהן יהיו עצות טובות שהגו רציונליסטים. אך אחרות הומצאו כדי להגן על שקר או הונאה עצמית: הן נולדו מתוך *הצד האפל*.

"לכל אחד יש זכות לדעה משלו." כשחושבים על זה, היכן נוצרה האמרה הזו? האם זה משפט שמישהו היה אומר במהלך הגנה על האמת, או במהלך הגנה *מפני* האמת? אבל אנשים לא מזדקפים ואומרים, "אהא! אני חש בנוכחותו של *הצד האפל*!" למיטב התרשמותי, לא רבים מודעים לכך *שהצד האפל* נמצא שם בחוץ.

אך כיצד יתכן אחרת? בין אם אתם מרמים אחרים או רק את עצמכם, "השקר *שיש* להגן עליו" יתפשט באופן רקורסיבי דרך רשת הסיבתיות האמפירית, דרך רשת החוקים האמפיריים הכלליים, דרך כללי החשיבה עצמם, וההבנה שעומדת בבסיס הכללים הללו. אם ישנה תורת הכרה *טובה* בעולם, וגם שקרים או הונאות־עצמיות שאנשים מנסים להגן עליהם, אזי תבוא לעולם תורת הכרה רעה כדי להילחם בַּטובה. קשה לצַפות, בעולם הזה, למצוא את *הצד המואר* ללא *הצד האפל*. יש את השמש, ויש את מה שנרתע מפניה ומטיל צללים.

שימו לב, אלה לא בהכרח אנשים *רעים*. הרוב המכריע מאותם המהלְכים ומדקלמים את "החוכמה העמוקה" הם מרומים יותר מאשר רמאים, משלים את עצמם יותר משהם מטעים את זולתם. אני חושב.

ובוודאי שאין בכוונתי להציע לכם טיעון־נגד כללי לחלוטין, כך שבכל פעם שמישהו יציע לכם תורת הכרה שאינה לרוחכם, תאמרו: "אה, מישהו *מהצד האפל* המציא את זה." זהו אחד מכללי

* ראו פרק 90.

הצד המואר, שעליכם להפריך את הטענה לגופה, ולא על ידי האשמת הממציא שלה בכוונות זדון.

אבל *הצד האפל* נמצא שם בחוץ. פחד הוא הנתיב המוביל אליו, ובגידה אחת עלולה להמיר אתכם. לא כל מי שלובש גלימות הוא גֶ׳דַיי או מתחזה: ישנם גם שׂרי סית׳, מאסטרים ושוליות הפועלים שלא מדעת. היזהרו ועמדו על המשמר.

באשר לעריכת רשימה של תבניות מחשבה שנולדו *בצד האפל* – לא אמונות שגויות אקראיות, שימו לב, אלא תורת הכרה גרועה, ״המגננות הכלליות של הכישלון״ – ובכן, האם תרצו לנסות את כוחכם בכך, קוראים יקרים?

חלק ח'

נגד דוּחוֹשֵׁב

חלק ניכר מתהליך ההסקה שלנו הוא למעשה רציונליזציה – טוויית סיפורים שגורמים לאמונותינו הנוכחיות להיראות עקביות ומוצדקות יותר, מבלי לשפר בהכרח את מידת הדיוק שלהן. הרצף "נגד רציונליזציה" עוסק בבעיה זו, ואחריו "נגד דוּחוֹשֵׁב" [רצף זה] (על הונאה עצמית) ו"לראות בעיניים חדשות" (על האתגר שבזיהוי ראיות שאינן תואמות את הציפיות וההנחות שלנו).

– רוב בנסינגר,
רציונליות: מבוא

81

חַדחוֹשֶׁב

זכור לי הרגע המדויק בו התחלתי את המסע שלי כרציונליסט.

זה לא קרה כשקראתי את *אתה בטח מתלוצץ, מיסטר פיינמן!* או ספר אחר העוסק ברציונליות. את אלה פשוט קיבלתי כמובנים מאליהם. המסע מתחיל כאשר אתה מבחין בפגם מהותי ביכולות שלך, ומגלה איתו גם דחף להשתפר, לרכוש כישורים *חדשים*, מעבר לכישורים המועילים־אך־לא־מספיקים שמצאת בספרים.

ברגעים האחרונים של החלק הראשון בחיי, הייתי בן חמש עשרה, ושיחזרתי בעיני רוחי רגע מתוק של הצדקה־עצמית מילדותי. זכרונותיי מאותה תקופה עמומים: יש לי תמונה מנטלית, אך איני זוכר את גילי המדויק. אני חושב שהייתי בן שש או שבע, ושהאירוע המדובר התרחש במחנה קיץ.

באותו אירוע, המדריך – נער צעיר – הציע לנו את המשחק הבא: אנו החניכים נסתדר בשורה. הילד בקצה השורה יזחל בין רגלינו, ואנו ניתן לו 'פליקים' בעודו עובר. לאחר מכן יהיה תורו של החניך הבא בתור. (ייתכן שאבדה לי תמימותי הילדותית, אך איני יכול שלא לתהות...) סירבתי לשחק את המשחק הזה, ונאמר לי לשבת בפינה בזמן שהשאר משחקים.

זיכרון זה – הסירוב להפליק ולקבל הפלקות בתורי – סימל עבורי שאפילו בגיל צעיר מאוד סירבתי להפיק הנאה מהסבת כאב לאחרים. לא הייתי מוכן לקבל עסקה שבה אני זוכה להכות בישבנו של אחר

בתמורה לספיגת מכה בישבני שלי. סירבתי להזדמנות להסב כאב תמורת כאב. סירבתי לשחק משחק בעל תוחלת שלילית.

ואז, בגיל חמש עשרה, הבנתי לפתע שזה לא היה נכון. *לא* סירבתי בשל איזושהי עמדה עקרונית נגד משחקים בעלי תוחלת שלילית. אמנם כבר בגיל צעיר שמעתי על "דילמת האסיר", אך זה לא היה בגיל שבע. סירבתי פשוט כי לא רציתי שיכאב לי, והישיבה בפינה הייתה מחיר נסבל לשם כך.

חשוב מכך, התבהר לי *שתמיד* ידעתי זאת. הזיכרון האמיתי הסתתר באיזו פינה בראשי, ומדי פעם הייתי מציץ בו בעיני רוחי לשבריר שנייה, ומיד מסיט את המבט.

אמרתי לעצמי — "אז *כך* זה מרגיש, להדחיק אמת בלתי־רצויה לירכתי התודעה! כעת אבחין בכל פעם שאני עושה זאת, ואנקה את כל הפינות המודחקות בתודעה שלי!" זה היה הצעד הראשון שלי בדרך: תחושה זו, והכללה שלה מתוך החוויה הסובייקטיבית.

לשיטה זו קראתי *חַדְחוּשֵׁב*, כפראפרזה על *דוּחוּשֵׁב* מהספר *1984* של ג'ורג' אורוול. בדוחושב, אתה שוכח, ואז שוכח ששכחת. בחדחושב, אתה מבחין בכך שאתה שוכח, ואז נזכר. אתה מחזיק במוחך רק מחשבה אחת, נטולת סתירות.

"חדחושב" היתה שיטת החשיבה הרציונלית הראשונה *שיצרתי בעצמי* ולא למדתי עליה מספרים. אני בספק אם היא אכן מקורית במובן האקדמי, אך למרבה המזל אין כל דרישה כזו.

אה, ועצמי בן החמש־עשרה אהב לתת שמות לדברים.

להטיית האישוש יש עומק אימתני שנמשך עוד ועוד. לא לנָצח, שכן מורכבות המוח היא סופית, אבל מספיק כדי שזה ירגיש כמו נצח. אתה לא מפסיק לגלות מנגנונים חדשים בהם משתמש המוח כדי לדחוק דברים לפינה.

אבל אני־הצעיר הצליח לטאטא כמה וכמה מאותן פינות עם המטאטא הראשון שלו.

82

דוחוֹשֵב (הבחירה להיות מוטה)

פיסת־עיתון מוארכת הופיעה בין אצבעותיו של אובריאן. כחמש שניות נמצאה בתחום הראיה של וינסטון. זה היה תצלום, ולא היה ספק בזהותו. זה היה ה ת צ ל ו ם. היה זה טופס אחר של התצלום שבו נראו ג׳ונס, ארונסון וראתרפורד באירוע המפלגתי בניו־יורק, שנזדמן לידיו לפני אחת־עשרה שנים והוא מיהר להשמידו. לרגע היה לפני עיניו, ונעלם. אבל הוא ראה אותו, בלי צל של ספק הוא ראה אותו! הוא עשה מאמץ נואש, מכאיב, לשחרר את המחצית העליונה של גופו. מן הנמנע היה לזוז סנטימטר אחד לכאן או לכאן. כדי רגע שכח אפילו את החוגה. הוא לא רצה אלא לשוב ולאחוז את התצלום באצבעותיו, או לפחות לראותו.

״הוא קיים!״ קרא.

״לא,״ אמר אובריאן.

הוא עבר לצדו השני של החדר. חור־זכרון היה בקיר שממול. אובריאן הרים את השבכה. סמויה מן העין הייתה רצועת־הנייר הדקה חגה ומתרחקת בתוך זרם אויר חם; ולבסוף נעלמה בלהבה מלחכת. אובריאן הסב פניו מן הכותל.

״אפר,״ אמר. ״אפילו לא אפר ממש, אבק. הוא אינו קיים. מעולם לא היה קיים.״

״אבל הוא היה קיים! הוא קיים! הוא קיים בזכרון. אני זוכר אותו. אתה זוכר אותו.״

״אני איננני זוכר,״ אמר אובריאן.

לבו של וינסטון היה חלל בקרבו. הרי לך דוחושב. היתה בו הרגשה של חוסר־ישע קטלני. אילו יכול להיות בטוח שאובריאן משקר, נדמה שלא היה אכפת לו. אבל ייתכן וייתכן שאובריאן שכח את התצלום באמת. ואם כך, ודאי כבר שכח את הכחשתו שהוא זוכר אותו, ושכח את מעשה השכחה. כיצד יוכל להיות בטוח שאין כאן אלא הונאה? אולי אותו שיבוש מטורף במוח אמנם אפשר שיתרחש; זו היתה המחשבה שהכריעה אותו.

— ג׳ורג׳ אורוול, *1984* *

מה אם אשליה עצמית מסייעת לנו להיות מאושרים? ומה אם הניסיון להתגבר על הטיות בכל מחיר רק הופך אותנו — אבוי! — *פחות* מאושרים? הדבר *החכם* לעשות, בוודאי, יהיה רציונליות *מסדר שני*: לבחור מתי להיות רציונלי. כך נוכל לתת להטיות קוגניטיביות מסוימות לפעול, ונמקסם את האושר שלנו.

גם אם נתעלם מן הצד המוסרי שבכך, אני בספק אם נֶקע תודעתי סהרורי שכזה יכול באמת להתרחש.

כשאני אומר ״רציונליות־מסדר־שני״ אני מתכוון לכך שבשלב מסוים תחשוב לעצמך, ״וכעת באופן בלתי רציונלי אני הולך להאמין שאזכה בלוטו, כדי שאהיה מאושר יותר.״ אך השליטה שלנו על אמונותינו אינה כה ישירה. אינך יכול לגרום לעצמך להאמין שהשמים ירוקים בכוח רצונך בלבד. יתכן שתוכל לגרום לעצמך להאמין שאתה מאמין בכך — אם כי הרגע הקשיתי עליך לעשות זאת, בכך שהסבתי את תשומת לבך להבדל. (אין בעד מה!) אולי אפילו היית *מאמין* שאתה מאושר ונתון להונאה עצמית. אבל *בפועל*, לא היית מאושר, וגם לא באמת היית נתון להונאה עצמית.

כדי שרציונליות־מסדר־שני תהיה באמת *רציונלית*, אתה צריך ראשית מודל טוב של המציאות, כדי לאמוד את ההשלכות של רציונליות ושל אי־רציונליות. אם לאחר מכן תבחר להיות אי־רציונלי

* בתרגום ג. אריוך (עם עובד, 2003).

מסדר־ראשון, תצטרך לשכוח את התפיסה המדויקת הזו. ואז לשכוח ששכחת. איני מתכוון לבצע את הכשל הלוגי של הכללה מראיות בדיוניות, אך אני סבור שאורוול עשה עבודה טובה בהמחשה לאן מובילה הדרך הזו.

לא תוכל לדעת את ההשלכות של ההטיה עד אשר תשחרר את עצמך ממנה. ואז כבר מאוחר מדי להונאה עצמית.

האלטרנטיבה היא לבחור לעצום עיניים ולהישאר מוטֶה, בלי לדעת מה יהיו ההשלכות. זוהי לא רציונליות־מסדר־שני. זוהי טיפשות מבחירה.

חֲיֵה אופטימי בצורה אי־רציונלית לגבי כישורי הנהיגה שלך, ותוכל להיות מאושר וחסר דאגות בעוד האחרים סביבך מזיעים ורועדים מפחד. לא תצטרך לסבול מאי־הנוחות של חגורת בטיחות. תהיה חסר דאגות ליום, שבוע, שנה. ואז *בום*, תבלה את שארית חייך מייחל לכך שיכולת להרגיע את הגירוד באיבר־הרפאים. או משותק מהצוואר ומטה. או מת. זה לא יקרה בהכרח, אבל זו אפשרות. עד כמה היא סבירה? לא תוכל לדעת באופן רציונלי אלא אם כן תדע מהם כישורי הנהיגה *האמיתיים* שלך, כך שתוכל להעריך את הסיכון שאתה לוקח על עצמך. לא תוכל להעריך זאת באופן רציונלי בלי להכיר הטיות כמו הזנחת הסתברות.*

זה לא משנה כמה ימים חולפים בבערות מאושרת. מספיקה טעות אחת כדי להרוס חיים, להפסיד כל שקל שאי־פעם הרמת ממסילת הרכבת של הטיפשות.**

אחת העצות המרכזיות שאני נותן לאלו השואפים לרציונליות היא "אל תנסה להתחכם," ו־"הקשב לספקות השקטים והמכרסמים המנקרים באחורי תודעתך." אם אתה לא יודע משהו, אתה לא יודע *מה* אתה לא יודע, ולא *כמה* אתה לא יודע, ולא כמה אתה *צריך* לדעת.

---

* ראו הערה 11 במאמר *הטיות: מבוא* בכרך הראשון של *רציונליות: מ-א ועד ת*.

** הביטוי *Picking up pennies in front of a steamroller* ("לאסוף מטבעות פני מול מַכבֵּש דרכים") מתאר אסטרטגיית השקעה או מסחר שבה המשקיע מרוויח סכומים קטנים ויציבים יחסית, אבל בסיכון גבוה לאובדן גדול ופתאומי.

אין רציונליות־מסדר־שני. יש רק קפיצה עיוורת לתוך מה שעשוי להיות, או שלא להיות, לוע הר געש. מהרגע *שתדע*, כבר יהיה מאוחר מדי לעצום עיניים.

אך יש אנשים שמתעלמים מכך, מכיוון שהם לא יודעים מה הם לא יודעים. מה שאיננו יודעים שאיננו יודעים אינו זמין לנו. אנשים לא חוקרים את החלקים החסרים במפה, אלא מתייחסים אליהם כאל כאלה שמייצגים שטח ריק. כאשר הם שוקלים קפיצה עיוורת, הם מחפשים את הסכנה בזיכרונם, אך אינם מוצאים שום לוע הר געש במפה הריקה. אז למה שלא לקפוץ?

הייתי שם. ניסיתי זאת. נכוויתי. אל תנסו להתחכם.

אמרתי פעם לחברה שאני חושד שמפריזים מאוד בערכו של האושר שבבורות. היא הנידה בראשה ואמרה "לא, לא מפריזים. באמת שלא."

אולי ישנם אנשים טיפשים ומאושרים שם בחוץ. אולי הם מאושרים יותר ממך. והחיים לא הוגנים, ולא תיעשה מאושר יותר באמצעות קנאה במשהו שלא תוכל להשיג. אני חושב שמרבית הקוראים של *Overcoming Bias* לא יוכל להגיע ל"אושר שבבורות" גם אם ינסה. דרך זו חסומה בפניך. לא תוכל להגיע לרמה כזו של בורות, לא תוכל לשכוח מה שאתה יודע, לא תוכל להיות עיוור למה שכבר ראית.

האושר שבבורות חסום בפניך. לא תגיע אליו ללא נזק מוחי, ואולי אפילו לא אז. תוכל לתהות, אני מניח, האם האושר שבבורות הוא *אופטימלי* – האם זהו האושר *הגדול ביותר* אליו יכול אדם לשאוף – אך אין זה משנה. דרך זו סגורה לפניך כעת, אם אי פעם היא היתה פתוחה.

כל שנותר כעת הוא לשאוף למרב האושר שרציונליסט יכול להשיג. אני סבור שבסופו של דבר, הוא עשוי להתגלות כגדול יותר. ישנם נתיבים מוגבלים ונתיבים פתוחים אל האופק; ישנם מישורים להשתרע עליהם ללא מאמץ, והרים להעפיל אליהם; וגם אם הטיפוס תובע מאמץ רב יותר, הרי שבסופו של דבר ההר מתנשא גבוה יותר.

בנוסף, אושר אינו הכל בחיים, והחלטותיך משפיעות גם על

אושרם של אחרים.

אך אין זה משנה. כשתבין שקיימת בחירה, היא כבר לא תהיה בידיך. לא תוכל להיות עיוור למה שכבר ראית. הדרך השניה סגורה בפניך.

83

תקשיב, אני באמת מצליחה להשלות את עצמי

שוחחתי לאחרונה עם אישה ש... קשה לי לנסח במילים. על פניו היא יהודיה דתיה אורתודוקסית. היא הייתה אינטליגנטית, הכירה חלק מהראיות הארכיאולוגיות נגד הדת שלה, והייתה מודעת לטיעונים הסטנדרטיים הרדודים נגד האמונה הדתית שכל אדם דתי מכיר. כך למשל, היא ידעה שמרדכי, אסתר, המן ושתי אינם קיימים ברישומי ההיסטוריה הפרסית, אך הייתה אגדה פרסית בה מופיעים האלים הבבליים מַרדוּך ואִשתָר, ויריביהם האלים העֵילמיים הוּמַן וַשְׁתִי.* היא יודעת זאת, ועדיין חוגגת את חג הפורים. היא אחת מאותם אנשים דתיים חכמים מאוד שמתבשלים עם הסתירות הפנימיות שלהם במשך שנים, מרחיבים ומתאימים את תפיסותיהם, עד שתודעתם נראית כמו ציור של מ. ק. אֶשֶׁר.

רוב האנשים במצב הזה יעמידו פני חכמים** מכדי לשוחח עם אתאיסטים, אך היא בכל זאת הסכימה לשוחח איתי לכמה שעות.

כתוצאה מכך, אני מבין כעת דבר נוסף על הונאה־עצמית שלא

---

* אין בידינו עדות לאגדה פרסית קדומה המתועדת במקורות, אלא מדובר בהשערה מחקרית מהמאה ה-19 המבוססת על הדמיון בין שמות גיבורי המגילה לשמות האלים הנזכרים.

** ראו פרק 15.

הבנתי קודם – כלומר, שאינך צריך *באמת* להצליח להשלות את עצמך, כל עוד *תאמין* שהצלחת להשלות את עצמך. אפשר לקרוא לכך "אמונה בהונאה־עצמית".

כאשר אישה זו הייתה בתיכון, היא חשבה שהיא אתאיסטית. אך היא החליטה, באותו הזמן, שהיא תעמיד פנים כאילו היא מאמינה באלוהים. ואז – היא אמרה לי ברצינות – לאורך הזמן, היא התחילה באמת להאמין.

למיטב שיפוטי, היא טועה לחלוטין לגבי זה. במהלך שיחתנו, היא שוב ושוב אמרה משפטים כגון "אני *מאמינה* באלוהים," אך אף לא פעם אחת "אלוהים *קיים*." כששאלתי אותה מדוע היא דתיה, היא לא דיברה על משמעות קיומו של אלוהים, אלא רק על משמעות *האמונה* בקיומו של אלוהים. לא "אלוהים יעזור לי," אלא "אמונה באלוהים עוזרת לי." כאשר אתגרתי אותה ואמרתי לה – "אדם אשר תר רק אחר האמת, מתבונן ביקום שלנו ומחפש הסבר, לא יעלה בדעתו את אלוהים אפילו כהשערה," היא הסכימה איתי.

היא לא *באמת* הצליחה להשלות את עצמה להאמין שאלוהים קיים או שהדת היהודית נכונה. למיטב שיפוטי, אפילו לא קרוב לכך.

מצד שני, אני חושב שהיא *באמת* מאמינה שהיא הצליחה להשלות את עצמה.

לכן, אף שהיא אינה מפיקה שום תועלת מן האמונה באלוהים – מפני שלמעשה היא אינה מאמינה בו – היא אכן מאמינה בכנות שהצליחה להונות את עצמה להאמין בו. ולכן, היא מצפה בכנות להפיק את התועלת שהיא מייחסת להונאה עצמית כזו. *וזה*, אני מניח, אמור ליצור כמעט את אותו אפקט פלצבו כמו אמונה *ממשית* באלוהים.

זה גם עשוי להסביר מדוע היא הגנה כה ברצינות על הטענה שהיא מאמינה באלוהים אל מול שאלותיי הספקניות, אף שמעולם לא אמרה, "אה, ודרך אגב, אלוהים באמת קיים," ואף לא נראתה מתעניינת ולו במעט בטענה הזאת עצמה.

84

אמונה בהונאה עצמית

סיפרתי בפרק הקודם על שיחתי עם אישה יהודיה ועל־פניו דתיה שהגנה בעוז על הטענה שהיא מאמינה באלוהים, בעוד שנראה היה שאין היא מאמינה באלוהים כלל.

בחלק של השיחה בו שאלתי אותה על התועלות שהיא חושבת שנובעות מאמונה באלוהים, הצגתי בפניה את *"המנטרה של טרסקי"* — שהיא למעשה משפחה אינסופית של מנטרות. לדוגמה:

אם השמים כחולים
רוצה אני להאמין שהשמים כחולים

אם השמים אינם כחולים
רוצה אני להאמין שהשמים אינם כחולים.

"זוהי לא הפילוסופיה שלי," היא ענתה לי.

"לא חשבתי אחרת," עניתי לה. "אני רק שואל — אם נניח עולם שבו אלוהים *אינו* קיים, ושזו עובדה ידועה, האם את עדיין סוברת, בעולם כזה, שעליך להאמין באלוהים?" היא היססה. להפתעתי, נראה שהיא אכן הקדישה מחשבה לשאלה.

"אז אם זו שאלה על מציאות חלופית..." היא אמרה באיטיות.

בזמן השיחה חשבתי שהיא מתקשה לאפשר לעצמה לדמות עולם־בלי־אלוהים, בשל היקשרותה לעולם־עם־אלוהים.

אולם כעת, אני חושד שהקושי שלה היה לדמות את *הניגוד* בין האופן שבו נראה עולם־עם־אלוהים אל מול האופן שבו נראה עולם־בלי־אלוהים, שכן מחשבותיה היו מוקדשות *לאמונה* באלוהים, אך רשת הסיבתיות שלה לא הכילה למעשה את *אלוהים* כצומת ברשת. כך שהיא יכלה בקלות לענות על השאלה "איך העולם היה שונה אילו לא האמנתי באלוהים?" אבל לא על השאלה "איך העולם היה שונה אילו אלוהים לא היה קיים?"

בסופו של דבר, היא לא ענתה על השאלה שלי. אך היא כן הציעה דוגמה נגדית למנטרה של טרסקי:

היא אמרה, "אני מאמינה שאנשים טובים יותר מכפי שהם באמת."

ניסיתי להסביר לה שאם היא אומרת "אנשים הם רעים," משמעות הדבר שהיא מאמינה שאנשים הם רעים, ואם היא אומרת "אני מאמינה שאנשים הם טובים," משמעות הדבר שהיא מאמינה שהיא מאמינה שאנשים הם טובים. אז לומר "אנשים הם רעים אבל אני מאמינה שהם טובים" משמעותו שהיא מאמינה שאנשים הם רעים, אבל מאמינה שהיא מאמינה שהם טובים.

ציטטתי לה:

> אילו היה פועַל שמשמעותו "להאמין בטעות", לא הייתה לו צורת גוף ראשון הווה בעלת משמעות ממשית.
>
> — לודוויג ויטגנשטיין[1]

היא אמרה בחיוך: "כן, אני אכן מאמינה שאנשים טובים יותר מכפי שהם באמת. פשוט חשבתי שכדאי שאציג את זה כך, בשבילך."

> אני חושבת שסבתא רוּחשַעווה צריכה להעיף בך חתיכת מבט, וולטר," אמרה נני אוג. "אני חושבת שהמוח שלך כולו הסתבך כמו פקעת חוטים שנפלה על הרצפה.
>
> — טרי פְּרַאצֶ'ט, *Maskerade*[2]

ואני יכול להקליד את המילים "נו, אני מניח שהיא לא מאמינה שהחשיבה שלה צריכה להיות עקבית תחת רפלקציה," אבל אני עדיין מתקשה לעכל את זה.

אני מזהה את דפוס המילים שיוצאות לה מהפה, אך לא מסוגל להבין את התודעה שמאחוריהן במידה שמאפשרת הזדהות. אני מסוגל לדמיין את עצמי במקומם של חייזרים אוכלי תינוקות והגבירה קיריטסוגו השלישית*, אך איני מסוגל לדמיין איך זה להיות היא. או שאולי אני פשוט לא *רוצה*?

זו הסיבה שלאנשים אינטליגנטים יש זמן קצוב (הנמדד כזמן סובייקטיבי בו חושב האדם על נושא הדת) להיהפך לאתיאיסטים. מעבר לנקודה מסוימת, אם אתה חכם, הקדשת זמן לחשיבה והגנה על הדת, ולא הצלחת לחמוק *מהאפיסטמולוגיה של הצד האפל*,** תודעתך נראית כמו ציור של אֶשֶׁר.

(רגע נוסף שהביא אותה לעצור רגע ולחשוב – ואני מציין זאת למקרה שתהיה לכם הזדמנות להשתמש בזה – הוא שכאשר היא דיברה על כמה זה טוב להאמין שלמישהו אכפת אם אתה עושה טוב או רע – *לא*, כמובן, על האם *ישנו* אלוהים שאכפת לו אם אתה עושה טוב או רע, זה לא חלק מהדת שלה –

ואני אמרתי, "אבל לי אכפת אם את עושה טוב או רע. אז מה שאת אומרת הוא שזה לבדו לא מספיק לך, ושאת צריכה גם להאמין במשהו *שמעל* לאנושות שאכפת לו אם את עושה טוב או רע." זה גרם לה לעצור רגע ולחשוב, כי כמובן שהיא לא חשבה על זה במונחים אלה בעבר. פשוט יישום סטנדרטי של כלי לא סטנדרטי.***)

בהמשך שאלתי אותה אם הייתה משנה *משהו* בהתנהגותה לו הוכח מעבר לכל ספק שאין אלוהים, והפעם היא ענתה "לא."

---

* דמויות מתוך הנובלה של אליעזר יודקובסקי, "עולמות מתנגשים", העוסקת במפגש בין תרבויות שונות *מאוד* אחת מהשניה. תרגום לעברית זמין לקריאה באתר 'רציונליות ישראל'.

** ראו פרק 80.

*** ראו פרק 96.

"אז," עניתי באי־אמון, "לקיומו או אי־קיומו של אלוהים אין כל השפעה על האופן שבו כדאי לאנשים לחשוב או להתנהג? אני חושב שאפילו לרב זה יישמע מוזר."

כעת נראה שהדת שלה מורכבת *אך ורק* מסגידה למושג הסגידה. כפי שמאמיני העבר האמינו שדמות אב יודעת־כל תציל אותם, כך אישה זו מאמינה שעצם האמונה באלוהים תציל אותה.

לאחר שהיא אמרה "אני מאמינה שאנשים טובים יותר מכפי שהם באמת," שאלתי, "אז האם את מופתעת פעם אחר פעם כאשר אנשים לא עומדים בציפיות שלך?" הייתה שתיקה ארוכה, ואז, לאט: "ובכן... האם אני *מופתעת* כאשר אנשים... לא עומדים בציפיות שלי?"

באותה נקודת זמן לא הבנתי את ההיסוס. אני התכוונתי להציע שאם המציאות מכה בה שוב ושוב, הרי שזהו חסרון שנגרם מאמונה שגויה. אך במקום זאת, נראה שהיא הופתעה מההשלכות של זה שהיא *לא* מופתעת.

כעת אני מבין שמהות הפילוסופיה שלה היא *האמונה שהיא מצליחה להונות את עצמה*, והאפשרות שלמעשה הערכותיה לגבי אנשים אחרים הן *דווקא מדויקות* מאיימת על מתודות *האפיסטמולוגיה של הצד האפל* שהיא בנתה סביב אמונות כמו "אנשים טובים יותר מכפי שהם באמת."

היא הסירה את האליל מהבמה, והמירה אותו בסגידה מפורשת *לאפיסטמולוגיה האפלה* שהומצאה אי־אז כדי להגן על הסגידה לאליל. היא סוגדת למאמץ של עצמה לבצע הונאה־עצמית. המאמץ כושל, אך היא באמת ובתמים אינה מודעת לכך.

וכך, שומרי־הסף של השפיות האנושית מטעם עצמם (מוטו: "מחרבים את המסיבה הקטנה והמטורללת שלכם מאז ימי אפיקורס") צריכים כעת להיאבק בסגידה הפעילה להונאה־עצמית – בסגידה *לתועלותיה כביכול של האמונה*, במקום לאלוהים.

למעשה זה גורם לי להבין משהו לגבי *עצמי* שלא הבנתי עד כה – מדוע אני מתרגז כאשר אנשים מדברים על הונאה עצמית כאילו היא דבר *קל*, ומדוע אני כותב מאמרים שלמים כדי לטעון שבחירה מודעת

להאמין שהשמים ירוקים היא קשה יותר מכפי שאנשים נוטים לחשוב.

הסיבה היא, שבעוד שאתה *לא יכול* פשוט לבחור להאמין שהשמים ירוקים – אם אינך *מודע* לעובדה זו, למעשה אתה *יכול* להונות את עצמך להאמין שהצלחת להונות את עצמך.

ומכיוון שאז אתה *מצפה* לקבל את התועלת שכרוכה בהונאה העצמית, אתה מקבל את אותו אפקט פלצבו שהיית מקבל אם אכן היית מבצע את ההונאה העצמית בהצלחה.

בכך שאני מסביר עד כמה *קשה* היא הונאה עצמית, אני למעשה מכוון ישירות לאפקט הפלצבו שאנשים מפיקים מהאמונה בכך שהצליחו להונות את עצמם. אני תוקף ישירות את הדת מהזן החדש – זו שאינה סוגדת לאלוהים, אלא רק לסגידה עצמה.

אני תוהה אם קרב זה יצמיח טיעונים חדשים: מדוע לא האמונה, אלא האמונה באמונה, היא דבר רצוי *כשלעצמו*? מדוע אנשים מפיקים תועלת כה רבה מהסגידה לסגידה? והאם נצטרך לחזור על כל זה שוב, לטעון נגד אמונה באמונה באמונה וסגידה לסגידה לסגידה? או שמא המאמינים האינטליגנטים יזנחו לבסוף קו טיעונים זה?

הלוואי ויכולתי להשתכנע שאף אחד לא יטען בעד אמונה באמונה באמונה, אך טיעון הזומבי* הפילוסופי הגיע אף ליותר מכך, ותומכיו עדיין מסרבים לנטוש אותו.

---

1. Ludwig Wittgenstein, *Philosophical Investigations*, trans. Gertrude E. M. Anscombe (Oxford: Blackwell, 1953).

2. Terry Pratchett, *Maskerade*, Discworld Series (ISIS, 1997).

---

* *טיעון הזומבי* הינו טיעון פילוסופי בעד הגישה הדואליסטית בשאלה הפסיכופיזית – דהיינו, בעד הימצאות הנשמה כישות נפרדת מהגוף. ראו פרק 221.

85

פרדוקס מור

פרדוקס מור הינו המושג המקובל עבור משפטים מסוג "יורד גשם בחוץ אך איני מאמין בכך."

בעקבות קריאת חלק מהתגובות באתר *LessWrong*, אני חושב שההבנה שלי לגבי פרדוקס מור השתפרה. Jimrandomh מציע:

> אנשים רבים מתקשים להפריד בין רמות תיווך שונות. עבורם "אני מאמין ב-**א**" ו"**א**" הינם זהים, ולכן הסיבות שבגללם האמונה ב-**א** מועילה הן גם הסיבות שבגללם **א** נכון.

אינני חושב שהאבחנה הזו נכונה – הרי ילדים מגיל צעיר מבינים את הרעיון של אמונה שגויה, ולשם כך דרושות קטגוריות נפרדות עבור המפה והשטח. אבל האבחנה כן מצביעה על רעיון דומה:

אנשים רבים אינם מפרידים בין *אמונה* במשהו לבין *תמיכה* בו.

אחרי הכל – "אני מאמין בדמוקרטיה" משמעו, בשפה המקובלת, שאתה תומך ברעיון הדמוקרטי, לא שאתה מאמין שדמוקרטיה קיימת. למילה "מאמין", אם כן, ישנה יותר ממשמעות אחת. ייתכן ואנו עומדים בפני מילה מבולבלת* הגורמת לחשיבה מבולבלת (או שאולי היא פשוט משקפת בלבול קיים).

---

* ראו פרק 169. בלשון הרמב"ם בתרגום אִבן־תִיבּוֹן: *שם משותף*, בבלשנות מודרנית: *פּוֹלִיסֶמיה* (מילה יחידה בעלת מספר משמעויות קשורות) ובלוגיקה: *אֶקוויבוֹקַציה* (כפל־לשון מטעה).

נתבונן, אם כן, בדוגמה המקורית מהפרק הקודם: "אני מאמינה שאנשים טובים יותר מכפי שהם באמת." האישה איתה שוחחתי הציעה סיבות מדוע מוטב להאמין שאנשים הם טובים – סיבות בריאותיות וכדומה – ומכיוון ש"האמונה בטוב ליבם של אנשים" מביאה אותה לחוות תחושת חמימות נעימה, היא מתבוננת בחוויה זו ומסיקה "אני מאמינה שאנשים טובים." כלומר, היא מפרשת בטעות את *האפקט החיובי* שיוצרת הטענה, כסימן לכך *שהיא מאמינה בטענה עצמה*. יחד עם זאת, בעולם עצמו נראה שהאנשים *אינם* כה טובים. לכן היא אומרת "אני מאמינה שאנשים טובים יותר מכפי שהם באמת."

וזה כבר גובל בטעות בתום לב – במידת מה – שהרי אנשים אינם מקבלים הדרכה מפורשת כיצד לזהות מתי הם מאמינים במשהו. כמו במשל הדרקון במחסן:* מי שטוען "יש דרקון במחסן שלי, אבל הוא בלתי נראה," לא מזהה *שהתחזית* שלו שלא יראה את הדרקון מהווה אינדיקציה לכך שהוא מחזיק במודל (מדויק) המכיל אפס דרקונים.

הרי אנשים אינם עוברים *הכשרה* בזיהוי האמונות שלהם. זה לא שאי־פעם מלמדים אותם בתיכון: "ההרגשה של אמונה אמיתית – של הימצאותה של טענה במאגר האמונות הפנימי שלכם – היא שפשוט נדמה לכם שכך העולם *הינו*. תוכלו לזהות את ההרגשה הזאת, שהיא אמונה ממשית, ללא מרכאות, ולהבדיל בינה לבין הרגשה טובה כלפי משהו שנדמית בעיניכם כ"אמונה" – כלומר, אמונה במרכאות."

הבנה זו מסייעת מאוד להפוך את המקרה המציאותי הזה של פרדוקס מור לפחות מוזר, ומציעה מנגנון נוסף באמצעותו אנשים יכולים להחזיק באמונות נכונות ושגויות בו זמנית.

בדומה לכך, המשתמש Kurige כתב:

> אני מאמין שיש אלוהים – ושהוא נטע בנו חוש מוסרי להבחין בין טוב ורע, שבאמצעותו אנו יכולים לשפוט את העולם סביבנו. בנוסף אני מאמין שחוש מוסרי הוטמע בנו בתהליך אבולוציוני – חוש

* ראו פרק 13.

מוסרי שמקורו, ככל הנראה, בהיווצרותן של קואליציות מטא־פוליטיות בקהילות בּוֹנוֹבּוֹ* מלפני זמן רב מאוד. שתי האמונות האלה אינן סותרות זו את זו, אך המורכבות טמונה ביישוב ביניהן.

אני חושד, Kurige, שהחלטת שיש לך *סיבות לתמוך* ב"אמונה" שאלוהים נטע בנו חוש מוסרי להבחין בין טוב ורע. ושיש לך סיבות לאמץ גם את פסיקת המדע. שתי הקהילות נראות כמו קהילות שכדאי להשתייך אליהן, הלא כן? יש יתרונות לשתי מערכות האמונות? אתה מתבונן פנימה ומגלה שיש לך תחושות חיוביות לגבי שתי האמונות?

אך לא אמרת:

"אלוהים נטע בנו חוש מוסרי להבחין בין טוב ורע, ובנוסף הוטמע בנו חוש מוסרי בתהליך אבולוציוני. שני ההסברים הללו של המציאות אינם סותרים זה את זה, אך המורכבות טמונה ביישוב ביניהם."

אם אתה קורא את זה, Kurige, אמור את המשפט לעיל בקול ובמהירות, כך שתוכל לשים לב שהוא לפחות קצת יותר קשה לבליעה – שים לב *להבדל הסובייקטיבי* – בטרם תטרח לבצע רציונליזציה מחדש.

זהו ההבדל הסובייקטיבי בין קיומן של סיבות לתמוך בשתי אמונות שונות, לבין מודל מנטלי של עולם יחיד, דרך אחת שבה הדברים הם כפי שהם.

* *בּוֹנוֹבּוֹ* (*pan paniscus*) הם מין של קופי־אדם שחולקים עם האדם אב קדמון משותף, ומחקר ההתנהגות שלהם משמש בנסיון לשחזור תכונותיו של אותו אב קדמון.

86

האמן שאינך מסוגל להונות את עצמך

אני לא רוצה להיתפס כמי שניטפל ל-Kurige, אך אני חושב שעליך לצַפות לכמות מסוימת של הרמת גבות אם אתה בא למקום כמו *LessWrong* ואומר:

> תובנה אחת שמסייעת לי להסביר את הפער שאני חש כשאני משוחח עם נוצרים אחרים היא העובדה שבשלב מסויים תפיסת העולם שלי פנתה הרחק מאמונה עיוורת, והתמקמה אי־שם בסביבת הדוּחוֹשב האוروליאני.

"אם אתה *יודע* שזה דוחושב...

...איך אתה עדיין מאמין בזה?" אני רוצה לומר בחוסר אונים.

או:

> בחרתי להאמין בקיומו של אלוהים – במודע ובמכוון. עם זאת, להחלטה זו אין כל קשר לעצם קיומו של אלוהים.

אם אתה *יודע* שהאמונה שלך אינה תואמת למציאות, איך אתה עדיין מאמין בה?

האם ההבנה ברמת *תחושת הבטן* "אה, רגע, מסתבר שהשמים *לא* ירוקים" לא אמורה לנבוע ישירות מההבנה "למפה שלי שאומרת

'השמים ירוקים' אין שום קשר לשטח"?

ובכן... מסתבר שלא.

חלק מפתרון החידה עשוי להיות ההסבר שלי לגבי פרדוקס מור ("יורד גשם בחוץ, אך איני מאמין בכך") מהפרק הקודם – אנשים מבלבלים בין תחושה חיובית לגבי טענה לבין נכונות הטענה.

אך יתכן שמרכיב נוסף בפתרון – בניגוד לתרעומת שרציתי להביע קודם – הוא שלמעשה די *קל* שלא לקפוץ מ-"מפה שמשקפת את השטח הייתה אומרת **א**" לאמונה ב-**א**. צריך להשקיע מאמץ כדי *להסביר* את הרעיון של תודעה כמנגנון הבונה התאמה בין מפה לשטח, וגם אז, עשויה להיות עבודה נוספת כדי להפנים את ההשלכות ברמת *תחושת הבטן*.

אני מבין כעת שכאשר כתבתי "אתה לא יכול לגרום לעצמך להאמין שהשמים ירוקים בכוח רצונך בלבד," לא מסרתי דיווח ענייני של העובדות בלבד. ניסיתי גם להטמיע נבואה שמגשימה את עצמה.*

אולי יהיה זה נבון להסתובב בעולם ולשנן שוב ושוב "איני יכול לבצע דוחשיבה בהצלחה! עמוק בפנים, אדע שזה לא נכון! אם אני יודע שאין קשר בין המפה שלי לשטח, הרי שאיני מאמין בה!"

כיוון שכך – אם אי פעם תתפתה לנסות – המחשבות "אבל אני יודע שזה לא באמת נכון!" ו-"איני יכול לשטות בעצמי!" יהיו תמיד זמינות במוחך. וכך, אכן יהיה זה פחות סביר שתצליח לשטות בעצמך. סביר יותר שתבין, ברמת תחושת הבטן, שלספר לעצמך **א** לא הופך את **א** לאמיתי: ולכן, באמת ובתמים, לא-**א**.

אם תספר לעצמך תדיר *שאינך יכול* פשוט לבחור להאמין שהשמים ירוקים – יהיה זה פחות סביר שאכן תצליח להונות את עצמך באחד המובנים: או בכך שבאמת תאמין, או שתיפול בפרדוקס מור, או

---

* "שיקרת לי."

הארי פקח את עיניו והביט בדראקו. "הייתי מעדיף לומר," אמר הארי, בקול לא ממש יציב, "שהדברים שאמרתי לך היו נכונים מנקודת מבט מסוימת... מה שאמרתי לך היה נבואה שמגשימה את עצמה; האמנת שאתה לא יכול לשטות בעצמך, ולכן לא ניסית."

אמונה באמונה, או אמונה בהונאה־עצמית.

אם תמשיך לספר לעצמך שעמוק בפנים אתה תדע –

אם תמשיך לספר לעצמך שאתה פשוט תסתכל במפה השגויה הבנויה בקפידה, ופשוט תדע שהיא מפה שגויה שאין סיבה לצַפות שתשקף את השטח, ולכן, למרות המורכבות שלה, לא תוכל לתת בה כל אמון –

אם תמשיך לספר לעצמך שעקביות רפלקטיבית* תיכנס לפעולה ותגרום לך להפסיק להאמין ברמת האובייקט, ברגע שתבין ברמת המטא** שהמפה אינה משקפת –

אזי כאשר יגיע רגע האמת, ייתכן שאכן תיכשל.

כשזה מגיע להונאה עצמית מכוונת, עליך *להאמין בחוסר היכולת שלך!*

אמור לעצמך שהמאמץ נידון לכישלון – *והוא אכן יהיה כזה!*

האם זהו כוחה של חשיבה חיובית או של חשיבה שלילית? כך או כך, זה נראה כמו אמצעי זהירות נבון.

---

* יודקובסקי משתמש במונח *עקביות רפלקטיבית* (reflective consistency) בדיונים שלו על אודות הבטיחות של בינה מלאכותית, כדי לתאר מערכת קבלת החלטות שמסוגלת לשנות את עצמה (או לבנות מערכת חדשה שתחליף אותה), ומסכימה לעשות זאת רק אם המערכת החדשה תקבל החלטות באותו אופן כמוה – כלומר, לפי אותם כללים ולאותן מטרות. בפרק זה הוא משתמש במונח באופן מטאפורי: ברגע שאתה מבין ברמת המטא שהמפה שלך אינה משקפת את המציאות, העקביות הזו אמורה לגרום לך להפסיק להאמין גם ברמת האובייקט.

** רמת האוביקט היא האמונה עצמה – למשל, "השמים ירוקים". רמת המטא היא החשיבה על האמונה – למשל, "המפה שלי שאומרת שהשמיים ירוקים אינה משקפת את השטח."

חלק ט'

לראות בעיניים חדשות

חלק ניכר מתהליך ההסקה שלנו הוא למעשה רציונליזציה – טוויית סיפורים שגורמים לאמונותינו הנוכחיות להיראות עקביות ומוצדקות יותר, מבלי לשפר בהכרח את מידת הדיוק שלהן. הרצף "נגד רציונליזציה" עוסק בבעיה זו, ואחריו "נגד דוּחוֹשֶׁב" (על הונאה עצמית) ו"לראות בעיניים חדשות" [רצף זה] (על האתגר שבזיהוי ראיות שאינן תואמות את הציפיות וההנחות שלנו).

– רוב בנסינגר,
רציונליות: מבוא

87

עוגנים ותיקונים

נניח שאני מסובב גלגל מזל מול עיניכם, והוא נוחת על המספר 65. ואז אני שואל: האם אתם מעריכים שאחוז המדינות האפריקאיות מתוך מדינות האו״ם הוא גבוה או נמוך מהמספר הזה? מהו לדעתכם אחוז המדינות האפריקאיות באו״ם? קחו רגע לחשוב על שתי השאלות האלה בעצמכם, אם תרצו. אל תשתמשו בגוגל בבקשה.

בנוסף, נסו לנחש, בתוך *חמש שניות*, את ערכו של הביטוי האריתמטי הבא. חמש שניות. מוכנים? היכון, הכן... *צא*!

$$1\times2\times3\times4\times5\times6\times7\times8$$

טְבֶרְסְקי וכָּהֲנמן תיעדו את הערכותיהם של נבדקים שראו מספרים שונים בגלגל המזל.[1] ההערכה החציונית של נבדקים שראו את הגלגל נוחת על 65 הייתה 45%. ההערכה החציונית של אלה שראו 10 הייתה 25%.

התיאוריה הנוכחית עבור תצפית זו ודומות לה היא שהנבדקים מתייחסים למספר הראשון, חסר המשמעות, בתור נקודת התחלה או *עוגן*. או אז, הם מתקנים את הערכתם מנקודה זו כלפי מעלה או מטה, והם ממשיכים לתקן עד שהם מגיעים למספר ש״נשמע סביר״. תוצאה טיפוסית של תהליך זה היא תיקון שמסתיים קרוב מדי לעוגן המקורי — גם מספרים רחוקים יותר עשויים ״להישמע סבירים״, אבל הנבדק

עוצר במספר הראשון שנשמע סביר. באופן דומה, סטודנטים שהוצג להם הביטוי 1×2×3×4×5×6×7×8 סיפקו אומדן חציוני של 512, בעוד סטודנטים שראו את הביטוי 8×7×6×5×4×3×2×1 סיפקו אומדן חציוני של 2,250. ההיפותזה שהובילה לתכנון הניסוי הייתה שסטודנטים ינסו לכפול (או לנחש) את ערכם של האיברים הראשונים כעוגן, ואז יבצעו תיקון כלפי מעלה. בשני המקרים התיקונים היו קטנים מדי ביחס לתוצאה האמיתית, 40,320. אך אוסף הניחושים הראשון היה רחוק יותר, מכיוון שהם התחילו עם עוגן נמוך יותר.

טברסקי וכהנמן מדווחים שהצעת תשלום תמורת דיוק רב יותר לא הפחיתה את אפקט העוגן.

שְׁטָרַאק ומוּסַוויילֶר שאלו נבדקים באיזו שנה ביקר איינשטיין בארצות הברית לראשונה.[2] עוגנים בלתי סבירים בעליל, כגון 1215 או 1992, גרמו לאפקט עיגון גדול באותה מידה כמו עוגנים סבירים יותר כמו 1905 או 1939.

ישנם יישומים ברורים לכך, למשל בדיוני שכר או ברכישת רכב. אינני מציע שתנצלו זאת, אבל היזהרו שאחרים לא ינצלו זאת נגדכם.

התבוננו בחשיבה שלכם, ושימו לב כאשר אתם *מתקנים* מספר כשאתם מנסים להעריך דבר מה.

ניסיונות לנטרל את הטיית העיגון לא הוכחו כיעילים. אני מציע שתי הצעות: ראשית, אם הניחוש הראשוני לא נשמע סביר, התעלמו ממנו לגמרי ונסו למצוא הערכה חדשה, במקום לתקן מהעוגן. אך זה לבדו לא מספיק – נבדקים שהונחו להתעלם מהעוגן עדיין נתנו הערכות שהושפעו ממנו.[3] אז שנית, אפילו אם אתם מנסים את ההצעה הראשונה, נסו לחשוב גם על עוגן בכיוון ההפוך – עוגן שהוא בעליל קטן מדי או גדול מדי, במקום גדול מדי או קטן מדי – והקדישו לו רגע של מחשבה.

1. Amos Tversky and Daniel Kahneman, "Judgment Under Uncertainty: Heuristics and Biases," *Science* 185, no. 4157 (1974): 1124–1131, doi:10.1126/science.185.4157.1124.

2. Fritz Strack and Thomas Mussweiler, "Explaining the Enigmatic Anchoring Effect: Mechanisms of Selective Accessibility," *Journal of Personality and Social Psychology* 73, no. 3 (1997): 437–446.

3. George A. Quattrone et al., "Explorations in Anchoring: The Effects of Prior Range, Anchor Extremity, and Suggestive Hints" (Unpublished manuscript, Stanford University, 1981).

88

הטרמה וזיהום

נניח שאתם מבקשים מנבדקים ללחוץ על כפתור אחד אם רצף האותיות שאתם מראים להם יוצר מילה, ועל כפתור אחר אם הוא לא יוצר מילה (למשל "שלס" לעומת "שלט"). ואז אתם מראים להם את המילה "מים". לאחר מכן, יקח להם פחות זמן לזהות את רצף האותיות "שתיה" בתור מילה. תופעה זו ידועה כ*הטרמה קוגנטיבית*. סוג זה של הטרמה יהיה "הטרמה סמנטית" או "הטרמה קונספטואלית".

הדבר המרתק לגבי הטרמה הוא שהיא מתרחשת ברובד כה בסיסי – הטרמה מאיצה את *זיהוי האותיות כמרכיבות מילה*, מה שהייתם מצפים שיקרה *לפני* שחשבתם על משמעות המילה.

הטרמה חושפת גם את האופי המקבילי של התפשטות הגירויים במוח: אם המילה "מים" מפעילה את המילה "שתיה", היא כנראה מפעילה גם את המילים "נהר", "כוס", או "התָזה"... וההפעלה הזו מתפשטת, דרך הקישורים הסמנטיים של המושגים, כל הדרך עד לזיהוי רצף האותיות.

הטרמה היא תהליך תת־מודע ובלתי ניתן לעצירה, תוצר לוואי של הארכיטקטורה העצבית במוח האנושי. לנסות לעצור הטרמה זה כמו לנסות לעצור התפשטות גירויים במעגלים העצביים שלך. נסו להגיד בקול את הצבע – לא את המשמעות, אלא את הצבע – של רצף האותיות הבא:

ירוק

בניסוי של מוּסַוויילֶר וּשְׁטְרַאק, הנבדקים נשאלו שאלת עיגון: "האם הטמפרטורה השנתית הממוצעת בגרמניה נמוכה או גבוהה מ-5 מעלות / 20 מעלות צלזיוס?"[1] לאחר מכן, במשימת זיהוי מילים, הנבדקים להם הוצג העוגן של 5 מעלות זיהו במהירות רבה יותר מילים כגון "קור" ו-"שלג", בעוד נבדקים להם הוצג העוגן הגבוה יותר זיהו מהר יותר "חום" ו-"שמש". זה מראה מנגנון עיגון שאינו קשור למנגנון התיקון שהוצג בפרק הקודם: הטרמה אוטונומית של מחשבות וזכרונות התואמים לעוגן.

התוצאה הכללית יותר היא ש"מידע" *חסר ערך, שגוי* או *לא רלוונטי* יכול להשפיע על הערכות והחלטות. בתחום ההיוריסטיקות וההטיות, תופעה כללית זו מכונה "זיהום".[2]

מחקר ראשוני של היוריסטיקות והטיות גילה את אפקטי העיגון, כגון נבדקים שמעריכים את אחוז המדינות האפריקאיות המיוצגות באו"ם כנמוך (גבוה) יותר, כתלות בשאלה שנשאלו תחילה האם האחוז גבוה או נמוך מ-10 (65). אפקט זה יוחס במקור לכך שהנבדקים משתמשים בעוגן כנקודת התחלה, מתקנים משם, ונעצרים כשהם מגיעים לערך הראשון שנראה להם סביר, ואז הם בתת־תיקון כיוון שהם נעצרים באחד הקצוות של רווח הסֶמֶך.*[3]

נראה שבחלק מהנסיבות, השערה זו של טְבֶרְסְקי וְכָּהַנמן עדיין נחשבת לנכונה, במיוחד כאשר הנבדקים מייצרים את ההערכה הראשונית בעצמם.[4] אך המחקר המודרני מראה שרוב הטיות העיגון נגרמות למעשה כתוצאה מזיהום ולא ממנגנון תיקון. (תודה למשתמש Unnamed שהזכיר לי זאת – קראתי זאת במאמר של אֶפְּלי וגילוביץ׳ לפני שנים, בספר *היוריסטיקות והטיות*, אך שכחתי.)

בחנות המכולת הקרובה לביתכם בטח יש שלט מעצבן בסגנון

---

* *רווח סֶמֶך* (confidence interval) הוא מושג מתורת האמידה הסטטיסטית המתאר טווח ערכים שסביר להניח שהערך הרצוי האמיתי נמצא בתוכו, ברמת ביטחון מסוימת (למשל 95%).

"עד 12 ללקוח" או "5 ב-10". האם שלטים אלה מצליחים לגרום ללקוחות לקנות יותר? אתם בוודאי חושבים שאינכם מושפעים מהם. אך *מישהו* מושפע, כיוון שהוכח ששלטים אלה עובדים, וזו הסיבה שחנויות ממשיכות להשתמש בהם.[5]

אך ההיבט המפחיד ביותר של הזיהום הוא שהוא משמש בתור פן נוסף מבין אלף הפנים של הטיית האישוש. כאשר רעיון נכנס לראשכם, הוא מטרים מידע תואם – ובכך מבטיח את המשך קיומו. תשכחו מלחצים אבולוציוניים של ניצחון בויכוחים פוליטיים: הטיית האישוש מוטמעת ישירות אל תוך החומרה שלנו, רשתות אסוציאטיביות המטרימות מחשבות וזכרונות תואמים. תופעת לוואי מצערת של קיומנו כיצורים נוירוניים.

תמונה רגעית בודדת עשויה להספיק כדי להטרים מילים קשורות עבור זיהוי מהיר. אל תחשבו שדרוש יותר מכך כדי לגרום להטיית האישוש. כל שדרוש הוא הבלח אחד, והמסקנה כבר רשומה, שכן אנו משנים את דעתנו פחות מכפי שנדמה לנו...*

---

1. Thomas Mussweiler and Fritz Strack, "Comparing Is Believing: A Selective Accessibility Model of Judgmental Anchoring," *European Review of Social Psychology* 10 (1 1999): 135–167, doi:10.1080/14792779943000044.

2. Gretchen B. Chapman and Eric J. Johnson, "Incorporating the Irrelevant: Anchors in Judgments of Belief and Value," in Gilovich, Griffin, and Kahneman, *Heuristics and Biases*, 120–138.

3. Tversky and Kahneman, "Judgment Under Uncertainty."

4. Nicholas Epley and Thomas Gilovich, "Putting Adjustment Back in the Anchoring and Adjustment Heuristic: Differential Processing of Self-Generated and Experimentor-Provided Anchors," *Psychological Science* 12 (5 2001): 391–396, doi:10.1111/1467-9280.00372.

* ראו פרק 97.

5. Brian Wansink, Robert J. Kent, and Stephen J. Hoch, "An Anchoring and Adjustment Model of Purchase Quantity Decisions," *Journal of Marketing Research* 35, no. 1 (1998): 71–81, http://www.jstor.org/stable/3151931.

89

האם אנו מאמינים לכל מה שאומרים לנו?

כמה ניסויים מוקדמים על עוגנים ותיקונים בדקו האם *הסחת דעתם* של הנבדקים – למשל, יצירת עומס קוגניטיבי על ידי בקשה לחפש את הספרה 5 בתוך רצפים של מספרים – תפחית את מידת התיקון, וכך תגדיל את השפעת העוגנים. רוב הניסויים נראו כמאששים את הרעיון שעומס קוגניטיבי מגביר את השפעת העיגון, ובאופן כללי יותר, את ה"זיהום".

כאשר דניאל גילברט סקר את צבר תוצאות הניסויים – עוד ועוד ממצאים של זיהום המוחמר עקב עומס קוגניטיבי – התהווה לעיניו דפוס מטורף לחלוטין: האם יתכן שאנו מאמינים *לכל* מה שאומרים לנו?

טבעי לחשוב שכאשר מציגים בפנינו טענה, ראשית *נבין* את משמעותה, לאחר מכן *נשקול* אותה, ולבסוף *נקבל* או *נדחה* אותה. מודל זה של תהליך העיבוד הקוגניטיבי, אשר נדמה כמובן מאליו, מקורו בדקארט. אך יריבו של דקארט, שפינוזה, חלק עליו: הוא הציע *שקודם כל אנו מקבלים טענה באופן סביל תוך כדי תהליך ההבנה שלה*, ורק לאחר מכן אנו *חדלים להאמין באופן פעיל* בטענות שנדחו עקב מחשבה מודעת.

במאות השנים האחרונות הפילוסופים נטו לכיוונו של דקארט,

שכן השקפתו נראתה יותר, ובכן, הגיונית. אך גילברט זיהה דרך לבחון את השערותיהם של דקארט ושל שפינוזה באופן ניסויי.

אם דקארט צודק, הרי שהסחת דעתם של נבדקים אמורה לשבש הן את קבלתן של טענות אמת והן את דחייתן של טענות שקר. אם שפינוזה צודק, הרי שהסחת דעתם של הנבדקים אמורה לגרום להם לזכור טענות שקר כאילו היו אמת, אך לא אמורה לגרום להם לזכור טענות אמת כאילו היו שקר.

גילברט, קְרוּל ומָאלון אוששו את התוצאה הזו. הם הראו כי בקרב נבדקים שהוצגו להם טענות חדשות שסומנו כאמת או כשקר, להסחת הדעת לא הייתה השפעה על זיהוי טענות אמת (55% הצלחה ללא הפרעה, לעומת 58% עם הפרעה), אך היא כן השפיעה על זיהוי טענות שקר (55% הצלחה ללא הפרעה, לעומת 35% עם הפרעה).[1]

המחשה דרמטית הרבה יותר התקבלה מניסויי המשך שערכו גילברט, טָפַרודי ומאלון.[2] הנבדקים קראו בקול מתוך מסך דיווחים על אודות פשעים, כאשר צבע הטקסט ציין אם דיווח מסוים הוא אמת או שקר. דיווחים מסוימים הכילו טענות שקר שהעלו את חומרת הפשע, בעוד דיווחים אחרים הכילו טענות שקר שהקלו בחומרתו. חלק מהנבדקים נדרשו להקדיש תשומת לב גם לרצפי ספרות, בחיפוש אחר הספרה 5, בזמן שקראו את דיווחי הפשע — זו הייתה משימת הסחת הדעת שנועדה ליצור עומס קוגניטיבי. לבסוף, הנבדקים התבקשו להמליץ על משך תקופת המאסר לכל פושע, מ-0 ועד 20 שנה.

נבדקים בקבוצת ה"עמוסים קוגניטיבית" המליצו על 11.15 שנות מאסר בממוצע לפושעים בתנאי ה"החמרה", כלומר, פושעים שדיווחיהם הכילו טענות אשר העלו את חומרת הפשע וסומנו כשקר. אותם נבדקים המליצו על 5.83 שנות מאסר בממוצע לפושעים שדיווחיהם הכילו טענות אשר הקלו בחומרת הפשע וסומנו כשקר. הפרש זה — כמעט פי־שניים — היה, כפי שוודאי ניחשתם, מובהק סטטיסטית.

המשתתפים שלא היו תחת עומס קוגניטיבי קראו את אותם

הדיווחים בדיוק, עם אותם הסימונים, ואותם רצפי מספרים שנעו מדי פעם על המסך, אלא שהם לא נדרשו לחפש אחר הספרה 5. לכן, הם יכלו להקדיש יותר קשב ל"שלילת האמונה" בטענות שסומנו כשקריות. נבדקים בקבוצה זו המליצו על 7.03 שנים לעומת 6.03 שנים עבור פושעים שדיווחי הפשע שלהם הוחמרו או הוקלו בשל הטענות השקריות, בהתאמה.

המאמר של גילברט, טפרודי ומאלון נשא את הכותרת "אינכם יכולים שלא להאמין לכל מה שאתם קוראים".

הדבר מרמז – בלשון המעטה – שעלינו להיות זהירים יותר כאשר אנו נחשפים למידע לא אמין, בייחוד אם אנו עסוקים בדבר מה נוסף באותו הזמן. היזהרו כשאתם מציצים בטלפון הנייד שלכם תוך כדי קניות בסופרמרקט.

נ"ב: לפי שמועה בלתי מבוססת שכרגע המצאתי, אנשים יהיו פחות ספקנים כלפי הנאמר בפרק זה בגלל שינויי הצבעים המסיחים את הדעת.

---

1. Daniel T. Gilbert, Douglas S. Krull, and Patrick S. Malone, "Unbelieving the Unbelievable: Some Problems in the Rejection of False Information," *Journal of Personality and Social Psychology* 59 (4 1990): 601–613, doi:10.1037/0022-3514.59.4.601.

2. Gilbert, Tafarodi, and Malone, "You Can't Not Believe Everything You Read."

90

מחשבות מטמון

אחת התעלומות הגדולות ביותר בנוגע למוח האנושי היא איך הדבר הארור הזה *פועל בכלל*, כשרוב הנוירונים מופעלים 20-10 פעמים בשנייה, או 200 הֶרץ* לכל היותר. בנוירולוגיה, "כלל מאה הצעדים" קובע שכל מודל תיאורטי לפעולה של המוח חייב לכלול 100 צעדים עוקבים *לכל היותר*. אפשר לבצע פעולות במקביל ככל שתרצו, אך אינכם יכולים להניח יותר מ-100 פולסים עצביים המתרחשים בזה אחר זה – ורצוי פחות.

תארו לעצמכם שהייתם נאלצים לתכנת תוך שימוש במעבדי 100 הרץ בלבד, רבים ככל שתרצו. הייתם זקוקים למאה מיליארד מעבדים רק כדי להצליח לבנות תוכנה שמבצעת *משהו* בזמן אמת.**

אם אכן הייתם צריכים לכתוב תוכנות זמן־אמת תוך שימוש במאה מיליארד מעבדי 100 הרץ, טכניקה אחת שהייתם מנצלים ככל יכולתכם היא שימוש בזיכרון־מטמון. השיטה היא שמירה של תוצאות החישובים הקודמים ושליפה שלהן בהמשך, במקום חישוב בכל פעם מחדש. זהו דפוס פעולה *נוירונ*י טיפוסי – זיהוי, אסוציאציה, השלמת תבניות.

ניתן בהחלט לשער *שרוב* הקוגניציה האנושית מורכבת בפועל

---

* *הֶרץ* (Hz) היא יחידת מידה לתדירות, המציינת מספר אירועים בשנייה.

** לשם השוואה, מעבדים מודרניים במחשבים אישיים בעת הוצאת ספר זה מסוגלים להגיע למהירות של 4 ואף 5 GHz – דהיינו, 5,000,000,000 פעולות בשנייה.

משליפות מתוך המטמון.

המחשבה הזו נוטה לחלוף בראשי מדי פעם.

היה לי סיפור שהמחיש זאת היטב, שחשבתי ששמרתי בסימניות שלי, אך אינני מצליח למצוא אותו: זה היה סיפור על אדם ששכנו המומחה־לכל־דבר טען פעם, כבדרך אגב, שהדרך הטובה ביותר להסיר ארובה מהבית היא לשבור את הקמין, לחכות שהלבנים מהארובה ייפלו מפלס אחד מטה, לשבור את הלבנים האלה, ולחזור על הפעולה עד שהארובה תוסר כליל. שנים לאחר מכן, כשאותו אדם רצה להסיר את תנור הארובה שלו עצמו, מחשבת המטמון הזו ארבה שם, מחכה לזנק...

כפי שאותו אדם ציין לאחר מכן – ואתם יכולים לנחש שזה לא נגמר בטוב – השכן לא היה בקי במיוחד בעניינים אלה, ולא ממש היווה מקור מידע מהימן. אילו היה אותו אדם *מטיל ספק* ברעיון, סביר להניח שהיה מבין שזה רעיון גרוע. ישנן תוצאות של חישובים בזיכרון המטמון שמוטב היה לחשב מחדש. אבל המוח משלים את התבנית באופן אוטומטי – ואם אינכם מבינים במודע שהתבנית דורשת תיקון, תישארו עם התבנית הקיימת.

אני חושד שאילו המחשבה הייתה עולה בדעתו של האיש עצמו – אילו *הוא עצמו* היה הוגה את הרעיון המבריק הזה להסרת ארובה – הוא היה בוחן את הרעיון בקפידה רבה יותר. אבל אם מישהו *אחר* כבר חשב על הרעיון עד תומו, אפשר לחסוך בכוח מחשוב על ידי שמירת *המסקנה* שלו במטמון – הלא כן?

במיוחד בתרבות המודרנית, איש אינו יכול לחשוב מספיק מהר כדי לחשוב את המחשבות של עצמו. אילו הייתי ננטש ביער בינקותי, ומגודל בידי זאבים או רובוטים אילמים, בקושי היה ניתן לזהות אותי כאנושי. אף אחד לא יכול לחשוב מהר מספיק כדי לשחזר את חוכמתו של שבט ציידים־לקטים במהלך תקופת חיים אחת, אם עליו להתחיל מאפס. באשר לחוכמתה של תרבות אוריינית – תשכחו מזה.

אבל הצד השני של המטבע הוא שאני רואה כל הזמן אנשים ששואפים לחשיבה ביקורתית, שמשתמשים במחשבות מטמון שלא

הומצאו על ידי הוגים ביקורתיים.

דוגמה טובה לכך היא אותו ספקן שמודה, "ובכן, אי אפשר להוכיח או להפריך דת באמצעות ראיות עובדתיות." כפי שציינתי במקום אחר, זה פשוט שקרי מבחינת תורת ההסתברות.* וזה גם פשוט שקרי ביחס לפסיכולוגיה האמיתית של הדת – לפני כמה מאות שנים, אמירה כזו הייתה מביאה לכך שיעלו אתכם על המוקד. אמא שבּתה חלתה בסרטן מתפללת, "אלוהים, אנא רפא את בתי," ולא, "אלוהים היקר, אני יודעת שאסור שיהיו לדת השלכות שניתנות להפרכה, מה שאומר בפועל שאינך יכול לרפא את בתי, אז... ובכן, בעיקרון, אני מתפללת כדי לגרום לעצמי להרגיש טוב יותר, במקום לעשות משהו שבאמת יעזור לבת שלי."

אך אנשים קוראים על כך ש"אי אפשר להוכיח או להפריך דת באמצעות ראיות עובדתיות," ואז, בפעם הבאה שהם נתקלים בראיה שמפריכה דת, המוח שלהם משלים את התבנית. ישנם אפילו אתאיסטים שמצטטים את האבסורד הזה ללא היסוס. אילו היו הוגים את הרעיון בעצמם, במקום לשמוע אותו ממישהו אחר, הם היו ביקורתיים יותר.

מוות. השלימו את התבנית: "המוות נותן משמעות לחיים."

זה מתסכל, לשוחח עם אנשים טובים והגונים – אנשים שלעולם לא יעלו על דעתם *מיוזמתם* את רעיון הכחדת המין האנושי – להעלות את נושא הסכנה הקיומית, ולשמוע אותם אומרים, "ובכן, אולי המין האנושי לא ראוי לשרוד." הם אנשים שלעולם, בשום מצב, לא היו יורים בילד שלהם, שהוא חלק מהמין האנושי. אבל המוח שלהם משלים לבד את התבנית.

אילו תבניות המוח שלכם משלים – תבניות שמעולם לא בחרתם שיהיו שם?

רציונליות. השלימו את התבנית: "האהבה אינה רציונלית."

אילו הרעיון הזה היה עולה בדעתכם, כמחשבה חדשה לחלוטין,

* ראו פרק 16.

כיצד הייתם בוחנים אותו בצורה ביקורתית? אני יודע מה *אני* הייתי חושב, אבל מה *אתם* הייתם חושבים? זה יכול להיות קשה לראות את הדברים בעיניים חדשות. נסו למנוע ממוחכם להשלים את התבנית בדרך הסטנדרטית, הלא־מפתיעה, הידועה. ייתכן שאין תשובה טובה יותר מזו הידועה, אבל אינכם יכולים *לחשוב* כראוי על תשובה עד שתצליחו לעצור את מוחכם מלהשלים את התבנית באופן אוטומטי.

כעת, לאחר שקראתם זאת, בפעם הבאה שתשמעו מישהו חוזר ללא היסוס על אמרה שאתם חושבים שהיא טיפשית או שקרית, אתם תחשבו: "מחשבות מטמון." האמונה שלי נמצאת עכשיו במוחכם, מחכה להשלים את התבנית. אך האם היא נכונה? אל תתנו למוח שלכם להשלים את התבנית לבד! *חישבו!*

91

קופסת ה"מחוץ לקופסה"

בכל פעם שמישהו מפציר בכם "לחשוב מחוץ לקופסה", הוא בדרך כלל, *לנוחיותכם*, מצביע בדיוק על המקום שבו נמצא ה"מחוץ לקופסה". האין זה מצחיק שכל אלו ש"הולכים נגד הזרם" מתלבשים אותו דבר...

בבינה מלאכותית, לכל מי שלא מצוי בתחום יש תוצאת מטמון מוכנה מראש עבור *רעיון AI מהפכני, חדש ומבריק* – רשת נוירונים, שעובדת בדיוק כמו המוח האנושי! רעיון AI חדש. השלימו את התבנית: "בינה מלאכותית לוגית, חרף כל ההבטחות הגדולות, כשלה מלספק אינטליגנציה אמיתית במשך עשורים – מה שאנחנו צריכים זה רשת נוירונים!"*

מחשבת המטמון הזו קיימת כבר שלושה עשורים. עדיין אין אינטליגנציה מלאכותית כללית. אבל, משום מה, כל מי שלא בתחום יודע שרשתות נוירונים הן הרעיון־החדש־שיחליף־את־הפרדיגמה־השלטת, מאז שהומצא מנגנון הפעפוע־לאחור** בשנות ה-70. ממש היפים מזדקנים.

---

* כאמור בהערה למאמר "הטיות: מבוא" בכרך א', ספר זה נכתב לפני מהפכת ה-LLM ששינתה ללא היכר את תחום המחקר של בינה מלאכותית.

** *פעפוע־לאחור* (backpropagation) הוא אלגוריתם המשמש לאימון רשתות נוירונים על ידי חישוב השגיאה בפלט והעברתה אחורה דרך שכבות הרשת. מנגנון זה משמש עד היום כמעט בכל אימון של מערכת למידת מכונה עמוקה.

דימויים מרדניים, מטבעם, אינם מאפשרים שום סטייה מהנורמה. אם אינך לובש שחור, איך אנשים יידעו שאתה אמן מיוסר? איך אנשים יזהו ייחודיות אם אינך מתאים לתבנית הסטנדרטית של איך שייחודיות אמורה להיראות? איך מישהו יזהה שיש לך רעיון AI מהפכני, אם הוא לא קשור לרשתות נוירונים?

דוגמה נוספת לאותה קלישאה היא ספרות "חתרנית", שכולה נשמעת אותו דבר, ומגובה על ידי חבורת מורדים זעירה ומתריסה ששולטת בכל המחלקה לספרות אנגלית באוניברסיטה. כפי ששואל המשתמש Anonymous בבלוג של סקוט ארונסון:

> האם ספרות חתרנית כלשהי שקראת גרמה לך לשנות אף במקצת את דעותיך הפוליטיות?

או כפי שמציין Lizard:

> המהפכה כבר שודרה בטלוויזיה. המהפכה *מוסחרה*. המהפכה היא מוצר, סגנון חיים ארוז מראש, שזמין בקניון הקרוב לביתך. 19.95 דולר יקנו לך את המסכה השחורה, פחית הספריי, שלט ההפגנה "למחוץ את הפשיסטים", והבלוג האישי שבו תוכל לכתוב על האלימות המשטרתית שסבלת ממנה כשקשרת את עצמך לברז כיבוי אש. הקפיטליזם למד איך למכור אנטי־קפיטליזם.

רבים בעמק הסיליקון הבחינו שרובם המכריע של משקיעי הון־סיכון, בכל זמן נתון, רודפים כולם אחר אותה חדשנות מהפכנית, וזו אותה חדשנות מהפכנית שהונפקה בבורסה שישה חודשים קודם לכן. זוהי אבחנה קטלנית *במיוחד* בתחום ההון־סיכון, משום שיש מניע כלכלי ישיר שלא ללכת אחרי העדר – או שמישהו אחר כבר מָפתח את אותו המוצר, או שמישהו אחר מציע מחיר גבוה מדי עבור אותו הסטארט־אפ. סטיב ג'וּרבֶטסון סיפר לי פעם שבקרן הון־הסיכון Draper Fisher Jurvetson מספיקה הסכמה של שני שותפים בלבד על מנת לממן כל סטארט־אפ עד 1.5 מיליון דולר. אבל אם כל השותפים מסכימים שמשהו נשמע כמו רעיון טוב, הם לא ישקיעו בו.

הלוואי שוועדות מענקים היו שפויות עד כדי כך.

הבעיה עם מקוריות היא שצריך באמת *לחשוב* כדי להשיג אותה, במקום לתת למוח להשלים בעצמו את התבנית. אין שום מקום נוח ומסומן בשלט "מחוץ לקופסה" אליו אפשר להגיע מיד. יש בזה איכות כמעט זן־בודהיסטית – כמו העובדה שאי אפשר ללמד 'סאטורי' במילים, משום שסאטורי היא החוויה שבה המילים מכזיבות אותך. ככל שתנסו לעקוב אחר מילותיו של המורה, כך תתרחקו מהשגת תודעה ריקה.

יש סיבה, לדעתי, לכך שאנשים אינם משיגים חדשנות באמצעות חתירה אליה. תכונות כמו אמת או עיצוב טוב אינן תלויות בחדשנות: שתיים ועוד שתיים שווה ארבע, אכן, באמת ובתמים, למרות שזה מה שכולם חושבים. אנשים ששואפים לגלות אמת או להמציא עיצובים טובים, אכן עשויים להשיג יצירתיות עם הזמן. לא כל שינוי הוא שיפור, אבל כל שיפור הוא שינוי.

כל שיפור הוא שינוי, אבל לא כל שינוי הוא שיפור. זה שאומר "אני רוצה לבנות מלכודת עכברים מקורית!" ולא "אני רוצה לבנות את מלכודת העכברים האופטימלית!" כמעט תמיד רק רוצה *להיתפס* כמקורי.

"מקוריות" במובן זה היא תכונה חברתית במהותה, כי ניתן לקבוע אותה רק באמצעות השוואה לאנשים אחרים. אז המוח פשוט משלים את התבנית הסטנדרטית למה שנתפס כ"מקורי", והחברים מהנהנים בהסכמה ואומרים שזה חתרני.

ספרי עסקים תמיד אומרים לכם, לנוחיותכם, לאן הגבינה שלכם הוזזה. אחרת הקוראים היו נותרים לתהות, "איפה נמצא ה'מחוץ לקופסה' הזה שאליו אני אמור ללכת?"

חשיבה אמיתית, בדומה לסאטורי, היא פעולת תודעה נטולת מילים.

הפילוסופים הדגולים של מונטי פייתון ניסחו זאת בצורה הטובה ביותר ב*בריאן כוכב עליון*:[1]

"אתם מוכרחים לחשוב בעצמכם! כולכם יחודיים!"
"כן, כולנו יחודיים!"
"כולכם שונים!"
"כן, כולנו שונים!"
"כולכם חייבים להבין זאת בעצמכם!"
"כן, כולנו חייבים להבין זאת בעצמנו!"

---

1. Graham Chapman et al., *Monty Python's The Life of Brian (of Nazareth)* (Eyre Methuen, 1979).

92

מבט מקורי

היות שרוברט מ. פירסיג ניסח זאת היטב, אעתיק את דבריו. איני יודע אם זהו סיפור אמיתי או לא, אך בכל אופן, הוא נכון.

למשל, הציקה לו העובדה שיש סטודנטים שאין להם מה לומר. תחילה סבר שהדבר נובע מעצלותם, אך אחר־כך התברר לו שאין הדבר כן. הם פשוט לא יכלו לחשוב על משהו לומר.

אחת מהם, בחורה בעלת משקפיים עבי־זגוגיות, הצהירה שברצונה לכתוב חיבור בן חמש מאות מילה על אודות ארה״ב. הוא כבר התרגל לרגשי האכזבה המתעוררים למשמע הצהרות כאלה, והציע בלי לפגוע בכבודה, שתמצם את הנושא לבּוֹזֶמַן* בלבד.

כשהגיע מועד הגשת החיבור, לא הצליחה היא עדיין לכותבו – והיתה נסערת למדי. היא ניסתה וניסתה, אך פשוט לא יכלה לחשוב על משהו לומר.

הדבר הזה פשוט סיבך אותו. עכשיו *הוא* לא יכול היה לחשוב על משהו לומר. השתרר שקט, ואז באה תשובה משונה: ״צמצמי את הנושא *לדחוב הראשי* של בוזמן.״ היתה זו הברקה פתאומית.

היה הנהנה בצייתנות ויצאה. אולם קצת לפני השיעור הבא חזרה ובאה אליו כשדמעות בעיניה, והפעם במצוקה *אמיתית*, מצוקה שבוודאי קיננה בה זמן רב. היא עדיין לא יכלה לחשוב על משהו לומר, ולא הצליחה להבין מדוע; שהרי אם לא הצליחה לחשוב

* *בּוֹזֶמַן* היא העיר הרביעית בגודלה במונטנה שבארצות הברית.

משהו על כל *בוזמן*, מדוע שלא תצליח לחשוב משהו על *רחוב אחד* בלבד.

הוא זעם. "את אינך *מתבוננת!*" אמר. הוא נזכר בסילוקו מהאוניברסיטה עקב כך שהיה לו *יותר מדי* מה לומר. להסבר כל עובדה *יש אין־סוף* היפותזות. ככל שאדם *מתבונן* יותר, כן הוא *רואה* יותר. היא למעשה לא התבוננה כלל, ומשום־מה לא הרגישה בכך.

הוא אמר לה בכעס: "צמצמי את הנושא *לחזיתו של בית אחד* ברחוב הראשי של בוזמן. בית האופרה. התחילי בלבנה העליונה שמשמאל."

עיניה שמאחורי המשקפיים עבי־הזגוגית נפערו.

היא באה לשיעור הבא במבט נבוך והושיטה לו חיבור בן חמש מאות מילה על חזית בית־האופרה ברחוב הראשי של בוזמן, מונטאנה. "ישבתי ב'סטקיה' *שממול*," אמרה, "והתחלתי לכתוב על הלבנה הראשונה, אחר־כך על השניה, ואז בלבנה השלישית התחיל הכל לשטוף — ולא יכולתי לעצור. כולם חשבו שנטרפה דעתי ולא פסקו מללגלג עלי, אך הנה כל מה שכתבתי. אינני מבינה מה קרה."

גם הוא לא הבין; אך בצעדותיו הארוכות ברחובות העיר הרהר על כך והגיע למסקנה, שהיא נעצרה אל־נכון מכוחו של אותו מחסום ששיתק גם אותו ביום הראשון להוראה. היא נחסמה משום שהיא ניסתה לחזור בכתיבתה על דברים שכבר שמעה, ממש כשם שביום הראשון ניסה לחזור על דברים שהחליט מראש לומר. היא לא יכלה לחשוב על משהו לכתוב על אודות בוזמן, מפני שלא יכלה להיזכר במשהו ששמעה ושראוי לחזור עליו. באורח משונה לא היתה מודעת לכך שהיא יכולה להתבונן בעיניים רעננות, תוך כדי כתיבה, בלי להיות כבולה למה שמישהו כבר אמר. צמצום הנושא ללבנה אחת פרץ את המחסום, מפני שהיה ברור כל־כך שהיא *מוכרחה*, לפחות במידת מה, להתבונן ישירות במו עיניה.

— רוברט מ. פירסיג,

זן ואומנות אחזקת האופנוע[*]

* תרגום לעברית: יהודה רגבים.

93

יותר מוזר מההיסטוריה

נניח שהייתי אומר לכם שאני יודע בוודאות שההצהרות הבאות נכונות:

- אם תצבעו את עצמכם בצבע *מדויק* כלשהו בין כחול לירוק, כוח המשיכה הפועל עליכם יתהפך ויגרום לכם ליפול כלפי מעלה.
- בעתיד, השמיים יהיו מלאים במיליארדי כדורים שחורים מרחפים. כל כדור יהיה גדול יותר מכל ספינות הצפלין שהיו אי פעם גם יחד. אם תציעו לאחד הכדורים כסף, הוא יוריד מהשמיים זונה ממין זכר על חבל בנג׳י.
- הנכדים שלכם יחשבו שזה לא רק טיפשי, אלא *מדושע*, לכלוא גנבים במקום להכות אותם בישבן.

הייתם חושבים שאני משוגע, נכון?

עכשיו נניח שהשנה היא 1901, ואתם צריכים לבחור בין להאמין להצהרות לעיל, לבין הצהרות כגון אלו:

- המהירות המרבית שבה שני עצמים יכולים להיראות כְּנָעִים זה ביחס לזה היא בדיוק 1,079,252,848.8 קמ״ש. אם תקפצו על רכבת שנוסעת כמעט במהירות הזו ותירו באקדח מבעד לחלון, יחידות האורך היסודיות ישתנו, כך *שלכם* ייראה שהקליע טס

קדימה, אבל אנשים אחרים יראו משהו שונה. אה, וגם הזמן משתנה בהתאם לצופה.

- בעתיד, תהיה רשת עולמית מקושרת היטב של מיליארדי מכונות חישוב, שלכל אחת מהן יש יותר כוח חישוב מכל המכונות שקדמו לשנת 1901 גם יחד. אחד השימושים העיקריים של רשת זו יהיה העברת תמונות נעות של יחסי־מין לסביים באמצעות העמדת פנים שהן עשויות ממספרים.
- הנכדים שלכם יחשבו שזה לא רק טיפשי, אלא *מרושע*, לומר למישהי שאסור לה להיות נשיאת ארצות הברית מפני שהיא שחורה.

מבוסס על הערה של רובין הנסון: *"אני תוהה אם אפשר לתאר בפירוט מספיק סיפור בדיוני על מציאות חלופית – מציאות שהדורות הקודמים לא היו מסוגלים להבחין בינה לבין המציאות – כדי להמחיש עד כמה האמת שהתבררה לבסוף הייתה מפתיעה."*

94

הכשל הלוגי של הכללה מראיות בדיוניות

כשאני מתחיל לדבר עם אנשים על בינה מלאכותית מתקדמת, מה הדבר הראשון שאומרים לי ברוב הפעמים?

"אה, אתה מתכוון כמו בסרטי *שליחות קטלנית* / *המטריקס* / הרובוטים של אסימוב!"

ואני משיב, "ובכן, לא, לא בדיוק. אני משתדל להימנע מהכשל הלוגי של הכללה מראיות בדיוניות."

יש אנשים שמבינים מיד וצוחקים. אחרים מגֵנים על עצם השימוש בדוגמה, וטוענים שלא מדובר בכשל.

מה רע בשימוש בסרטים או סיפורים בדיוניים בתור נקודת פתיחה לדיון? הרי אף אחד לא טוען שזה *אמיתי*. איפה השקר, מהו החטא הרציונליסטי? מדע בדיוני מייצג את הניסיון של המחבר לתאר את העתיד. למה לא לנצל את החשיבה שנעשתה כבר למעננו, במקום להתחיל מאפס?

לא כל מעידה בריקוד המדויק של הרציונליות מורכבת מאמונה מפורשת בשקר. ישנן דרכים מעודנות יותר לטעות.

ראשית, הבה ניפטר מהרעיון שמדע בדיוני מייצג ניסיון רציונלי של ממש לחזות את העתיד. אפילו הקפדנים שבסופרי המדע הבדיוני הם, בראש ובראשונה, מספרי סיפורים. דרישותיה של מלאכת הסיפור

אינן זהות לאלו של מלאכת החיזוי. כפי שמציין ניק בוֹסטרוֹם:[1]

> מתי בפעם האחרונה ראיתם סרט על כך שהאנושות פשוט נכחדת פתאום (ללא אזהרה ומבלי שתוחלף על ידי ציוויליזציה אחרת)? אף שתרחיש כזה עשוי להיות סביר בהרבה מתרחיש שבו גיבורים אנושיים הודפים בהצלחה פלישה של מפלצות או לוחמים רובוטיים, זה לא יהיה כיף גדול לצפות בו.

מכאן שישנם עיוותים מסוימים בספרות הבדיונית. אבל ניסיון נקודתי לתקן את העיוותים הללו אינו מספיק. סיפור *לעולם* אינו ניסיון רציונלי לניתוח, אפילו לא אצל סופרי המדע הבדיוני הקפדנים ביותר, משום שסיפורים אינם משתמשים בהתפלגויות הסתברות. אני מדגים זאת כך:

> בוב מֶרקֶלתָ׳אד החליק בזהירות דרך דלת החללית החייזרית, מביט ימינה ואז שמאלה (או שמאלה ואז ימינה) כדי לראות אם נותרו עוד ממפלצות החלל האימתניות. בידיו היה הנשק היחיד שנמצא כיעיל נגד מפלצות החלל: חרב־חלל מחושלת מטיטניום טהור בסבירות של 30%, לום ברזל רגיל בסבירות של 20%, ודיסקוס שחור ובוהק שנמצא בהריסות העשנות של סטוֹנהֶנג׳ בסבירות של 45%, כאשר ה־5% הנותרים מתחלקים על פני יותר מדי תוצאות משניות מכדי לפרט כאן.
> מרקלת׳אד (אף שיש סיכוי משמעותי שסוזן וויפֶלפוּפֶר הייתה שם במקומו) עשה שני צעדים קדימה או צעד אחד אחורה, כששאגה אדירה פלחה את השקט של מנעל האוויר השחור! או את זמזום הרקע החרישי של מנעל האוויר הלבן! למרות שאַמפֶר ו־וופי (1997) טוענים שמרקלת׳אד נטרף בשלב זה, ספֶּקֶלבֶּקֶל (2003) מציין ש—

דמויות יכולות להיות בורות, אבל *המחבר* לא יכול לומר את שלוש מילות הקסם "אני לא יודע." גיבור הסיפור מוכרח לטוות חוט יחיד דרך העתיד, מלא בפרטים שגורמים לסיפור לקרום עור וגידים, החל מהיחס העתידני כהלכה של וויפלפופר כלפי פמיניזם וכלה בצבע

העגילים שלה.

ואז, כל הפרטים המכבידים וההנחות המפוקפקות הללו נארזים יחד וזוכים לתווית קצרה, שיוצרת אשליה שהם חבילה אחת בודדת.

בבעיות עם מרחבי תשובות גדולים, הקושי הגדול ביותר הוא לא *אימות* התשובה הנכונה, אלא האיתור שלה במרחב התשובות מלכתחילה. אם מישהו מתחיל בשאלה האם בינה מלאכותית תכניס אותנו לקפסולות כמו ב*מטריקס* או לא, הוא קופץ להנחה של 100 ביט, ללא 98 ביטים תואמים של ראיות כדי לאתר אותה במרחב התשובות כאפשרות שראויה להתייחסות מפורשת. דרוש רק קומץ ראיות אחרי 98 הביטים הראשונים כדי לקדם את האפשרות הזו לכמעט־ודאות, מה שמלמד משהו על המקום בו מתבצעת כמעט כל העבודה.

הצעד ה"מקדים" של איתור אפשרויות שראויות להתייחסות מפורשת כולל צעדים כמו: שקילת מה שאתם יודעים ומה שאינכם יודעים, מה שאתם יכולים ולא יכולים לחזות, מאמץ מכוון להימנע מהטיית האבסורד והרחבה של רווחי הסמך, הרהור באילו שאלות הן החשובות, ניסיון להתכונן ל"ברבורים שחורים" אפשריים ומחשבה על דברים ש(עד כה) לא ידעתם שאינכם יודעים. קפיצה לשאלה "*המטריקס*: כן או לא?" מדלגת על כל זה.

כל נושא ונותן מקצועי יודע ששליטה בתנאי הדיון משמעה שליטה כמעט מוחלטת בתוצאות הדיון. אם מתחילים מלחשוב על *המטריקס*, עולה המחשבה על צבאות רובוטים צועדים בסך שמביסים את בני האדם לאחר מאבק ארוך – ולא אינטליגנציית־על שמקישה באצבעות ננו־טכנולוגיות. הדיון מתמקד במאבק של "אנחנו או הם", ומפנה את תשומת הלב לשאלות כמו "מי ינצח?" "מי צריך לנצח?" ו"האם בינה מלאכותית באמת תהיה כזאת?" האווירה הכללית שנוצרת היא של בידור, של "מהו חזון העתיד המופלא שלך?"

מה שאובד בתוך הריק המהדהד הם: שיקולים על יותר מעיצוב אפשרי אחד לתודעה ש"בינה מלאכותית" עשויה לממש; התלות של העתיד בתנאי הפתיחה; העוצמה של אינטליגנציה חכמה־מאדם והטיעון לגבי היעדר היכולת לחזות אותה; אנשים שלוקחים את

העניין כולו ברצינות ומנסים לעשות לגביו משהו.

אילו משחית־דיונים ערמומי היה מחליט שהתוצאה המועדפת עליו תושג בצורה הטובה ביותר בכך שיגרום למתדיינים להתחיל בהפרכת *שליחות קטלנית*, הוא היה עושה עבודה טובה בהטיית מסגרת הדיון. בוויכוח אודות פיקוח על נשק, דובר איגוד הרובאים אינו רוצה להיות מוצג כ"משוגע להדק", והצד השני אינו רוצה להיות מוצג כ"תומך בפירוק קורבנות מנשקם". למה שתאפשרו הטיית מסגרת חמורה באותה מידה על ידי תסריטאים מהוליווד, אפילו בטעות?

עיתונאים לא אומרים לי, "העתיד יהיה כמו בסרט *2001*." אבל הם שואלים, "האם העתיד יהיה כמו בסרט *2001*, או כמו בסרט *אינטליגנציה מלאכותית*?" זוהי בעיית מסגור עצומה בדיוק כמו לשאול "האם עלינו לקצץ קצבאות לנזקקים, או להעלות מסים לעשירים?"

בסביבה הקדמונית, לא היו תמונות נעות. מה שראיתם במו עיניכם היה אמיתי. הצצה רגעית במילה בודדת מסוגלת להטרים אותנו ולהפוך מחשבות הקשורות למילה לזמינות יותר, עם השפעה חזקה ומוכחת על אומדני ההסתברות שלנו. כמה הרס אתם חושבים שסרט בן שעתיים יכול לזרוע בכושר השיפוט שלכם? יהיה זה מספיק קשה לתקן את הנזק באמצעות מאמץ מחשבתי מכוון – למה להזמין את הערפד אליכם הביתה?* בשחמט או בגו, כל מהלך מבוזבז הוא הפסד. ברציונליות, כל השפעה שאינה מבוססת־ראיות היא (בממוצע) אנטרופית.

האם צופי קולנוע מצליחים לא־להאמין למה שהם רואים? עד כמה שאני יכול לשפוט, מעט מאוד צופים מתנהגים כאילו צפו *ישירות* בעתיד של כדור הארץ. אנשים שצפו בסרטי *שליחות קטלנית* לא הסתתרו במקלטים גרעיניים ב-29 באוגוסט 1997.** אבל אלו שנופלים בכשל הלוגי הזה מתנהגים כאילו ראו את אירועי הסרט

---

* על פי הפולקלור, ערפד אינו יכול להיכנס לבית אלא אם בעליו הזמינו אותו פנימה.

** ספוילר: תאריך זה הוא 'יום הדין' בסרטי *שליחות קטלנית*, בו התרחשה שואה גרעינית כתוצאה ממאבק בבינה מלאכותית שיצאה משליטה.

מתרחשים על *כוכב לכת אחר*. לא כדור הארץ, אך מקום דומה לכדור הארץ.

אתם אומרים, "נניח שנבנה בינה מלאכותית חכמה מאוד," והם אומרים, "אבל האם לא זה מה שהוביל למלחמה גרעינית ב*שליחות קטלנית*?" ככל שאני יכול לשפוט, זהו היגיון זהה לחלוטין, עד לרמת טון הדיבור, לזה של מישהו שעשוי לומר: "אבל האם לא זה מה שהוביל למלחמה גרעינית באלפא קנטאורי?" או "האם לא זה מה שהוביל לנפילת עיר־המדינה האיטלקית פיקולו במאה הארבע־עשרה?" לא מאמינים לסרט, אבל הוא זמין. הוא זוכה להתייחסות, לא כנבואה אבל כן כתרחיש היסטורי. האם ההיסטוריה תחזור על עצמה? אין לדעת.

בדיון שנערך לאחרונה על פיצוץ אינטליגנציה,* מישהו ציין שוִינג'י כנראה לא חושב שממשקי מוח־מחשב יגדילו בהרבה את האינטליגנציה, וציטט את *Marooned in Realtime* ואת טונץ' בלומנטל, שהיה הנוסע המתקדם ביותר אבל לא נראה חזק במיוחד.** השבתי בזעם, "אבל טונץ' איבד את רוב החומרה שלו! הוא היה נכה!" ואז תפסתי את עצמי וחשבתי: מה *לכל הרוחות* אני אומר.

האם הסוגיה לא אמורה להיבחן לגופה, ללא קשר לאופן שבו וינג'י עיצב את הדמויות שלו? טונץ' בלומנטל אינו "נכה", הוא *דמיוני*. יכולתי לומר "וינג'י בחר לתאר את טונץ' כנכה, מסיבות שאולי היו קשורות ואולי לא היו קשורות לתחזית האישית שלו לגבי העתיד," וזה היה מעניק לבחירה הספרותית שלו משקל ראייתי הולם. אני לא יכול לומר "טונץ' היה נכה." לא היה דבר, כשמדובר בטונץ' בלומנטל.

בכוונה השארתי טעות שעשיתי בתחילת הטיוטה הראשונה של פרק זה: "אחרים מגנים על עצם השימוש *בדוגמה*, וטוענים שלא

---

* ראו הערה בפרק 62.

** ורנר סטיפן וינג'י (1944-2024) היה מדען וסופר מדע בדיוני. טונץ' בלומנטל (Tunç Blumenthal) הוא דמות מספרו *Marooned in Realtime*, אשר נוסעת עשרות מליוני שנים קדימה בזמן.

מדובר בכשל.״ אבל *המטריקס אינו* דוגמה!

פגם קרוב לכך הוא הכשל הלוגי של טיעון מראיות דמיוניות: ״ובכן, אם *כן* היית מגיע לקצה הקשת בענן, *היית* מוצא קדירת זהב — מה שמוכיח את הטענה שלי!״ (עדכון על סמך ראיות שנחזו אך לא נצפו הינו תמונת הראי המתמטית של הטיית החוכמה בדיעבד.)

למוח ישנם מנגנונים רבים עבור הכללה מתוך תצפית, לא רק היוריסטיקת הזמינות. אתם רואים שלוש זברות, יוצרים את הקטגוריה ״זברה״, והקטגוריה הזו מגלמת בתוכה הסקה תפיסתית אוטומטית. יצורים דמויי־סוס עם פסים לבנים ושחורים מסווגים כ״זברות״, ולכן הם מהירים וטובים למאכל. מצופה מהם להיות דומים לזברות אחרות שנצפו.

לכן, כשאנשים רואים (בתמונות נעות) שלושה פועלי בּוֹרג,* המוח שלהם יוצר אוטומטית את הקטגוריה ״בורג״, והם מסיקים אוטומטית שבני אדם עם ממשקי מוח־מחשב שייכים למחלקת ״בורג״ ויהיו דומים לבורג אחרים שנצפו: קרים, נטולי חמלה, לבושים בעור שחור, והולכים בצעדים מכניים כבדים. עיתונאים לא מאמינים שהעתיד *יכיל* בורג — הם לא מאמינים *שמסע בין כוכבים* הוא נבואה. אבל כשמישהו מדבר על ממשקי מוח־מחשב, הם חושבים: ״האם העתיד יכיל בורג?״ הם לא חושבים, ״איך אני יודע שטלפתיה מבוססת־מחשב הופכת אנשים לפחות נחמדים?״ או, ״מעולם לא ראיתי בורג, וכך גם אף אחד אחר.״ או, ״אני בונה סטריאוטיפ גזעי שמבוסס *מילולית* על אפס ראיות.״

כפי שאמר ג׳ורג׳ אורוול על קלישאות:[2]

> מה שנחוץ יותר מכל הוא לתת למשמעות לבחור את המילה, ולא להפך... כשאתם חושבים על משהו מופשט אתם נוטים יותר

* *הבּוֹרג* (Borg) הוא גזע בדיוני המופיע כיריב מרכזי ביקום של סדרת הטלוויזיה והסרטים *מסע בין כוכבים*. יצורים אלו משלבים ביולוגיה (לרוב של יצורים דמויי אדם שכבשו) עם שתלים טכנולוגיים מורכבים, כולל ממשקי מוח־מחשב וחלקי גוף מכניים.

להשתמש במילים מהרגע הראשון, ואם לא תתאמצו במודע למנוע זאת, הניב הקיים יסתער פנימה ויעשה את העבודה בשבילכם, על חשבון טשטוש או אפילו שינוי המשמעות המקורית של מחשבתכם.

אולם להערכתי, ההיבט *המזיק ביותר* של שימוש בדמיון של אנשים אחרים הוא בכך שהוא מונע מאנשים להשתמש בדמיון של עצמם. כפי שאמר רוברט פירסיג:[3]*

היא נחסמה משום שהיא ניסתה לחזור בכתיבתה על דברים שכבר שמעה, ממש כשם שביום הראשון ניסה לחזור על דברים שהחליט מראש לומר. היא לא יכלה לחשוב על משהו לכתוב על אודות בּוֹזמֶן, מפני שלא יכלה להיזכר במשהו ששמעה ושראוי לחזור עליו. באורח משונה לא היתה מודעת לכך שהיא יכולה להתבונן בעיניים רעננות, תוך כדי כתיבה, בלי להיות כבולה למה שמישהו כבר אמר.

בִּדיוֹנות שעולים בזיכרון מסתערים פנימה וחושבים במקומכם. הם מהווים תחליף להתבוננות – הנוחות ההרסנית מכל.

---

1. Nick Bostrom, "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards," Journal of Evolution and Technology 9 (2002), http://www.jetpress.org/volume9/risks.html.
2. Orwell, "Politics and the English Language."
3. Pirsig, Zen and the Art of Motorcycle Maintenance.

* ראו פרק 92.

95

סוד הצמצום

מה שנכון עבור תפוח אחד עשוי שלא להיות נכון עבור תפוח אחר. בכך, ניתן לומר יותר לגבי תפוח יחידי מאשר על כל תפוחי העולם יחדיו.

— שנים־עשר מעלות הרציונליות

בתחום התמחותם, אנשים מבינים את החשיבות שבצמצום. מכונאי מבין את ההבדל בין קרבורטור ורדיאטור, ולא יתייחס לשניהם כאל "חלקי רכב". צייד־לקט מבדיל בין אריה לפנתר. שרת לא שוטף את הרצפה עם מנקה חלונות, אפילו אם הבקבוקים נראים דומים למי שאינו שולט ברזי המקצוע.

מחוץ לתחום התמחותם, אנשים שוגים לעיתים קרובות בניסיון להרחיב את משמעותה של מילה כך שתתפוס כמה שיותר שטח. האין זה משובח יותר, חכם יותר, מרשים יותר לדבר על כל תפוחי העולם יחדיו? כמה נעלה זה בוודאי *להסביר את פעולת המחשבה האנושית בכללותה*, ללא התעסקות בשאלות קטנות המסיחות את הדעת, כגון כיצד בני אדם ממציאים שיטות לפתרון קוביה הונגרית.* בהחלט, נראה שכמעט אין צורך לבחון שאלות *ספציפיות*. האם תיאוריה

* ראו פרק 37.

כללית איננה הישג ראוי בפני עצמו?

זהו טבעו של האדם הסקרן להרים אבן אחת מִבֵּינות מיליוני חלוקי נחל על החוף, ולראות בה משהו חדש, משהו מעניין, משהו שונה. אתם קוראים לאבנים מסוג זה "יהלום", ושואלים מה עשוי לייחד אותם – אילו תכונות אולי יש להם במשותף, מלבד הנצנוץ שלכד את תשומת לבכם. ואז בא מישהו ואומר:

"למה שלא נקרא גם לאבן *הזו* יהלום? וגם לזו, וגם לזו?" הוא נלהב, ויש לו כוונה טובה. שכן זה נראה לא־דמוקרטי, מתבדל, אליטיסטי ובלתי־הוליסטי לקרוא לחלק מהאבנים "יהלומים" ולאחרות לא. זוהי קצת... *צרות מוחין*... אם תמחלו לי על הביטוי. לא מאוד *פתוח*, בקושי *מכיל*, כלל לא *קהילתי*.

אולי תחשבו שזה פואטי, לתת למילה אחת משמעויות רבות, ובכך לזרוע סביבה שלל הקשרים שונים ומגוונים. אך אפילו משוררים, אם הם טובים, לומדים לראות את העולם בצורה מדויקת. אין זה מספיק להשוות אהבה לפרח. אהבה קנאית ונכזבת אינה דומה לאהבתם של בני זוג שנשואים עשורים. אם אתם צריכים פרח שיסמל אהבה קנאית, עליכם לצאת אל הגן, להתבונן, ולבצע הבחנות מעודנות – למצוא פרח עם ריח משכר, צבע בוהק וקוצים. אפילו אם בכוונתכם ליצור גוונים של משמעות ולרמוז להקשרים שונים, עליכם לעקוב במדויק אחר המשמעויות אליהן אתם רוצים לגוון ולרמוז.

זהו חלק חיוני מכישוריו של הרציונליסט – או אפילו של המשורר! – להתמקד באופן צר באבנים לא רגילות שיש להן תכונה ייחודית. להתבונן בתכונות שאותן אבנים אלה – ואבנים אלה בלבד! – חולקות אחת עם השניה. זהו אינו פשע.

לגמרי מותר לביולוגים אבולוציוניים להסביר *רק* את תופעותיהם של יצורים חיים, ולא את "אבולוציית" הכוכבים או "אבולוציית" הטכנולוגיה. אבוי, ישנן נשמות טועות שמשתמשות באותה המילה "אבולוציה" כדי לתאר את התופעות הנובעות מבריה טבעית בקרב יצורים מתרבים, *וגם* את האופן האקראי בו מתהווים כוכבים, *וגם* את ההתפתחות המתוכננת־תבונית של טכנולוגיה. וכפי שכולנו יודעים, אם

אנשים משתמשים באותה המילה, זה הרי אותו הדבר. נוכל להכליל כל מה שנדמה לנו שאנחנו יודעים על אבולוציה ביולוגית, גם על תחום הטכנולוגיה. כל מי שאומר לנו אחרת הוא בסך הכל טרחן חסר תועלת. לא ייתכן שהבורות שלנו בתורת האבולוציה המודרנית היא כה מוחלטת עד שלא נוכל להבדיל בין קרבורטור לרדיאטור. זה בלתי מתקבל על הדעת. אין זאת אלא שהאדם *האחר* – אתם יודעים, זה שלמד את המתמטיקה הרלוונטית – פשוט טיפש מכדי לראות את ההקשרים.

ומה יכול להיות נעלה יותר מראיית הקשרים? ללא ספק, החכמים בבני האדם הם אותם גורואים של ניו־אייג׳ שאומרים, ״הכל קשור להכל.״ אם אי פעם תאמרו זאת בקול רם, עליכם להשתהות מעט, כדי לאפשר לכולם להתאושש מההלם המוחלט המלווה שמיעת חוכמה עמוקה זו.

קיימת התאמה טריוויאלית בין גרף לבין הגרף המשלים שלו.* גרף שלם (קשיר לחלוטין), שבו יש קשת בין כל שני צמתים, מעביר בדיוק את אותה כמות מידע כמו גרף ללא קשתות כלל. הגרפים החשובים הם אלה שבהם דברים מסוימים *אינם* מחוברים לדברים אחרים.

כאשר אלו שאינם מוארים מנסים להיות עמוקים, הם עורכים השוואות מילוליות אינסופיות בין הנושא הזה לנושא ההוא, שדומה לזה, שדומה להוא... עד שהגרף שלהם מחובר לחלוטין וגם חסר תועלת לחלוטין. התרופה לכך היא ידע ממשי ולמידה מעמיקה. כאשר אתם מבינים דברים לפרטי פרטים, אתם מסוגלים לראות במה הם *אינם* דומים, ולהתחיל *להסיר* בחדווה קשתות מהגרף שלכם.

בדומה לכך, הקטגוריות החשובות הן אלה שאינן מכילות את כל מה שקיים ביקום. השערות טובות מסוגלות להסביר רק חלק מהתוצאות האפשריות, ולא אחרות.

זה היה בסדר גמור מצדו של אייזק ניוטון להסביר *רק* את כוח המשיכה, *רק* את האופן שבו דברים נופלים מטה – וכיצד כוכבי הלכת

* בתורת הגרפים, *גרף משלים* הוא אותה קבוצת צמתים, כאשר קשת קיימת בגרף משלים אם ורק אם היא חסרה בגרף המקורי. כך, שני הגרפים הם למעשה אותו מבנה מידע – ישר והפוך.

מקיפים את השמש, ואיך הירח מחולל את הגאות והשפל – אבל *לא* את תפקיד הכסף בחברה האנושית או כיצד הלב מזרים דם. הבוז כלפי הצמצום מזכיר למדי את היוונים הקדמונים שסברו שלצאת החוצה וממש *להתבונן* בדברים זו עבודת כפיים, ועבודת כפיים נועדה לעבדים.

כפי שניסח זאת אפלטון, בספר ז׳ של *המדינה*:

> וגם אם יסתכל אדם בקישוטים שבתקרה, ויבחין במה שיבחין אגב הטיית ראשו כלפי מעלה, תהיה סבור, כפי הנראה, שהוא מעיין במחשבתו, ולא בעיניו. ואולי יפה סברתך, ושלי – תמימה. כי שוב אינני יכול להאמין שלימוד כל שהוא יגרום לנשמה שתביט אל על, חוץ מאותו שעניינו – מה שהווה־ואיננרנראה; ואם יביט אדם כלפי מעלה בפה פעור, או כלפי מטה בפה סגור, וינסה ללמוד דבר מה מוחש, לעולם לא אגיד עליו שהוא למד, באשר לא תחול ידיעה על שום ענין מעין זה, אף לא אגיד שנשמתו מביטה כלפי מעלה, כי אם כלפי מטה, ואפילו הוא שוחה על גבו, ביבשה או בים, ולומד פרקדן.*

רבים כיום עושים טעות דומה, וחושבים שמושגים מתוחמים הם נחותים, לא־נשגבים ובלתי־פילוסופיים בדיוק כמו, נניח, לצאת החוצה ולהתבונן בדברים – פעולה שראויה לפשוטי העם. אך רציונליסטים – וגם משוררים – זקוקים למילים קונקרטיות כדי להביע מחשבות מדויקות. הם צריכים קטגוריות שכוללות רק דברים מסוימים, ואינן כוללות אחרים. אין כל פסול במיקוד המחשבה, צמצום הקטגוריות, פסילת אפשרויות וחידוד הטענות. באמת שאין! אם תרחיבו את משמעות מילותיכם יותר מדי, תסיימו עם משהו שאינו נכון ואפילו לא מאפשר שירה טובה.

ואפילו אל תתחילו איתי *לגבי אנשים שחושבים ש״ויקיפדיה״ היא ״בינה מלאכותית״, המצאת ה־LSD היא ״סינגולריות״ או שתאגידים הם בעלי ״אינטליגנציית־על״*!

* תרגום לעברית: יוסף ג. ליבס.

96

כיצד להיראות (ולהיות) עמוקים

לאחרונה השתתפתי בקבוצת דיון שהנושא שלה, באותו מפגש, היה מוות. הנושא עורר רגשות עמוקים. אני סבור שמכל מפגשי הצהריים שבהם השתתפתי בעמק הסיליקון, מפגש זה היה הכֵּן ביותר. המשתתפים דיברו על מוות במשפחה, מוות של חברים, ועל מה שהם חושבים על המוות שלהם עצמם. הם באמת הקשיבו אחד לשני. הלוואי וידעתי כיצד לשחזר באופן אמין את התנאים לשיחה כזו.

אני הייתי הטרנס־הומניסט היחיד, ומאוד השתדלתי שלא להיות מעצבן. (״פנאט הוא זה שאינו מסוגל לשנות את דעתו, וגם לא מוכן לשנות את הנושא.״ אני שואף לפחות להיות מוכן לשנות את הנושא.) באופן לא מפתיע, המשתתפים שוחחו על המשמעות שנותן המוות לחיים, או על איך המוות הוא למעשה ברכה סמויה. אבל הסברתי, בזהירות רבה, שבאופן כללי טרנס־הומניסטים הם בעד החיים ונגד המוות.

לאחר מכן, ניגשו אליי כמה מהמשתתפים ואמרו לי שהייתי ״מאוד עמוק״. ובכן, זה נכון, אבל זה גם הוביל אותי לחשוב על מה גורם לאנשים *להישמע עמוקים*.

בשלב מסוים בשיחה, אישה אמרה שהמחשבה על מוות מביאה אותה להיות נחמדה יותר לאנשים, כי, מי יודע, אולי היא לא תראה אותם שוב. ״כשיש לי משהו טוב לומר על מישהו, עכשיו אני פשוט אומרת לו את זה, ולא מחכה.״

"זוהי מחשבה יפה," אמרתי, "ואפילו אם יום אחד איום המוות יוסר ממך, אני מקווה שתמשיכי לעשות זאת —"

לאחר מכן, אישה זו הייתה אחת מאלו שאמרו לי שאני עמוק.

בשלב אחר בשיחה, גבר דיבר על תועלת כלשהי של המוות, אינני זוכר איזו תועלת בדיוק. אמרתי: "אתה יודע, בהינתן טבע האדם, אילו אנשים היו חוטפים מכה בראש ממחבט בייסבול כל שבוע, עד מהרה הם היו ממציאים סיבות מדוע מכה בראש ממחבט בייסבול היא דבר טוב. אבל אם היית שואל מישהו שאינו סופג מכות בראשו מאלת בייסבול אם הוא מעוניין לקבל מכות, הוא היה אומר שלא. אני חושב שאם היית שואל בני אלמוות אם הם מוכנים למות בשביל התועלת הזו, הם היו אומרים שלא."

לאחר מכן, גבר זה אמר לי שאני עמוק.

מתאם אינו סיבתיות. אולי קולי היה עמוק באותו יום, ולכן נשמעתי חכם.

אך אני חושד שנתפסתי כ"עמוק" מכיוון שהפרתי באופן עקבי את תבנית־המטמון* עבור "חוכמה עמוקה" בדרך שנשמעה הגיונית באופן מיידי.

ישנו סטריאוטיפ של *חוכמה עמוקה*. *מוות*. השלימו את התבנית: "המוות נותן משמעות לחיים." כולם מכירים את התשובה הסטנדרטית *והחכמה* הזאת, וכך היא מקבלת תכונות מסוימות של נורת מחיאות כפיים.** אם תאמרו זאת, אנשים עשויים להנהן, כיוון שמוחם משלים את התבנית והם יודעים שהם אמורים להנהן. יתכן שאפילו יאמרו "איזו חוכמה עמוקה!" — אולי בתקווה להצטייר כעמוקים בעצמם. אבל הם לא יהיו *מופתעים*. הם לא שמעו שום דבר מחוץ לקופסה. הם לא שמעו משהו שלא יכלו לחשוב עליו בעצמם. אולי ניתן לקרוא לכך *אמונה בחוכמה* — המחשבה מתוייגת כ"חוכמה עמוקה", ומשלימה תבנית סטנדרטית של "חוכמה עמוקה", אך אינה מביאה עמה את חוויית התובנה.

---

* ראו פרק 90.

**ראו פרק 19.

אלו שמנסים *להיראות* בעלי *חוכמה עמוקה* מצטיירים בסופו של דבר כחלולים, מהדהדים כביכול, משום שהם מנסים להיראות בעלי *חוכמה עמוקה* במקום לבצע אופטימיזציה.

כמה חשיבה הייתי צריך לעשות, כדי להיראות עמוק? תדירותו של המוח האנושי היא 100 הרץ, ואני הגבתי בזמן אמת, אז רוב החשיבה שלי בהכרח נעשתה מראש. החלק שהיה כרוך במאמץ מבחינתי היה לבחור תגובה שמובנת בצעד היסקי אחד, ואז לנסח אותה כך שהשפעתה תהיה מרבית.

מבחינה פילוסופית, רוב העבודה שלי נעשתה מראש. השלימו את התבנית: מצב **א** מוצדק כיוון שהוא מביא לתועלת **ב**. "הכשל הנטורליסטי?" / "הטיית הסטטוס קוו?" / "האם נוכל להשיג את **ב** ללא **א**?" / "לו מעולם שמענו על **א** קודם לכן, האם היינו מאמצים אותו מרצון כדי להגיע ל-**ב**?" אני חושב שהוגן לומר שאלו תבניות חשיבה שמתקיימות אצלי באותה רמת אוטומציה כמו נשימה. אחרי הכל, רוב החשיבה האנושית מוכרחה להתבסס על שליפה ממטמון כדי שהמוח יוכל בכלל לתפקד.

וכבר החזקתי בתפיסה פילוסופית מפותחת של טרנס־הומניזם. גם לטרנס־הומניזם יש מחשבות־מטמון בנושא המוות. *מוות*. השלימו את התבנית: "מוות הוא טרגדיה חסרת טעם לה אנשים עושים רציונליזציה." זו הייתה מחשבת־מטמון לא־סטנדרטית, שאותה המאזינים שלי לא הכירו. היו לי מספר הזדמנויות להשתמש במחשבות־מטמון לא־סטנדרטיות, וכיוון שהן כולן היו חלק מהפילוסופיה המפותחת של הטרנס־הומניזם, כולן השתייכו לאותו המוטיב. זה גרם לי להיראות *עקבי*, כמו גם מקורי.

אני חושד שזו אחת הסיבות שפילוסופיה מהמזרח נראית עמוקה לאנשי המערב – יש לה מחשבות־מטמון לא־סטנדרטיות אך קוהרנטיות של *מחשבה עמוקה*. באופן סימטרי, ביצירות בדיוניות יפניות, לעיתים מופיעים נוצרים המוצגים כבעלי חוכמה עמוקה ו/או סודות מיסטיים.

אם זכרוני אינו מטעה, כלכלן אחד העיר פעם שהציבור הכללי

כה בּוּר בנושאי כלכלה, שכאשר הוא נקרא להופיע בטלוויזיה, כל שעליו לעשות כדי להישמע כמו הוגה מקורי ומבריק הוא לחזור על העקרונות מ-"מבוא לכלכלה".

נקודה חשובה נוספת היא שהמאזינים שלי יכלו לראות באופן *מיידי* שהתגובה שלי הגיונית. אולי הם לא הסכימו איתה, אבל היא לא הייתה קפיצה לוגית חסרת הקשר. אני מכיר טרנס־הומניסטים שלא מצליחים להיתפס כחכמים באותה צורה מכיוון שהם לא מצליחים להעריך מה המאזינים שלהם לא יודעים. אם אתם רוצים להישמע עמוקים, הימנעו מלומר משהו שדורש יותר מצעד־היסק אחד ממצבו המנטלי הנוכחי של המאזין שלכם. פשוט כך.

כדי *להיראות* עמוקים, לִמדו שיטות פילוסופיות לא־סטנדרטיות. חפשו דיונים בנושאים שיתנו לכם הזדמנויות להיראות עמוקים. בצעו את החשיבה הפילוסופית מראש, כך שבמהלך הדיון תוכלו להתרכז בהסברים טובים. ומעל הכל, תַּרגלו הישארות במרחק של צעד־היסק אחד בלבד מהקהל שלכם.

כדי *להיות* עמוקים, חשבו בעצמכם על נושאים "חכמים", חשובים, או קשים מבחינה רגשית. מחשבה עצמאית אין משמעה בהכרח מציאת תשובות בלתי רגילות, אך היא כן דורשת התבוננות עצמאית, במקום לתת למוחכם להשלים תבניות. אם לא תיעצרו בתשובה הראשונה, ותשליכו תשובות שמותירות תחושה מעורפלת של חוסר סיפוק, עם הזמן מחשבותיכם יתגבשו לכדי מבנה עקבי, הנובע מהמקור היחיד שהוא אתם עצמכם, במקום להוות חזרות מקוטעות על מסקנותיהם של אנשים אחרים.

97

אנו משנים את דעתנו פחות מכפי שנדמה לנו

במהלך השנים האחרונות, פנינו בדיסקרטיות לאנשים שניצבו בפני בחירה בין הצעות עבודה שונות, וביקשנו מהם להעריך את ההסתברות שיבחרו בעבודה אחת על פני האחרת. מידת הביטחון הממוצעת בבחירה הצפויה היתה 66% בלבד, אך רק 1 מתוך 24 המשיבים בחר באפשרות שאותה העריך תחילה בהסתברות נמוכה יותר, מה שהניב שיעור דיוק כולל של 96%.

— דֶיִיל גריפין ועמוס טְבֶרסקי[1]

כשקראתי לראשונה את המילים הללו — ב-1 באוגוסט 2003, בסביבות השעה שלוש אחר הצהריים — הדבר שינה את אופן החשיבה שלי. הבנתי *שברגע שאני יכול לנחש מה תהיה התשובה שלי* — ברגע שאני יכול לייחס הסתברות גבוהה יותר להחלטה אחת מאשר לאחרת — הרי שקרוב לוודאי שכבר החלטתי. אנו משנים את דעתנו פחות מכפי שנדמה לנו. וברוב המקרים, אנו מסוגלים לנחש מה תהיה התשובה שלנו בתוך חצי שנייה מרגע ששמענו את השאלה.

כמה מהר חולף לו אותו רגע בלתי מורגש, שבו עדיין איננו יכולים לנחש מה תהיה התשובה שלנו. אותו חלון הזדמנויות זעיר שבו האינטליגנציה מסוגלת לפעול. בשאלות בהן עלינו לבחור, כמו גם

בשאלות עובדתיות.

עיקרון השורה התחתונה הוא שרק הגורמים האמיתיים לאמונות שלכם קובעים את האפקטיביות שלכם כרציונליסטים. ברגע שהאמונה שלכם התקבעה, שום כמות של טיעונים לא תשנה את ערך־האמת שלה. ברגע שההחלטה שלכם התקבעה, שום כמות של טיעונים לא תשנה את ההשלכות.

אולי אתם סבורים שאתם מסוגלים להגיע לאמונה או להחלטה באמצעים לא־רציונליים, ואז לנסות להצדיק אותה, ואם תגלו שאינכם יכולים להצדיק אותה – פשוט תדחו אותה.

אבל אנו משנים את דעתנו פחות – *הרבה* פחות – מכפי שנדמה לנו.

אני בטוח שאתם יכולים לחשוב על לפחות מקרה אחד בחיים שלכם שבו שיניתם את דעתכם. כולנו יכולים. מה לגבי כל המקרים שבהם לא שיניתם את דעתכם? האם הם זמינים באותה המידה, כשאתם מעריכים באופן היוריסטי את היכולת שלכם?

בין הטיית החוכמה בדיעבד, סיבתיות מזויפת, הטיה חיובית, עיגון/הטרמה וכו׳ וכו׳, ומעל לכל – הטיית האישוש האימתנית – ברגע שרעיון נכנס לכם לראש, הוא כנראה הולך להישאר שם.

---

1. Dale Griffin and Amos Tversky, "The Weighing of Evidence and the Determinants of Confidence," *Cognitive Psychology* 24, no. 3 (1992): 411–435, doi:10.1016/0010-0285(92)90013-R.

98

המתן לפני שתציע פתרונות

מתוך ספרו של רובין דוס, *בחירה רציונלית בעולם של חוסר ודאות*.[1] ההדגשות שלי.

נורמן ר. פ. מאייר מציין שכאשר קבוצה ניצבת בפני בעיה, הנטייה הטבעית של חבריה היא להציע פתרונות אפשריים ברגע שהם מתחילים לדון בבעיה. כתוצאה מכך, הדיון מתמקד ביתרונות ובחסרונות של הפתרונות שהוצעו, אנשים נקשרים רגשית להצעות שלהם, ופתרונות טובים יותר כלל אינם מוצעים. מאייר ניסח כלל במטרה לשפר פתרון בעיות קבוצתי: "**אל תציעו פתרונות לפני שהבעיה נידונה באופן המעמיק ביותר האפשרי מבלי שהוצע אף פתרון.**" קל להראות שכלל זה עובד בהקשרים שבהם ישנם לבעיות פתרונות טובים המוגדרים באופן אובייקטיבי.

מאייר הגה את ניסוי "משחק התפקידים" הבא כדי להדגים את טענתו. שלושה עובדים בעלי יכולות שונות עובדים בפס ייצור. הם מתחלפים בין שלוש משימות הדורשות רמות יכולת שונות, משום שהעובד המוכשר ביותר — שהוא גם הדומיננטי ביותר — נחוש מאוד להימנע משעמום. למרות שהעובד הכי פחות מוכשר יודע שאין ביכולתו לבצע את המשימות הקשות באותה רמה כמו שני האחרים, הוא מסכים לסבב, בשל הדומיננטיות של עמיתו המוכשר. "מומחה התייעלות" קובע שאם העובד המוכשר ביותר יקבל את המשימה הקשה ביותר והעובד הפחות מוכשר יקבל את המשימה הקלה ביותר, אפשר יהיה להגדיל את התפוקה ב-20%, והמומחה

ממליץ לעובדים להפסיק את הסבב. שלושת העובדים ו... אדם רביעי שמונה לגלם את תפקיד מנהל העבודה, מתבקשים לדון בהמלצת המומחה. חלק מהקבוצות במשחק התפקידים מקבלות את הכלל של מאייר, שלא לדון בפתרונות עד שידונו בבעיה לעומק, בעוד שאחרות לא. אלו שלא קיבלו את הכלל מתחילות מיד להתווכח על החשיבות של תפוקה לעומת אוטונומיית העובד והימנעות משעמום. לקבוצות שקיבלו את הכלל יש סיכוי גבוה בהרבה להגיע לפתרון שבו שני העובדים המוכשרים יותר מתחלפים ביניהם, בעוד העובד הפחות מוכשר נשאר במשימה הפחות תובענית – פתרון שמניב עלייה של 19% בתפוקה.

השתמשתי בכלל הזה לעתים קרובות עם קבוצות שהנחיתי – **במיוחד כשהן ניצבו בפני בעיה קשה מאוד, שכן זהו המצב שבו לחברי הקבוצה יש את הנטיה החזקה ביותר להציע פתרונות באופן מיידי.** למרות שאין לי קריטריון אובייקטיבי שעל פיו ניתן לשפוט את איכות פתרון הבעיות של הקבוצות, נראה שהכלל של מאייר מסייע להגעה לפתרונות טובים יותר.

זה כל כך נכון שזה אפילו לא מצחיק. וזה הופך לגרוע יותר ויותר ככל שהבעיה הופכת קשה יותר. קחו לדוגמה בינה מלאכותית. נראה שמספר מפתיע של אנשים שאני פוגש יודעים בדיוק איך לבנות בינה מלאכותית כללית, מבלי לדעת, למשל, איך לבנות מזהה תווים אופטי או מערכת סינון שיתופי* (שהן בעיות קלות בהרבה). ובאשר לבניית בינה מלאכותית בעלת השפעה חיובית על העולם – "בינה מלאכותית ידידותית", בניסוח חופשי – ובכן, הבעיה *הזו* כל כך קשה עד *שמרבית* האנשים פותרים את הסוגיה כולה בתוך חמש־עשרה שניות. *תעשו לי טובה.*

הבעיה הזו אינה ייחודית כלל לבינה מלאכותית. פיזיקאים נתקלים בשלל אנשים שאינם פיזיקאים המחזיקים בתאוריות פיזיקליות משלהם, וכלכלנים זוכים לשמוע המון תאוריות כלכליות חדשות

* *מערכת סינון שיתופי* (collaborative filtering) היא שיטה להמלצה על תוכן או מוצרים על סמך ההעדפות של אנשים אחרים בעלי טעם דומה – למשל, המלצה על השיר הבא באפליקציית מוזיקה על סמך ההתנהגות של משתמשים דומים.

ומדהימות. אם אתם ביולוגים אבולוציוניים, כל אדם שתפגשו יוכל לפתור באופן מיידי כל בעיה פתוחה בתחומכם, בדרך כלל על ידי הנחת ברירה קבוצתית.* וכן הלאה.

העצה של מאייר מהדהדת את עיקרון השורה התחתונה: האפקטיביות של ההחלטות שלנו נקבעת אך ורק על ידי הראיות ועיבוד הנתונים שביצענו כשהגענו להחלטות שלנו לראשונה. אחרי שכתבתם את השורה התחתונה, מאוחר מדי לכתוב סיבות נוספות מעליה. אם תקבלו את ההחלטה שלכם בשלב מוקדם מאוד, היא תהיה מבוססת, למעשה, על מעט מאוד חשיבה, לא משנה כמה טיעונים נהדרים תמציאו לאחר מכן.

וחִשבו גם על כך שאנו משנים את דעתנו פחות מכפי שנדמה לנו: 24 אנשים ייחסו הסתברות ממוצעת של 66% לבחירה העתידית שנחשבה לסבירה יותר, אך רק 1 מתוך 24 בחר בפועל באפשרות שנחשבה לסבירה פחות. **ברגע שאתם יכולים לנחש מה תהיה התשובה שלכם, סביר להניח שכבר החלטתם.** אם אתם יכולים לנחש את התשובה שלכם חצי שנייה אחרי ששמעתם את השאלה, הרי שיש לכם חצי שנייה שבה אתם יכולים להיות אינטליגנטיים. זה לא הרבה זמן.

הרציונליות המסורתית מדגישה את *ההפרכה* – היכולת *לוותר* על דעה קיימת כאשר ניצבות בפניכם ראיות ברורות נגדה. אבל ברגע שרעיון נכנס לכם לראש, קרוב לוודאי שיידרשו הרבה יותר מדי ראיות כדי להוציא אותו משם. גרוע מכך, לא תמיד יש לנו את הפריווילגיה של ראיות מכריעות.

אני חושד ששיטה עוצמתית יותר (וקשה יותר) היא *להשהות את המחשבה על תשובה*. לעכב ולמשוך את אותו רגע זעיר שבו אנחנו עדיין לא מסוגלים לנחש מה תהיה התשובה שלנו, ובכך להעניק

---

* *ברירה קבוצתית* (group selection) היא הרעיון שהברירה הטבעית יכולה לפעול גם ברמת הקבוצה ולא רק ברמת הפרט. לפי גישה זו, קבוצות שיש בהן יותר שיתוף פעולה או התנהגות אלטרואיסטית עשויות לשרוד ולהתרבות טוב יותר מקבוצות אנוכיות.

לאינטליגנציה שלנו זמן ארוך יותר לפעול.
אפילו חצי דקה תהיה שיפור לעומת חצי שנייה.

---

1. Dawes, *Rational Choice in An Uncertain World*, 55–56.

99

הכשל הגנטי

ברשימות של כשלים לוגיים, תוכלו למצוא את "הכשל הגנטי" – הכשל של תקיפת אמונה על בסיס הסיבות שמביאות אדם להחזיק בה.

במבט ראשון, זהו רעיון משונה מאוד – אם הסיבות לאמונה לא קובעות את מידת המהימנות שלה, אז מה כן? אם מנוע השחמט *כחול עמוק* ממליץ לנו על מהלך, אנו סומכים עליו בהתבסס על הבנתנו את *הקוד* שסורק את עץ המשחק, מאחר שאיננו מסוגלים להעריך את עץ המשחק בעצמנו. מה יכול להצדיק ייחוס "רציונליות" להסתברות כלשהי, מלבד העובדה שהיא הופקה על ידי תהליך מהימן באופן שיטתי?

מאמרים על הכשל הגנטי יספרו לכם שהסקה 'גנטית' אינה תמיד כשל – מקור הראיות *עשוי* להיות רלוונטי להערכת ההשפעה שלהן על ההסתברות, כמו במקרה של מומחה מהימן. אך בפעמים אחרות, טוענים המאמרים, מדובר בכשל: הכימאי קֶקוּלֶה ראה לראשונה את מבנה הטבעת של מולקולת הבֶּנזֶן בחלום, אך אין זה אומר שלעולם לא נוכל לסמוך על אמונה זו.

אז לפעמים הכשל הגנטי הוא כשל, ולפעמים לא?

הכשל הגנטי הוא מבחינה פורמלית כשל, משום *שהסיבה המקורית* לאמונה אינה זהה *למידת הביסוס הנוכחית* שלה, שהיא סך כל התמיכה וההתנגדות הידועים *כיום*.

עם זאת, אנו משנים את דעתנו פחות מכפי שנדמה לנו. להאשמות גנטיות יש השפעה על בני אדם, השפעה שלא הייתה מתקיימת על בייסיאנים אידיאליים.

'איפוס' המחשבה שלכם היא *היוריסטיקה עוצמתית* כשאתם חושדים שרבים מהרעיונות שלכם הגיעו ממקור פגום. ברגע שרעיון חודר לראש שלנו, לא תמיד קל לעקור אותו באמצעות ראיות. חישבו על כל אותם אנשים שגדלו על אמונה בתנ"ך. מאוחר יותר הגיעו למצב שבו דחו (ברמה מודעת ומחושבת) את הרעיון שהתנ"ך נכתב בידי אלוהים, אך הם עדיין סבורים שהתנ"ך מכיל חוכמה מוסרית שאי אפשר לוותר עליה. הם נכשלו באיפוס המחשבה שלהם. מצבם היה טוב משמעותית אילו היו מטילים ספק בכל מה שהתנ"ך אמר, *פשוט משום שהתנ"ך אמר זאת.*

בו זמנית, יהיה עליהם לזכור היטב את העיקרון שלפיו טיפשות הפוכה איננה חוכמה. המטרה היא לשחרר את המחשבה שלכם באמת ולחשוב באופן עצמאי, לא לשלול את התנ"ך ולהפוך את השלילה הזו לאלגוריתם שלכם.

ברגע שנכנס לכם רעיון לראש, אתם נוטים למצוא לו תמיכה בכל כיוון אליו תביטו – ולכן, כאשר המקור הראשוני מוטל לפתע בספק, יהיה זה חכם מצידכם לחשוד בכל העלים שצמחו מתוך הענף הזה...

אם אתם מסוגלים ! לא קל לאפס את המחשבה. נדרש מאמץ כביר כדי לשקול דברים מחדש *באמת*, במקום לתת למוח ליפול לדפוס של חזרה על טיעוני־מטמון. "זה לא משבר אמונה אמיתי אלא אם הדברים יכלו להסתיים כך או אחרת באותו הקלות," אמר תור שֶׁנקֶל.

עליכם להיות *חשדנים במיוחד* אם יש לכם רעיונות רבים שנובעים מאותו המקור שאתם יודעים כעת שאינו אמין, אך, למרבה הפלא, נראה שכל הרעיונות עדיין עומדים בתוקפם – התנ"ך הוא הדוגמה הארכיטיפית לכך.

מצד שני... קיים מצב בו הראיות כה חזקות שזה כבר לא באמת משנה מהיכן הגיע הרעיון. צבירת ראיות מסוג זה היא מהות המדע. זה כבר לא משנה שקקולה ראה את מבנה הטבעת של הבנזן לראשונה

בחלום – זה גם לא היה משנה אם היינו מוצאים את ההשערה באמצעות ייצור תמונות אקראיות במחשב, או ממיסטיקן שהתגלה כנוכל, או אפילו מהתנ"ך. מבנה הטבעת של הבנזן מעוגן במספיק ראיות ניסוייות כדי להפוך את המקור ללא רלוונטי.

בהיעדר ראיות חזקות שכאלה, עליכם בהחלט לשים לב למקור ממנו הגיע הרעיון – לתת למומחים משקל גדול יותר מהדיוטות, אם התחום שלהם הוכח כראוי – לחשוד ברעיונות שקיבלתם ממקורות חשודים – לא לתת אמון באלו שמניעיהם אינם אמינים, *אם* הם אינם יכולים להציג טיעונים בלתי תלויים בסמכותם־שלהם.

הכשל הגנטי הוא *כשל* כאשר קיימות הצדקות *מעבר* למקור עליו מבוססת הטענה, אך הטענות בדבר מקור הטענה מוצגות כאילו הן מכריעות את הדיון. הַאל פִּיני מציע שנקרא להסתמכות נכונה על מקורותיה של טענה "ההיוריסטיקה הגנטית".

כמה כללי אצבע טובים (לבני אדם):

- הֱיו חשדנים כלפי 'האשמות גנטיות' נגד אמונות שאינכם מחבבים, במיוחד אם מי שתומך באותן האמונות מצדיק אותן באופן החורג מסמכותו של מקור האמונה. "תעופה היא רעיון דתי, ולכן האחים רייט בהכרח שקרנים" זו אחת הדוגמאות הקלאסיות הניתנות לכך.
- באותה מידה, אל תחשבו שאתם יכולים לקבל מידע אמין לגבי סוגיה טכנית רק באמצעות פסיכואנליזה מתחכמת של האישים המעורבים והמניעים הפגומים שלהם. אם קיימים טיעונים טכניים, הם עדיפים.
- כאשר מתעורר חשד חדש כלפי אחד מהמקורות הבסיסיים שלכם, *עליכם* בהחלט לפקפק בכל הענפים והעלים שצמחו מאותו השורש. אין הכרח לדחות אותם מיד על הסף, מכיוון שטיפשות הפוכה איננה חוכמה, אבל...
- היו חשדנים במיוחד אם תגלו שאתם עדיין מאמינים באותם הרעיונות הנובעים מהמקור אותו דחיתם.

חלק ו'

מעגלי מוות

התקדמות מבחינה רציונלית משמעה מפגש עם רעיונות רבים חדשים, מעניינים ועוצמתיים. במקרים רבים, משמעות הדבר היא גם רכישת חברים שעימם ניתן להחליף דעות, ומציאת קהילות המעודדות אתכם להשתפר. "מעגלי מוות" דן בכמה סכנות מרכזיות שעלולות לפקוד קבוצות המאוחדות סביב תחומי עניין משותפים ורעיונות מושכים ומדהימים – סכנות שיש להתגבר עליהן אם ברצוננו להפיק את מלוא התועלת מקהילות רציונליסטיות.

– רוב בנסינגר,
רציונליות: מבוא

100

היוריסטיקת הרגש

היוריסטיקת הרגש מתרחשת כאשר רשמים סובייקטיביים של טוב/רע מתפקדים כהיוריסטיקה – מקור לשיפוטים תפיסתיים מהירים. תחושות נעימות ובלתי נעימות תופסות מקום מרכזי בתהליכי ההיסק האנושיים, והיוריסטיקת הרגש מגיעה עם הטיות נהדרות – כמה מהאהובות עליי.

נתחיל עם אחת ההטיות הפחות מטורפות, יחסית. אתם מתעתדים לעבור לעיר חדשה, ועליכם לשלוח בהובלה שעון אורלוגין עתיק. במקרה הראשון, השעון היה מתנה שקיבלתם מסבא וסבתא ליום ההולדת החמישי שלכם. במקרה השני, השעון היה מתנה מקרוב משפחה רחוק ואין לכם רגשות מיוחדים כלפיו. כמה תהיו מוכנים לשלם עבור פוליסת ביטוח שתשלם לכם 100 דולר אם השעון יאבד במשלוח? לפי סי וקונרויטר, נבדקים הצהירו על נכונות לשלם יותר מכפול במקרה הראשון.[1] זה עשוי להישמע רציונלי – למה לא לשלם יותר כדי להגן על חפץ בעל ערך רב יותר? – עד שאתם מבינים שהביטוח לא באמת *מגן* על השעון, הוא רק משלם אם השעון אובד, ומשלם בדיוק את אותו הסכום עבור כל אחד מהשעונים. (וכן, צוין שהביטוח מתבצע על ידי חברה חיצונית, כך שהוא לא מספק מניע מיוחד למובילים.)

בסדר, אבל זה לא *נשמע* מטורף מדי. אולי תוכלו לתרץ זאת בטענה שהנבדקים ביטחו את עצמם מהשלכות רגשיות, לא השלכות

כלכליות – סוג של קניית נחמה.

אז מה דעתכם על זה? יאמאגישי הראה שנבדקים התייחסו למחלה כמסוכנת יותר כאשר נאמר להם שהיא גורמת למותם של 1,286 אנשים מתוך כל 10,000 הלוקים בה, לעומת מחלה בעלת סיכויי תמותה של 24.14%.[2] ככל הנראה, התמונה המנטלית של אלף גופות מעוררת הרבה יותר חרדה, בהשוואה לאדם בודד שרוב הסיכויים שישרוד.

אבל חכו, זה נהיה גרוע יותר.

נניח ששדה־תעופה צריך להחליט האם להוציא כסף על רכישת ציוד חדש, או להשקיע בהיבטים אחרים של בטיחות השדה – כפי שטוענים מתנגדי המהלך. סָלוֹבִיק ואחרים הציגו לשתי קבוצות של נבדקים את הטיעונים בעד ונגד רכישת הציוד, עם סולם תשובות שנע בין 0 (לא אתמוך כלל) ל-20 (אתמוך מאוד).[3] בפני קבוצה אחת הוצג הציוד ככזה שיציל את חייהם של 150 בני אדם. בפני הקבוצה השנייה הוא הוצג ככזה שיציל 98% מתוך 150 בני אדם. ההשערה שבבסיס הניסוי הייתה שהצלת 150 אנשים נשמעת כמו משהו טוב באופן מעורפל – האם זה הרבה? מעט? – בעוד שהצלת 98% ממשהו היא בבירור דבר טוב מאוד, משום ש-98% כל כך קרוב לקצה העליון של סולם האחוזים. הפלא ופלא, הצלת 150 בני אדם זכתה לתמיכה ממוצעת של 10.4, בעוד שהצלת 98% מתוך 150 בני אדם זכתה לתמיכה ממוצעת של 13.6.

או קחו לדוגמה את הניסוי של דֶנֶשׁ־ראז׳ ואפשטיין:[4] נבדקים שהובטח להם דולר אחד עבור כל פעם שישלפו באקראי סוכריית ג׳לי אדומה מתוך קערה, העדיפו לעיתים קרובות קערה שבה ישנן יותר סוכריות אדומות אך שיעורן היחסי נמוך יותר. לדוגמה, קערה שבה 7 סוכריות אדומות מתוך 100 הועדפה על פני קערה שבה 1 מתוך 10.

לפי דנש־ראז׳ ואפשטיין, נבדקים אלו הסבירו לאחר מכן שאף על פי שידעו שההסתברות פועלת לרעתם, הם הרגישו שיש להם סיכוי טוב יותר לזכות כשהיו יותר סוכריות אדומות בקערה. זה עשוי להישמע לכם מטורף, הו קוראים־בקיאים־ברזי־הסטטיסטיקה, אך אם

תקדישו לכך מחשבה נוספת תבינו שזה הגיוני לחלוטין. הסתברות של 7% לעומת 10% היא אולי נתון מאכזב, אך הכמות הגדולה יותר של סוכריות אדומות מפצה על כך ויותר. ההסתברות אמנם נמוכה יותר, כן, אבל עדיין יש סיכוי גבוה יותר *שתזכו*, אתם מבינים. עליכם להתייחד עם המחשבה הזו עד שתרד עליכם ההארה לגבי האופן שבו שאר תושבי כדור הארץ חושבים על הסתברות.

פִּינוּקְיין ואחרים מצאו שעבור כורים גרעיניים, גז טבעי וחומרים משמרים במזון, הצגת מידע על תועלת גבוהה גרמה לאנשים לתפוס את הסיכונים כנמוכים יותר; הצגת מידע על סיכונים גבוהים יותר גרמה לאנשים לתפוס את התועלת כנמוכה יותר; וכן הלאה על פני כל הרביעים.[5] אנשים כורכים יחד את השיפוט הנקודתי שלהם לגבי היבטים חיוביים או שליליים של דבר מה לכדי תחושה כללית (חיובית או שלילית) כלפי הדבר.

פינוקיין ואחרים מצאו גם שלחץ זמן *הגביר* משמעותית את הקשר ההפוך בין תפיסת הסיכון לתפיסת היתרון, ממצא שעולה בקנה אחד עם הממצא הכללי לפיו לחץ זמן, חוסר במידע או הסחת דעת – כולם מגבירים את הדומיננטיות של היוריסטיקות תפיסתיות על פני שיקול דעת אנליטי.

גַּנְזָך מצא את אותו האפקט בתחום הפיננסים.[6] על פי התיאוריה הכלכלית הרגילה, תשואה וסיכון אמורים להיות במתאם *חיובי*: אנשים משלמים מחיר גבוה יותר על השקעות בטוחות, מה שמוריד את התשואה, ומנגד מניות מניבות תשואות גבוהות יותר מאגרות חוב, אך כרוכות בסיכון גדול יותר בהתאם. כשאנליסטים בחנו מניות *מוכרות*, ההערכות שלהם לגבי סיכונים ותשואות היו במתאם חיובי, כמצופה. אך כשבחנו מניות *לא מוכרות*, האנליסטים נטו לשפוט אותן כאילו היו טובות או רעות באופן כללי – סיכון נמוך ותשואות גבוהות, או סיכון גבוה ותשואות נמוכות.

לקריאה נוספת, אני ממליץ על מאמרו המצוין של סלוביק המסכם את הנושא, "שחקנים רציונליים או שוטים רציונליים: ההשלכות של היוריסטיקת הרגש על כלכלה התנהגותית".[7]

1. Christopher K. Hsee and Howard C. Kunreuther, "The Affection Effect in Insurance Decisions," *Journal of Risk and Uncertainty* 20 (2 2000): 141–159, doi:10.1023/A:1007876907268.

2. Kimihiko Yamagishi, "When a 12.86% Mortality Is More Dangerous than 24.14%: Implications for Risk Communication," *Applied Cognitive Psychology* 11 (6 1997): 461–554.

3. Paul Slovic et al., "Rational Actors or Rational Fools: Implications of the Affect Heuristic for Behavioral Economics," *Journal of Socio-Economics* 31, no. 4 (2002): 329–342, doi:10.1016/S1053-5357(02)00174-9.

4. eronika Denes-Raj and Seymour Epstein, "Conflict between Intuitive and Rational Processing: When People Behave against Their Better Judgment," *Journal of Personality and Social Psychology* 66 (5 1994): 819–829, doi:10.1037/0022-3514.66.5.819.

5. Finucane et al., "The Affect Heuristic in Judgments of Risks and Benefits."

6. Yoav Ganzach, "Judging Risk and Return of Financial Assets," *Organizational Behavior and Human Decision Processes* 83, no. 2 (2000): 353–370, doi:10.1006/obhd.2000.2914.

7. Slovic et al., "Rational Actors or Rational Fools."

101

יכולת הערכה (וקניות זולות לחגים)

כעת, כשהחלק *היקר* של עונת החגים הולך ומתקרב, שאלה מרכזית בוודאי מנקרת במוחם של קוראינו:

"לכבוד כותבי *Overcoming Bias** היקרים, האם ישנן הטיות שאוכל לנצל כדי להיראות נדיב מבלי לבזבז *בפועל* הרבה כסף?"

אני שמח לבשר שהתשובה היא כן! על פי כריסטופר סְי – במאמר שכותרתו "פחות זה יותר טוב: מתי חלופות בעלות ערך נמוך מוערכות יותר מאשר חלופות בעלות ערך גבוה" – אם תקנו למישהו צעיף ב-45 דולר, סביר שתיראו נדיבים יותר מאשר אם תקנו לו מעיל ב-55 דולר.[1]

זהו מקרה פרטי של תופעה כללית יותר. בניסוי מוקדם יותר,[2] סי שאל נבדקים כמה יהיו מוכנים לשלם עבור מילון מוזיקה יד־שנייה:

- מילון **א**, משנת 1993, בעל 10,000 ערכים, במצב כמו־חדש.
- מילון **ב**, משנת 1993, בעל 20,000 ערכים, במצב כמו־חדש מלבד קרע בכריכה.

הטריק היה שחלק מהנבדקים ראו את שני המילונים זה לצד זה, בעוד שנבדקים אחרים ראו רק מילון *אחד*...

נבדקים שראו רק *אחת* מהאפשרויות הללו היו מוכנים לשלם

* הבלוג בו נכתב במקור פרק זה.

בממוצע 24 דולר עבור מילון **א** ו-20 דולר עבור מילון **ב**. נבדקים שראו את *שתי* האפשרויות, זו לצד זו, היו מוכנים לשלם 27 דולר עבור מילון **ב** ו-19 דולר עבור מילון **א**.

כמובן שמספר הערכים במילון חשוב יותר מאשר השאלה אם כריכתו קרועה, לפחות אם אתם מתכננים אי פעם להשתמש בו. אבל אם מוצג בפניכם רק מילון אחד, ויש בו 20,000 ערכים, המספר 20,000 לא אומר הרבה. האם זה מעט? הרבה? מי יודע? זהו נתון *שלא ניתן להעריך*. הכריכה הקרועה, לעומת זאת, בהחלט תופסת את העין. יש לה מטען רגשי שלילי מובהק.

אולם, כאשר הם מוצגים זה לצד זה, מספר הערכים הופך *לניתן להערכה*, מכיוון שישנן שתי כמויות בנות־השוואה. וברגע שמספר הערכים ניתן להערכה, ההיבט הזה מגמד את חשיבותה של הכריכה הקרועה.

מתוך סלוביק ואחרים:[3] מה הייתם מעדיפים?

1. סיכוי של 29/36 לזכות ב-2 דולר.
2. סיכוי של 7/36 לזכות ב-9 דולר.

בעוד *שהמחירים* הממוצעים (ערכי השקילות*) שניתנו לאפשרויות אלו עמדו על 1.25 דולר ו-2.11 דולר בהתאמה, דירוגי האטרקטיביות** הממוצעים שלהם היו 13.2 ו-7.5. לנבדקים נאמר כי הם יקבלו שני תורות משחק בהימור בעל המחיר הגבוה יותר או דירוג האטרקטיביות הגבוה יותר. (כך לנבדקים היה מניע לדרג הימורים כאטרקטיביים יותר, או לתמחר אותם גבוה יותר, אם הם באמת יעדיפו לשחק בהם.)

ההימור שהיה שווה יותר כסף נראה פחות אטרקטיבי, היפוך העדפות קלאסי.*** החוקרים שיערו שבמשימת קביעת

---

* *ערך שקילות* (equivalence value) הוא סכום הכסף שאדם תופס כשווה ערך להימור מסוים, כך שמבחינתו אין זה משנה אם יקבל את הסכום לידיו או ישתתף בהימור.

** *דירוג אטרקטיביות* (attractiveness rating) הוא מדד סובייקטיבי שבו הנבדק מתבקש לציין על גבי סולם ערכים (למשל מ-1 עד 20) עד כמה האפשרות נתפסת בעיניו כמושכת ורצויה.

*** *היפוך העדפות* (preference reversal) הוא תופעה שבה אנשים מביעים העדפה

המחיר הנבדקים הסתמכו על סכום הזכיה, בעוד שבמשימת דירוג האטרקטיביות הם נשענו על ההסתברות לזכיה. לכן (כך חשבו החוקרים), מדוע לא לנסות להפוך את סכום הזכיה לבולט יותר מבחינה רגשית – ליותר ניתן להערכה – ולאטרקטיבי יותר?

וכיצד הם עשו זאת? על ידי הוספה של הפסד קטן מאוד להימור. בהימור הישן היה סיכוי של 7/36 לזכות ב-9 דולר. בהימור החדש היה סיכוי של 7/36 לזכות ב-9 דולר וסיכוי של 29/36 להפסיד 5 סנט. בהימור הישן, אתם מעריכים את האטרקטיביות של 9 דולר כשלעצמם. בהימור החדש אתם מעריכים את האטרקטיביות של זכייה ב-9 דולר *לעומת* הפסד של 5 סנט.

"התוצאות," אמרו סלוביק ואחרים, "עלו על הציפיות שלנו." בניסוי חדש, ההימור הפשוט עם סיכוי של 7/36 לזכות ב-9 דולר קיבל דירוג אטרקטיביות ממוצע של 9.4, ואילו ההימור המורכב שכלל סיכוי של 29/36 להפסיד 5 סנט קיבל דירוג אטרקטיביות ממוצע של 14.9.

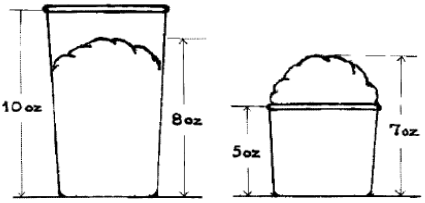
ניסוי המשך בדק האם הנבדקים יעדיפו את ההימור הישן על פני זכייה ודאית ב-2 דולר. רק 33% מהסטודנטים העדיפו את ההימור הישן. בקרב קבוצה אחרת שהתבקשה לבחור בין 2 דולר ודאיים לבין ההימור החדש (עם תוספת האפשרות להפסיד 5 סנט), לא פחות מ-60.8% העדיפו את ההימור. אחרי הכל, 9 דולר זה לא סכום כסף כל כך אטרקטיבי, אבל היחס בין 9 דולר ל-5 סנט הוא יחס זכייה/הפסד אטרקטיבי *להפליא*.

אפשר להפוך הימור לאטרקטיבי יותר על ידי הוספה של הפסד ממשי! נכון שפסיכולוגיה זה כיף? זו הסיבה שאיש מאלו שבאמת מעריכים את המורכבות המופלאה של האינטליגנציה האנושית, לא רוצה לעצב בינה מלאכותית דמוית־אדם.

כמובן, זה עובד רק אם הנבדקים לא רואים את שני ההימורים זה לצד זה.

שונה או הפוכה בין אותן אפשרויות בדיוק, אך ורק בעקבות שינוי בשיטת ההערכה או המדידה שלהן.

בדומה לכך, איזו משתי הגלידות הללו העדיפו לדעתכם הנבדקים במחקרו של סִי מ-1998?



From Hsee, © 1998 John Wiley & Sons, Ltd.

באופן טבעי, התשובה תלויה בשאלה האם הנבדקים ראו גלידה אחת בלבד, או את שתיהן זו לצד זו. נבדקים שראו גלידה אחת היו מוכנים לשלם 1.66 דולר למוכר H ו-2.26 דולר למוכר L. נבדקים שראו את שתי הגלידות היו מוכנים לשלם 1.85 דולר למוכר H ו-1.56 דולר למוכר L.

מה ההשלכות לגבי קניות החגים שלכם? שאם תשקיעו 400 דולר בקניית אייפוד טאצ' 16GB, מקבל המתנה יראה את נגן המוזיקה היקר ביותר. אם תשקיעו 400 דולר בקניית נינטנדו Wii, מקבל המתנה יראה את קונסולת המשחקים הזולה ביותר. איזה מוצר נותן תמורה טובה יותר לכסף? אה, אבל זו שאלה שקיימת רק אם רואים את שני המוצרים זה לצד זה. *אתם* תחשבו עליהם זה לצד זה בזמן שתעשו קניות, אבל המקבל יראה רק את מה שהוא יקבל.

אם יש לכם סכום כסף נתון להוציא על מתנה – והמטרה שלכם היא להפגין את הידידות שלכם, ולא באמת *לעזור* למקבל המתנה – עדיף לכם שלא לחפש את התמורה הטובה ביותר לכסף. החליטו כמה כסף אתם רוצים להוציא על מנת להרשים את המקבל, ואז חפשו את החפץ חסר הערך ביותר שזהו מחירו. ככל *שקטגוריית* החפצים זולה יותר, כך חפץ *מסוים* מתוכה ייראה יקר יותר, בהינתן שאתם מוציאים סכום נתון. מה נחרט יותר בזיכרון, חולצה ב-25 דולר או נר ב-25

דולר?

זה מעניק משמעות חדשה לגמרי למנהג היפני של קניית מלונים ב-50 דולר, הלא כן? אתם מסתכלים, מנידים בראשכם ואומרים "מה *הקטע* של היפנים?" ובכל זאת הם נתפסים כנדיבים להפליא, אולי אפילו כפזרנים, בעודם מוציאים 50 דולר בלבד. אתם עשויים להוציא 200 דולר על ארוחת ערב יוקרתית ולא להיראות עשירים כפי שיכולתם להיראות על ידי הוצאת 50 דולר על מלון. לו רק היה מנהג של הענקת קיסמי שיניים ב-25 דולר או גרגרי אבק ב-10 דולרים... היפנים היו יכולים לצאת אפילו יותר בזול.

נ.ב.: אם אתם באמת משתמשים בטריק הזה, אני רוצה לדעת מה קניתם.

---

1. Christopher K. Hsee, "Less Is Better: When Low-Value Options Are Valued More Highly than High-Value Options," *Behavioral Decision Making* 11 (2 1998): 107–121.
2. Christopher K. Hsee, "The Evaluability Hypothesis: An Explanation for Preference Reversals between Joint and Separate Evaluations of Alternatives," *Organizational Behavior and Human Decision Processes* 67 (3 1996): 247–257, doi:10.1006/obhd.1996.0077.
3. Slovic et al., "Rational Actors or Rational Fools."

102

סולמות אינסופיים, פיצויי עתק של חבר מושבעים ועתידנות

"פסיכופיזיקה", למרות שמה, היא התחום המכובד המקשר בין השפעות פיזיקליות להשפעות תחושתיות. אם אתם מכניסים אנרגיה אקוסטית לחלל האוויר – כלומר, מרעישים – *עד כמה חזק* נשמע הרעש לאדם, כפונקציה של האנרגיה האקוסטית? כמה אנרגיה אקוסטית אתם צריכים להוסיף לפני שהרעש יישמע חזק פי שניים לאוזן אנושית? התשובה איננה פי שניים – אלא קרוב לפי שמונה.

אנרגיה אקוסטית או פוטונים הם פשוטים יחסית למדידה. כאשר אתם רוצים לגלות עד כמה רועש גירוי אקוסטי או עד כמה מקור אור *מרגיש* בוהק, אתם בדרך כלל שואלים את המאזין או את הצופה. ניתן לעשות זאת באמצעות סולם תָחוּם וסופי, הנע בין "שקט מאוד" ועד "רועש מאוד", או מ"עמום מאוד" ועד "בוהק מאוד". אפשר גם להשתמש בסולם אינסופי, שנקודת האפס שלו היא "לא נשמע כלל" או "לא נראה כלל", אך הוא עולה משם ללא הגבלה. כאשר משתמשים בסולם אינסופי, בדרך כלל מוצג לצופה גירוי קבוע, *המודולוס*, שמקבל דירוג קבוע. לדוגמה, צליל שמקבל דירוג עוצמה של 10. לאחר מכן הצופה יכול לציין צליל שנשמע חזק פי שניים מהמודולוס באמצעות הדירוג 20.

טכניקה זו הוכחה כאמינה למדי. אבל מה קורה אם הנבדקים

מקבלים סולם אינסופי, אך ללא מודולוס? מאפס עד אינסוף, ללא נקודת ייחוס קבועה? הם ממציאים לעצמם מודולוס משלהם, כמובן. *היחס* בין הגירויים ימשיך להיות במתאם מהימן בין הנבדקים. נבדק אחד אומר שלצליל **א** יש עוצמה של 10 ולצליל **ב** יש עוצמה של 15. אם נבדק שני יאמר שלצליל **א** יש עוצמה של 100, ניתן לשער בסבירות גבוהה שהוא יעריך את עוצמתו של צליל **ב** באזור ה-150. אך אם אינכם יודעים במה משתמש נבדק שלישי כמודולוס שלו – מהי נקודת הייחוס שלו – הרי שאין שום דרך לנחש מה הוא יגיד על צליל **א**. זה יכול להיות 1, וזה יכול להיות 1,000.

עבור נבדק המדרג צליל *יחיד* על פני סולם *אינסופי* ללא נקודת ייחוס מוגדרת, כמעט כל השונות נובעת מהבחירה השרירותית של המודולוס, ולא מהצליל עצמו.

"הממ," אתם חושבים לעצמכם, "זה נשמע דומה להפליא לחֶבר מושבעים הדן בפיצויים עונשיים. לא פלא שהשונות שם גדולה כל כך!" אנלוגיה מעניינת, אבל האם תוכלו להוכיח אותה בניסוי?

כהנמן ואחרים הציגו בפני 867 נבדקים הכשירים לכהן כמושבעים תיאורי מקרים משפטיים (למשל, ילד שבגדיו עלו באש) וביקשו מהם את אחת מהאפשרויות הבאות:

1. לדרג את מידת השערורייתיות של פעולות הנתבע, על פני סולם סופי,
2. לדרג עד כמה ראוי להעניש את הנתבע, על פני סולם סופי, או
3. לקבוע את סכום הפיצויים העונשיים.

והנה הפלא ופלא, בעוד שהיה מתאם טוב מאוד בין דירוגי השערורייתיות והענישה של הנבדקים, סכומי הפיצויים העונשיים היו מפוזרים לחלוטין. עם זאת, *הדירוג היחסי* של הפיצויים העונשיים – סידורם מהפיצוי הנמוך ביותר ועד לגבוה ביותר – הראה מתאם טוב בין הנבדקים.

אם הייתם שואלים כמה מתוך השונות בסולם ה"ענישה" ניתן להסביר על ידי התרחיש הספציפי – המקרה המשפטי המסוים, כפי שהוצג לנבדקים – התשובה הייתה, אפילו עבור הנתונים הגולמיים,

0.49.* עבור *הדירוג היחסי* של סכום הפיצויים, כמות השונות המוסברת עמדה על 0.51. עבור *הסכומים הכספיים הגולמיים*, השונות המוסברת הייתה 0.06 !

כלומר: אם אתם יודעים מהו התרחיש שהוצג – אותו ילד שבגדיו עלו באש – אתם יכולים לנחש ניחוש מושכל לגבי דירוג הענישה, וניחוש מושכל לגבי *הדירוג היחסי* של הפיצוי הכספי ביחס לתיקים אחרים, אבל הסכום הכספי עצמו הינו בלתי ניתן לחיזוי לחלוטין.

גם חישוב החציון של שתים־עשרה תשובות שנבחרו באקראי לא עזר הרבה.**

כך שפסיקת הפיצויים העונשיים על ידי חבר מושבעים אינה הערכה כלכלית של ממש, אלא יותר הבעת עמדה – מדד פסיכופיזי של תרעומת, המובע על פני סולם אינסופי ללא מודולוס קבוע.

אני מוצא כי באופן דומה, גם *תחזיות עתידניות* רבות ניתנות להבנה בצורה הטובה ביותר כהבעת עמדה. קחו למשל את השאלה, "כמה זמן יעבור עד שתהיה לנו בינה מלאכותית ברמה אנושית?" התשובות שראיתי לשאלה זו מפוזרות לכל עבר. באחד המקרים הזכורים לי, אדם בולט מתחום הבינה המלאכותית הערין בפניי: "חמש מאות שנים." (!!!)

ובכן, הסיבה לכך שכל כך קשה לחזות מתי נגיע לבינה מלאכותית היא דיון ארוך בפני עצמו. אבל זה לא שהאדם שאמר "חמש מאות שנים" בחן את העתיד כדי להגיע לתשובה. ואין סיכוי שהוא הגיע למספר הזה באמצעות השיטה המפוקפקת והמוכרת של שימוש בחוק מור.*** אז מהי *המשמעות* של המספר 500?

---

* *שיעור השונות המוסברת* (explained variance) הוא החלק מתוך השונות (ההבדלים בדירוגים שהנבדקים השונים העניקו) שנובע מעובדות המקרה עצמו, ולא מהבדלים אישיים בין הנבדקים או מתנודות אקראיות. במדעי החברה ובפסיכולוגיה, 0.49 – דהיינו 49% מההבדלים – נחשב לנתון גבוה מאוד עבור משתנה בודד.

** בארצות הברית, חבר מושבעים מונה לרוב 12 איש. החוקרים ניסו לקחת 12 נבדקים שונים ולחשב את החציון של סכומי הפיצויים שפסקו, מתוך מחשבה שה'חוכמה הקולקטיבית' או ניטרול הקצוות החריגים ייצרו סכומים עקביים יותר בין קבוצות.

*** *חוק מור* (Moore's law) הוא התצפית ההיסטורית שלפיה מספר הטרנזיסטורים על שבב (ומכאן כוח המחשוב הזמין במחיר נתון) מכפיל את עצמו בערך כל

ככל שאני יכול לנחש, באותה מידה יכולתי לשאול: "על סולם שבו אפס משמעותו 'לא קשה כלל', עד כמה בעיית הבינה המלאכותית *מרגישה* לך קשה?" אילו היה מדובר בסולם סופי, כל משיב שפוי היה מסמן "קשה בצורה קיצונית" בקצה הסולם. כל דבר *מרגיש* קשה בצורה קיצונית אם אינך יודע איך לעשות אותו. אך במקום זאת, הסולם הוא אינסופי וללא מודולוס קבוע. לכן אנשים פשוט בוחרים מספר שייצג "קשה במיוחד", שעשוי להיות 50, 100, או אפילו 500. ואז הם מוסיפים את המילה "שנים" בסוף, וזוהי התחזית שלהם.

"עד כמה בעיית הבינה המלאכותית מרגישה לך קשה?" אינה השאלה החלופית היחידה. אחרים עונים כאילו שאלתי "עד כמה יש לך תחושה חיובית כלפי בינה מלאכותית?" אלא שבמקרה זה, מספרים נמוכים יותר משמעותם תחושות חיוביות יותר, וגם הם מוסיפים את המילה "שנים" בסוף. אם "הערכות הזמן" הללו מייצגות משהו מעבר להבעת עמדה על פני סולם אינסופי ללא מודולוס, הרי שלא הצלחתי להבחין בכך.

---

1. Daniel Kahneman, David A. Schkade, and Cass R. Sunstein, "Shared Outrage and Erratic Awards: The Psychology of Punitive Damages," Journal of Risk and Uncertainty 16 (1 1998): 48–86, doi:10.1023/A:1007710408413; Daniel Kahneman, Ilana Ritov, and David Schkade, "Economic Preferences or Attitude Expressions?: An Analysis of Dollar Responses to Public Issues," *Journal of Risk and Uncertainty* 19, nos. 1–3 (1999): 203–235, doi:10.1023/A:1007835629236.

שנתיים. למרות שמדובר במגמה אמפירית חזקה מאוד בעשרות השנים האחרונות, אין כל הבטחה שהמגמה לא תשתנה בעתיד. בנוסף, בינה מלאכותית היא בעיקרה בעיית תוכנה ולא רק חומרה, ולכן קצב הגידול בכוח המחשוב לבדו אינו מספיק כדי לקבוע מתי תיפתר הבעיה.

103

אפקט ההילה

היוריסטיקת הרגש מתארת כיצד תחושה כללית – חיובית או שלילית – משפיעה על שיפוטים רבים אחרים, בין אם זה הגיוני ובין אם לאו, בין אם אתם מודעים לכך ובין אם לאו. נבדקים שהוסברו להם היתרונות של אנרגיה גרעינית נוטים לדרג אותה כבעלת פחות סיכונים. אנליסטים המדרגים מניות שאינן מוכרות להם, מעריכים אותן לרוב כטובות או רעות באופן כללי – סיכון נמוך ותשואה גבוהה, או סיכון גבוה ותשואה נמוכה – בניגוד לתיאוריה הכלכלית המקובלת, לפיה המתאם בין סיכון לתשואה צריך להיות חיובי.

אפקט ההילה הוא ביטויה של יוריסטיקת הרגש בפסיכולוגיה החברתית. רוברט צ׳יאַלדיני מסכם זאת בספרו ״השפעה: מדע ומעשה״:[1]

> מחקרים הראו שאנו מייחסים לאנשים נאים תכונות חיוביות כגון כישרון, אדיבות, כנות וחכמה באופן אוטומטי (לסקירה של הראיות, ראו איגלי, אֶשמוֹר, מַקהיֶנ׳אני ולוֹנגוֹ, 1991).[2] יתרה מכך, אנו מגבשים את דעתנו עליהם מבלי להיות מודעים לכך שמשיכה פיזית ממלאת תפקיד בתהליך. חלק מההשלכות של ההנחה הלא־מודעת הזאת כי ״יפה משמעו טוב״ מפחידות אותי. לדוגמה, מחקר על הבחירות הפדרליות בקנדה ב-1974 מצא שמועמדים מושכים פיזית קיבלו יותר מפי שניים וחצי קולות ממועמדים שאינם כאלה (אֶפרֶן ופֶטֶרסוֹן, 1976).[3] על אף הראיות הללו להעדפת פוליטיקאים

נאים, מחקרי המשך הוכיחו שהמצביעים אינם מודעים להטיה שלהם. למעשה, 73 אחוז מהמצביעים הקנדים שהשתתפו בסקר הכחישו בכל תוקף שהצבעתם הושפעה מהמראה הפיזי. רק 14 אחוז מהם היו מוכנים להכיר באפשרות של השפעה כזו (אפרן ופטרסון, 1976).[4] מצביעים יכולים להכחיש ככל שירצו את השפעת המראה הפיזי על סיכויי הבחירה, אך הראיות ממשיכות לאשר את נוכחותה המטרידה (בּוּדֶסחַיים ודֶה־פָּאוֹלָה, 1994).[5]

אפקט דומה נמצא גם בתהליכי גיוס עובדים. באחד המחקרים, הופעה מטופחת של מועמדים בראיון עבודה מדומה השפיעה על ההחלטה לקבלם לעבודה יותר מאשר כישוריהם המקצועיים – וזאת למרות שהמראיינים טענו שלמראה החיצוני היה תפקיד שולי בלבד בהחלטותיהם (מָאק וריני, 1990).[6] היתרון שבו זוכים עובדים נאים אינו מוגבל רק לראיון הקבלה: כלכלנים שבחנו מדגמים מארצות הברית ומקנדה מצאו שאנשים נאים משתכרים בממוצע 14-12 אחוז יותר מעמיתיהם שאינם כאלה (הָמֶרמֶש ובִידְל, 1994).[7]

מחקרים מטרידים לא פחות מצביעים על כך שגם המערכת המשפטית שלנו חשופה באופן דומה להשפעות של מידות הגוף ומבנה העצמות. אנו רואים כי אנשים נאים צפויים לקבל יחס אוהד במיוחד במערכת המשפט (לסקירה, ראו קַסטֶלוֹ, וֶנְש ומוּר, 1991, ו־דָאונס ולָיְיוֹנס, 1990).[8] לדוגמה, במחקר שנערך בפנסילבניה (סטיוארט, 1980),[9] חוקרים דירגו את המראה הפיזי של 74 נאשמים גברים בפתח המשפטים הפליליים שלהם. מאוחר יותר, כשבדקו את גזרי הדין ברישומי בית המשפט, התברר שהגברים הנאים יותר קיבלו עונשים קלים בהרבה. למעשה, לנאשמים נאים היה סיכוי כפול להימנע ממאסר בהשוואה לנאשמים שאינם כאלה. במחקר אחר, בנושא פיצויים שנפסקו במשפט מדומה על רשלנות, נאשם שנראה טוב יותר מקורבנו חויב בפיצוי ממוצע של 5,623 דולר – אך כאשר הקורבן היה הנאה מבין השניים, הפיצוי הממוצע עמד על 10,051 דולר. יתרה מזאת, ההעדפה המבוססת על מראה חיצוני ניכרה הן בקרב המושבעים והן בקרב המושבעות (קוּלקָה וקֶסלֶר, 1978).[10]

ניסויים אחרים הוכיחו שלאנשים נאים יש סיכוי גבוה יותר לקבל סיוע בעת מצוקה (בֶּנסוֹן, קָרָאבֶּניק ולֶרנֶר, 1976)[11] והם מוצלחים יותר בשינוי דעתו של קהל באמצעות שכנוע (צַ׳ייקֶן, 1979)...[12]

ההשפעה של מראה פיזי על ההערכה של חכמה, כנות או אדיבות היא דוגמה מובהקת להטיה – במיוחד כאשר מעריכים את התכונות האחרות על סמך כתיבה בלבד – משום שאין אף סיבה עניינית לצַפות שמידת הכנות והמראה החיצוני יהיו כרוכים זה בזה. מצד שני, באיזו מידה אני נתפס בעיני אחרים כחכם בגלל הכנות שלי? באיזו מידה אני נתפס בעיני אחרים כאדם כֵּן בגלל האינטליגנציה שלי? בירור האמת ואמירת האמת, מטבען, אינן כה רחוקות זו מזו כמו להיראות יפה ולהיראות חכם...

אבל מחקרים אלה על אפקט ההילה של מראה חיצוני צריכים לעורר בנו חשד שייתכן שקיים אפקט הילה דומה גם לגבי אדיבות או חכמה. נניח שאתם מכירים מישהו שלא רק נראה לכם חכם מאוד, אלא גם כֵּן, אלטרואיסט, טוב־לב ושליו. עליכם לחשוד שחלק מהתכונות שאתם מייחסים לו משפיעות על תפיסתכם את האחרות. אולי האדם אכן חכם, כֵּן ואלטרואיסט, אך לא כל כך טוב־לב או שליו. עליכם לחשוד אם האנשים שאתם מכירים מתחלקים בצורה מושלמת מדי לצדיקים ולרשעים.

וגם – אני יודע שאתם לא חושבים שאתם צריכים לעשות זאת, אבל אולי אתם אכן צריכים – היו טיפ־טיפה יותר ספקנים כלפי המועמדים הפוליטיים הנאים יותר.

---

1. Robert B. Cialdini, *Influence: Science and Practice* (Boston: Allyn & Bacon, 2001).

2. Alice H. Eagly et al., "What Is Beautiful Is Good, But... A Meta-analytic Review of Research on the Physical Attractiveness Stereotype," *Psychological Bulletin* 110 (1 1991): 109–128, doi:10.1037/0033-

2909.110.1.109.

3. M. G. Efran and E. W. J. Patterson, "The Politics of Appearance" (Unpublished PhD thesis, 1976).

4. Ibid.

5. Thomas Lee Budesheim and Stephen DePaola, "Beauty or the Beast?: The Effects of Appearance, Personality, and Issue Information on Evaluations of Political Candidates," *Personality and Social Psychology Bulletin* 20 (4 1994): 339–348, doi:10.1177/0146167294204001.

6. Denise Mack and David Rainey, "Female Applicants' Grooming and Personnel Selection," *Journal of Social Behavior and Personality* 5 (5 1990): 399–407.

7. Daniel S. Hamermesh and Jeff E. Biddle, "Beauty and the Labor Market," *The American Economic Review* 84 (5 1994): 1174–1194.

8. Wilbur A. Castellow, Karl L. Wuensch, and Charles H. Moore, "Effects of Physical Attractiveness of the Plaintiff and Defendant in Sexual Harassment Judgments," *Journal of Social Behavior and Personality* 5 (6 1990): 547–562; A. Chris Downs and Phillip M. Lyons, "Natural Observations of the Links Between Attractiveness and Initial Legal Judgments," *Personality and Social Psychology Bulletin* 17 (5 1991): 541–547, doi:10.1177/0146167291175009.

9. John E. Stewart, "Defendants' Attractiveness as a Factor in the Outcome of Trials: An Observational Study," *Journal of Applied Social Psychology* 10 (4 1980): 348–361, doi:10.1111/j.1559-1816.1980.tb00715.x.

10. Richard A. Kulka and Joan B. Kessler, "Is Justice Really Blind?: The Effect of Litigant Physical Attractiveness on Judicial Judgment," *Journal of Applied Social Psychology* 8 (4 1978): 366–381, doi:10.1111/j.1559-1816.1978.tb00790.x.

11. Peter L. Benson, Stuart A. Karabenick, and Richard M. Lerner, "Pretty Pleases: The Effects of Physical Attractiveness, Race, and Sex on Receiving Help," *Journal of Experimental Social Psychology* 12 (5 1976): 409–415, doi:10.1016/0022-1031(76)90073-1.

12. Shelly Chaiken, "Communicator Physical Attractiveness and Persuasion," *Journal of Personality and Social Psychology* 37 (8 1979): 1387–1397, doi:10.1037/0022-3514.37.8.1387.

104

הטיית גיבורי-העל

תארו לעצמכם סוציופת חמוש מכף רגל ועד ראש המחזיק בבני ערובה, שזה עתה דחה את כל הבקשות למשא ומתן והכריז על כוונתו להתחיל להרוג. במציאות, החבר׳ה הטובים לרוב אינם בועטים בדלת ופורצים פנימה כשלחבר׳ה הרעים יש בני ערובה. אבל לפעמים – לעיתים נדירות *מאוד*, אבל לפעמים – החיים מחקים את הוליווד עד כדי כך שהטובים באמת נאלצים לפרוץ פנימה.

דמיינו שתי מציאויות נפרדות לחלוטין, שבהן שני גיבורים מסתערים אל תוך החדר והם הראשונים להתעמת עם הנבל.

במציאות אחת, הגיבור חזק דיו כדי להשליך מכוניות, מסוגל לירות מטחי אנרגיה מנחיריו, ניחן בשמיעת רנטגן, ועורו אינו סתם *חסין* לקליעים אלא משמיד אותם במגע. הנבל התבצר בבית ספר יסודי ומחזיק ביותר ממאתיים ילדים כבני ערובה. הוריהם ממתינים בחוץ וממררים בבכי.

במציאות שניה, הגיבור הוא שוטר במשטרת ניו יורק, ובני הערובה הן שלוש זונות שהנבל אסף מהרחוב.

שקלו את השאלה הבאה בזהירות רבה: מי מהם הוא הגיבור הגדול יותר? ולמי מהם יש סיכוי גבוה יותר לככב בחוברת קומיקס משלו?

אפקט ההילה קובע כי קיים מתאם בין התכונות החיוביות שאנו מייחסים לדברים. אנשים שמדורגים גבוה במדדי מראה חיצוני מדורגים גבוה גם במדדי כישרון, טוב לב, כנות וחוכמה.

ולכן, דמויות קומיקס שמצטיירות כחזקות ובלתי־פגיעות – שתי תכונות חיוביות – מצטיירות גם כמי שניחנו במידה רבה יותר של תכונות גבורה כגון אומץ והקרבה. ובכל זאת:

עד כמה קשה לשחק את הגיבור האמיץ כשאתה חסין לכל פגע?

– אדם וורן, *Empowered*, כרך 1.[1]

אינני זוכר אם קראתי את הנקודה הבאה במקום כלשהו, או שניסחתי אותה בעצמי: נראה כי *פרסום*, בפרט, מוסיף נדבך נלווה לכל שאר תכונות האישיות שאנו תופסים. קחו לדוגמה את גנדי. האם גנדי היה האדם *האלטרואיסטי ביותר* במאה העשרים, או פשוט האלטרואיסט *המפורסם ביותר*? גנדי צעד אל מול שוטרים חמושים באלות לפיזור הפגנות ומול חיילים חמושים ברובים. אבל גנדי היה אדם מפורסם, והוא היה מוגן מתוקף פרסומו. מה לגבי האחרים שצעדו איתו – האנשים שעמדו מול אלות ורובים אף שלא היו זוכים לכותרות בינלאומיות גם אילו הוכו עד אשפוז או נורו למוות?

מה חשב גנדי על הזכייה בכותרות, בפרסום, בתהילה, במקומו בהיסטוריה, על *הפיכתו לארכיטיפ* של התנגדות לא־אלימה, בשעה שלקח על עצמו סיכון פחות מכל אחד מהאנשים שצעדו לצידו? איך הוא הרגיש כשאחד מאותם גיבורים אלמוניים ניגש אליו, בעיניים נוצצות, ואמר לו עד כמה הוא נפלא? האם גנדי אי פעם ראה את הדברים באור הזה? אינני יודע. אני אינני גנדי.

אין בכך משום ביקורת על גנדי. תכליתה של התנגדות לא־אלימה אינה הפגנת אומץ לב. את זה קל הרבה יותר להשיג באמצעות צניחה ממפלי הניאגרה בתוך חבית. גנדי לא יכול היה למנוע מפרסומו להגן עליו לפחות באופן חלקי. פעולותיו אכן דרשו אומץ – אולי לא כפי שנדרש כדי לצעוד באלמוניות, אך עדיין אומץ רב.

ההטיה שאני מבקש להצביע עליה היא שנדמה כי "מדד הפרסום" של גנדי *מתווסף* באופן תפיסתי ל"מדד האלטרואיזם" שצבר בצדק. כשאתם חושבים על אי־אלימות, אתם חושבים על גנדי – ולא על

מפגינה אלמונית באחת הצעדות שלו, שעמדה מול אלות ורובים, הוכתה, פונתה לבית חולים, נותרה צולעת לשארית חייה, *ואיש מעולם לא זכר את שמה.*

באופן דומה, איזה מעשה *גדול* יותר – סיכון חייכם כדי להציל מאתיים ילדים, או סיכון חייכם כדי להציל שלושה מבוגרים?

התשובה תלויה בכוונתכם במילה "גדול". אם אי פעם תצטרכו *לבחור* בין הצלת מאתיים ילדים להצלת שלושה מבוגרים, בחרו באפשרות הראשונה. "כל המציל נפש אחת כאילו הציל עולם מלא" היא אולי אמרה נהדרת בתור נורת מחיאות כפיים, אבל זו עצה מוסרית איומה אם אתם חייבים לבחור בין שתי האפשרויות. לכן, אם ב"גדול יותר" אתם מתכוונים ל"מה חשוב יותר?" או "מהי התוצאה המועדפת?" או "מה עליי לבחור אם ביכולתי לעשות רק אחד מבין השניים?" – הרי שיהיה זה מעשה גדול יותר להציל מאתיים אנשים מאשר שלושה.

אך אם אתם שואלים על "גדוּלה" במובן של הפגנת סגולה מוסרית, הרי שמי שיסכן את חייו כדי להציל שלושה אנשים בלבד מפגין אומץ רב יותר ממי שמוכן לסכן את חייו כדי להציל מאתיים אנשים אך לא כדי להציל שלושה.

אין הכוונה שאתם יכולים לבחור במכוון לסכן את חייכם כדי להציל שלושה מבוגרים ולתת למאתיים תלמידי בית־ספר למות, רק כדי להפגין סגולה מוסרית גדולה יותר. מי שמסכן את חייו רק כי *הוא רוצה להיחשב לבעל מידות טובות*, מפגין סגולה מוסרית פחותה בהרבה ממי שמסכן את חייו *משום שהוא רוצה להציל אחרים*. מי שבוחר להציל שלושה אנשים ולא מאתיים מפני שהוא סבור שהדבר מצביע על סגולה מוסרית גדולה יותר, מסנוור מה"גדוּלה" של עצמו באנוכיות רבה כל כך עד שהוא מבצע מעשה השקול מוסרית להריגה.

זהו אחד מאותם תרחישי *וּ־וֵוי*:* אינכם יכולים להפגין סגולה

* *וּ־וֵוי* (wu-wei) הוא מונח טאואיסטי ל'אי־עשייה' – פעולה מתוך זרימה טבעית וללא מאמץ מכוון או כפייה. באופן פרדוקסלי, לא ניתן להגיע לוו־ווי מתוך ניסיון ישיר לעשות זאת.

מוסרית בעצם הניסיון להפגין אותה. אם ניצבת בפניכם הבחירה בין שיטה בטוחה להציל את העולם שאינה כרוכה כלל בהקרבה או חוסר נוחות אישית, לבין שיטה המסכנת את חייכם ודורשת מכם לסבול ייסורים ומחסור – לא תוכלו להפוך לגיבורים על ידי בחירה מכוונת בדרך השנייה. אין שום דבר הרואי ברצון להיראות גיבור. זה יהיה פספוס מוחלט של המטרה.

אנשים בעלי סגולה מוסרית אמיתית, שמנסים בכנות להציל חיים ולא מנסים להוכיח את מוסריותם, יחפשו תמיד דרכים להציל חיים רבים יותר בפחות מאמץ – ומשמעות הדבר היא שסגולתם המוסרית תבוא פחות לידי ביטוי. זה עשוי להיות מבלבל, אך אין בכך סתירה.

עם זאת, איננו תמיד יכולים לבחור להיות חסינים לקליעים. לאחר שעשינו את המיטב כדי להפחית את הסיכון ולהגדיל את מספר הניצולים, כל גבורה *שנותרת* אכן מתגלית במלוא עוצמתה.

השוטר ששם את חייו על הכף ללא כוחות־על, ללא ראיית רנטגן, ללא כוח פיזי על־אנושי, ללא יכולת תעופה, ומעל לכל ללא חסינות לקליעים, מפגין סגולה מוסרית גדולה לאין שיעור מזו של סופרמן – שהוא, בסופו של דבר, *בסך הכל גיבור־על*.

---

1. Adam Warren, *Empowered*, vol. 1 (Dark Horse Books, 2007).

105

בסך הכל משיח

לפני שני פרקים הראיתי כיצד אפקט ההילה – הגורם לנו לתפוס תכונות חיוביות כבעלות מתאם זו לזו (למשל, אנשים נאים נתפסים גם כטובי לב, כנים וחכמים יותר) – גורם לנו להעריץ גיבורים במידה רבה יותר אם הם חזקים להפליא וחסינים לקליעים. זאת על אף שמבחינה לוגית נדרש אומץ רב הרבה יותר כדי להיות גיבור *כשאינך* חסין לקליעים. יתרה מזאת, פעולה אמיצה להצלת חייו של אדם אחד מפגינה סגולה מוסרית גדולה יותר מפעולה להצלת העולם כולו (אף שאילו נדרשתם לבחור בין השניים, כמובן שעליכם להציל את העולם).

אך הבה נבחן מקרה קונקרטי.

ג׳ון פֶּרי היה קצין במשטרת ניו יורק שבמקרה היה גם אֶקסטרופיאני* וטרנס־הומניסט, וכך התוודעתי לשמו. פרי עמד לפרוש בתוך זמן קצר מתפקידו ולהקים משרד עורכי דין עצמאי, כשהגיעה הידיעה על מטוס שהתרסק לתוך מרכז הסחר העולמי. הוא נהרג בקריסת המגדל הצפוני. לא הכרתי את ג׳ון פרי אישית, ולכן איני יכול להעיד על אמונותיו ממקור ראשון. ואולם מעטים מאוד מקרב האקסטרופיאנים מאמינים באלוהים, ואני מניח שפרי גם הוא

* *אֶקסטרופיאניזם* (extropianism) הוא תת־זרם בתנועה הטרנס־הומניסטית, המדגיש שימוש פעיל במדע, בטכנולוגיה ובחשיבה רציונלית כדי להתגבר על המגבלות הביולוגיות של האדם – ובכלל זה הזדקנות ומוות.

היה אתאיסט.

משמעות הדבר היא שפרי ידע שהוא מסכן את עצם קיומו מדי שבוע במסגרת תפקידו. לא היה זה מצב שבו, כפי שהיה נכון לגבי רוב בני האדם לאורך ההיסטוריה, הוא חושב שהמוות הינו בלתי נמנע, והמרב שביכולתו לעשות הוא לבחור כיצד למות ולהעניק למוות הזה משמעות. פרי היה טרנס־הומניסט, והייתה לו תקווה אמיתית להמשיך לחיות. ובכל זאת פרי יצא למשימותיו ושם את חייו על הכף. לא משום שציפה לגמול אלוהי כלשהו. לא משום שציפה לחוות משהו לאחר מותו. הוא עשה זאת משום שהיו אנשים אחרים בסכנה, שגם להם לא היו נשמות בנות־אלמוות, ותקוותו שלו לחיים לא הייתה שווה יותר מתקוותם שלהם.

כאמור, לא הכרתי את ג׳ון פרי. איני יודע אם הוא אכן ראה את העולם באופן זה. אך העובדה שאתאיסט וטרנס־הומניסט עדיין יכול לשרת כשוטר, ועדיין מסוגל לרוץ אל תוך בניין בוער, מעידה על רוח האדם יותר מכל הקדושים המעונים שקיוו אי פעם להגיע לגן עדן.

אז זה היה שוטר ספציפי אחד...

...ועכשיו נעבור לגיבור־העל.

על פי המסורת הנוצרית, ישו היה מסוגל ללכת על המים, לשכך סערות ולגרש שדים במילה אחת. אלו בוודאי היו חיים נוחים מאוד. רעֵב? שכפל קצת לחם. עץ לא מוצא חן בעיניך? קלל אותו.* הרומאים עושים צרות? שַׁסֵּה בהם את אבא שלך. בסופו של דבר החיים הקסומים הללו הגיעו לקיצם כאשר ישו הסגיר את עצמו מרצונו לצליבה. מיסמור לצלב אינו דרך נוחה למות. אך על פי הנרטיב הנוצרי, ישו עשה זאת מתוך ידיעה שיחזור לחיים כעבור שלושה ימים ואז יעלה השמיימה. מה הוא האיום שגרם לישו לספוג את הסבל הזמני הזה, שבעקבותיו ציפו לו חיי נצח בגן עדן? האם היו אלה חייו של אדם בודד? האם הייתה זו השחיתות של הממסד הדתי ביהודה, או הדיכוי מצד רומא? לא: על פי הסיפור הנוצרי, גורלו

* ׳נס הלחם והדגים׳ ו׳קללת עץ התאנה׳ בברית החדשה.

הנצחי של כל אדם ואדם היה מונח על כף המאזניים בטרם ישו הניח לעצמו להיות ממוסמר זמנית לצלב.

אך איני רוצה לגנות אדם שאינו באמת אשם עד כדי כך. מה אם ישו – לא, הבה נֶהגה את שמו נכון: יֵשׁוּעַ – מה אם ישוע מנצרת מעולם לא הלך על המים, *ובכל זאת* יצא נגד הממסד הדתי ביהודה שזכה לגיבוי מכוחותיה של רומא?

האם מעשה כזה אינו ראוי לכבוד רב יותר מזה שמוענק לישו הנוצרי, שהיה בסך הכל משיח?

למרבה הצער, משום מה מעשה נתפס כגדול יותר כאשר הגיבור ניחן בעור פלדה ובכוחות דמויי־אל. משום מה נדמה שלָמות באופן זמני כדי להציל את העולם כולו מצביע על סגולה מוסרית גדולה יותר מאשר למות לתמיד במסגרת עימות מול ממסד דתי מושחת. מוות כזה נראה *בנאלי* כל כך, כאילו אנשים רבים אחרים לאורך ההיסטוריה עשו בדיוק אותו הדבר.

ממקומנו הבטוח אלפיים שנה בעתיד, אנו יכולים להטיח בישוע שלל ביקורות. אך ישוע פעל מתוך אמונה בצדקת דרכו, התעמת עם ממסד שראה כמושחת, ומת בשם אמונתו זו. ללא חוכמה בדיעבד, קשה לצַפות ממנו שיחזה את ההשפעה האמיתית של חייו על העולם. בהשוואה לרוב הנביאים האחרים בני תקופתו, סביר להניח שהוא היה פֶּן יותר, אלים פחות ואמיץ הרבה יותר. אם נשים בצד את ההשלכות הבלתי מכוּוונות, הדבר הגרוע ביותר שניתן לומר על ישוע הוא שאחרים בהיסטוריה עשו עבודה טובה יותר (אפיקורוס, בודהה ומרקוס אוֵרליוס עולים במחשבה). ישוע מת לתמיד, ומנקודת מבט מסוימת הוא עשה זאת בשמה של הכנוּת. אלף וחמש מאות שנים לפני הולדת המדע, כנות דתית לא הייתה אוקסימורון.

כפי שכתב סם האריס:[1]

> לא די בכך שישו היה אדם שחולל בקרבו תמורה כזו שדרשת ההר יכלה לבטא את רחשי ליבו הכנים. היה עליו להיות גם בן האלוהים, להיוולד לבתולה ולהיות מיועד לשוב לכדור הארץ עטור בענני תהילה. השפעתה של דוגמה כזו היא הרחקת המודל שמציג ישו

אל מחוץ להישג יד, לנצח. משנתו חדלה מלהיות אוסף של טענות אמפיריות על הקשר שבין אתיקה לתובנה רוחנית, ותחת זאת הופכת לאגדת ילדים מיותרת ומזוויעה למדי. על פי המודל הנוצרי, אי־אפשר להיות בדיוק כמו ישו. כל שנותר לאדם לעשות הוא לִמנות את חטאיו, להאמין בבלתי ייאמן ולהמתין לקיצו של העולם.

אני מסופק מאוד אם ישוע נשא אי פעם את דרשת ההר. ואף על פי כן, ישוע ראוי לכבוד. הוא ראוי לכבוד רב מזה שהנוצרים מעניקים לו.

אך מכיוון שישוע ככל הנראה ציפה שנשמתו תשרוד, הוא אינו ראוי ליותר כבוד מג׳ון פרי.

---

1. Sam Harris, *The End of Faith: Religion, Terror, and the Future of Reason* (WW Norton & Company, 2005).

106

מעגלי מוות רגשיים

רבים מאוד הם הפגמים בחשיבה האנושית שמובילים אותנו להפריז בהערכת המידה שבה התיאוריה האהובה עלינו מסבירה את העובדות. תאוריית הפְלוֹגיסטוֹן בכימיה הייתה יכולה להסביר כמעט הכל, כל עוד לא נדרשה לחזות זאת מראש. וככל שתשתמשו בתיאוריה המועדפת עליכם כדי להסביר תופעות רבות יותר, כך היא תיראה לכם נכונה יותר – הרי היא זכתה לאישוש מתצפיות רבות כל כך, הלא כן? ככל שהתיאוריה נראית נכונה יותר, כך עולה הסבירות שתטילו ספק בראיות הסותרות אותה. וככל שהתיאוריה המועדפת נראית כללית יותר, כך תנסו להחיל אותה על תחומים נוספים.

אם אתם מכירים אדם המאמין שבלגיה שולטת בחשאי במערכת הבנקאות האמריקאית, או שהוא מסוגל להשתמש בשדה־כוח רוחני כחול ובלתי נראה כדי לאתר מקומות חניה פנויים, סביר להניח שכך הדבר התחיל אצלו.

(רק תפקחו עיניים, ומיד תבחינו בשפע של ראיות המחזקות את התיאוריה הזו...)

מעגל המשוב החיובי הזה שבו אמון־יתר מזין אישוש ואישוש מזין אמון־יתר אכן מעורר אימה, ואחראי לטעויות רבות הן במדע והן בחיי היומיום.

אך הוא מתגמד לעומת מעגלי מוות* שמתחילים במטען של רגש חיובי – במחשבה *שמרגישה נהדר*.

שיטה פוליטית חדשה שיכולה להציל את העולם. מנהיג דגול, חזק, אציל וחכם. שיקוי פלא מדהים שיכול לרפא כאבי־בטן וסרטן.

בעצם, למה לא לבחור בשלושתם? מטרה דגולה זקוקה למנהיג דגול. מנהיג דגול אמור להיות מסוגל לרקוח איזה שיקוי קסם או שניים.

אפקט ההילה קובע שכל תכונה חיובית שאנו מזהים (כמו מראֶה מצודד או כוח פיזי) מעצימה את התפיסה של תכונות חיוביות אחרות (כמו חוכמה או אומץ). גם כשאין בכך כל היגיון, או גרוע מכך.

תכונות חיוביות מעצימות את התפיסה של תכונות חיוביות אחרות? הדבר מזכיר מאוד את האופן שבו אטום אורניום מתבקע פולט נייטרונים המבקעים אטומי אורניום נוספים.

רגש חיובי חלש הוא תת־קריטי** – הוא אינו יוצר סחרור שיוצא משליטה. אדם נאה נראה כֵּן יותר, מה שאולי גורם לו להיראות קצת נאה יותר – אך מקדם הכפלת הנייטרונים האפקטיבי קטן מאחד. מטאפורית, כמובן. התהודה מסבכת מעט את העניינים, אך לבסוף דועכת.

כאשר רגש חיובי עַז נקשר ל"דבר המופלא", התהודה מהדהדת בכל. קומוניסט אדוק רואה את חוכמתו של מרקס בכל המבורגר שנקנה במקדונלד'ס, בכל קידום שנשלל ממנו (ואשר היה מוענק לו בגן העדן האמיתי של הפועלים), בכל מערכת בחירות שאינה מתנהלת לרוחו, ובכל כתבה בעיתון ש"מוּטה לכיוון הלא נכון". בכל פעם שהוא משתמש ב"רעיון הגדול" כדי לפרש אירוע חדש, הרעיון הגדול זוכה לאישוש נוסף. התחושה נעשית טובה יותר – חיזוק חיובי – וכאשר

---

* *מעגל מוות* היא תופעה שבה נמלים מאבדות את דרכן ומתחילות לעקוב זו אחר זו במעגל סגור ואינסופי, עד שהן קורסות ומתות מתשישות.

** בפיזיקה גרעינית, מצב *תת־קריטי* (subcritical) מתאר תגובת שרשרת שאינה חזקה מספיק כדי להזין את עצמה, ולכן היא הולכת ודועכת במקום לגדול ולצאת מכלל שליטה.

משהו גורם לנו להרגיש טוב, הדבר מעורר בנו, למרבה הצער, רצון להאמין בו אפילו יותר.

כאשר הדבר המופלא מעורר תחושה טובה עד כדי כך שהוא גורם לכם *לחפש הזדמנויות חדשות* לשאוב ממנו תחושות טובות אף יותר, וליישם אותו לפרשנות של אירועים חדשים מדי יום, תהודת הרגש החיובי משולה לחדר מלא במלכודות עכברים טעונות בכדורי פינג־פונג.

אתם יכולים לקרוא לזה "מושך מאושר", "משוב חיובי מדי", "חוג נעול־מושפע", או "מלכודת מתיקות".* באופן אישי, אני מעדיף את המונח "מעגל מוות רגשי".

בפרק הבא: כיצד לעמוד בפני מעגל מוות רגשי. (רמז: לא על ידי סירוב להעריץ שום דבר שהוא מעתה והלאה, וגם לא באמצעות כליאת הדברים שאתם מעריצים בתוך תחום נפרד, מוגבל ובטוח.)

---

* *מושך מוזר* (תורת הכאוס ומערכות דינמיות) הוא סוג של *מושך* – דפוס שאליו המערכת נוטה להתכנס ולהישאר בו לאורך זמן, מתוך מגוון רחב של נקודות התחלה.

משוב חיובי (פסיכולוגיה התנהגותית ועוד) הוא מנגנון שבו תוצאת התהליך חוזרת ומזינה אותו, וכך מעצימה אותו ויוצרת מעגל שמגביר את עצמו.

חוג נעול־מופע (אלקטרוניקה) הוא מנגנון בקרה המאלץ את המערכת להסתנכרן ולהינעל על אות נכנס, כך שהיא מתקבעת עליו ואינה סוטה ממנו.

מלכודת דבש (אבטחת מידע ועוד) היא פיתיון שיוצר אשליה של מטרה מושכת וקלה להשגה כדי ללכוד תוקפים המנסים לנצל אותה.

107

התנגדו למעגל המוות המאושר

היה היה פעם אדם שהיה משוכנע שהגה רעיון נהדר. למעשה, ככל שחשב עליו יותר, כך הבין שזה אינו סתם רעיון *נהדר*, אלא *הרעיון המופלא ביותר אי פעם*. הרעיון הנהדר יפענח את סודות היקום, יחליף את סמכותו של הממסד המושחת והשגוי, יעניק כוחות כמעט־קסומים לאלו שישתמשו בו, יאכיל את הרעבים, ירפא את החולים, יהפוך את העולם כולו למקום טוב יותר, וכו׳ וכו׳ וכו׳.

האיש הזה היה פרנסיס בייקון. הרעיון הכביר שלו היה השיטה המדעית, והוא היה התמהוני היחיד לאורך ההיסטוריה שייחס לרעיון שלו תועלת עצומה כל כך לאנושות – והתברר כצודק לחלוטין. (בייקון לא המציא את המדע במו ידיו, כמובן, אך הוא בהחלט תרם לו, וייתכן שהיה הראשון להבין את עוצמתו.)

זוהי הבעיה בהחלטה הגורפת שלא להעריץ שום דבר לעולם: ישנם רעיונות שהם *באמת* עד כדי כך טובים. אף שאיש *טרם עמד* בהבטחות נועזות מאלו של בייקון, לפחות לא עד כה.

אבל אם כך, כיצד נוכל להתנגד למעגל המוות המאושר ביחס למדע עצמו? מעגל המוות המאושר מתחיל כשאתם מאמינים שמשהו הוא *כה* מופלא, עד שאפקט ההילה מוביל אתכם למצוא בו *עוד ועוד* תכונות חיוביות. התהליך הזה גורם לכם לראות אותו כמופלא *אף יותר*, וכן הלאה, בסחרור מעלה מעלה עד התחתית. מה אם המדע מועיל *עד כדי כך* שאיננו יכולים גם להכיר בתפארתו האמיתית וגם

לשמור על שפיותנו? נשמע כמו דבר נחמד לומר, לא? *אוי לא, זה מתחיל, תברחווווו...*

אם תשלפו את חוכמת־המטמון העמוקה המוכנה־מראש עבור *אל תגזימו בהערצת המדע*, תמצאו מחשבות כמו: "המדע נתן לנו מיזוג אוויר, אבל הוא גם יצר את פצצת המימן," או "המדע יכול ללמד אותנו על כוכבים ועל ביולוגיה, אבל הוא לעולם לא יוכל להוכיח או להפריך את קיומו של הדרקון במחסן שלי." ואולם, האנשים *שהגו* את המחשבות הללו לא ניסו להתנגד למעגל המוות המאושר. הם לא חששו שהערצת המדע שלהם תצא מכלל שליטה. סביר יותר להניח שהם פשוט לא אהבו את מה שהיה למדע לומר על אמונות המחמד שלהם, וחיפשו דרכים לערער על סמכותו.

האמירות השליליות *השגורות* שנהוג להטיח במדע כנראה לא ידברו לאלו שבאמת חשים את ההתעלות שהמדע מעורר – הם אינם קהל היעד. לכן, נצטרך לחפש דברים שליליים אחרים לומר במקום זאת.

אך אם תחפשו משהו שלילי לומר על המדע באופן סלקטיבי – אפילו בניסיון להתנגד למעגל המוות המאושר – האם אינכם מעידים על עצמכם מיד שאתם חוטאים ברציונליזציה? מדוע שתקשיבו למחשבות שלכם, אם אתם יודעים שאתם מנסים לבצע מניפולציה על עצמכם?

אני נוטה להיות סקפטי כלפי אנשים שטוענים כי ניתן להשתמש בהטיה אחת כדי לקזז אחרת. זה נשמע לי כמו מכונאי רכב שאומר שהמנוע של המגב הימני שלכם מקולקל, אבל במקום לתקן אותו, הוא פשוט ישבור את המגב השמאלי כדי לאזן את העניינים. זוהי בדיוק סוג ההתחכמות שתוביל לכך שתירו לעצמכם ברגל. הפתרון, יהיה אשר יהיה, חייב לכלול אמונה בדברים נכונים, ולא בניסיון לשכנע את עצמכם שאתם מאמינים בדברים שאתם יודעים שהם שקריים.

האם תוכלו למנוע את מעגל המוות המאושר על ידי הגבלת ההערצה שלכם לַמדע לתחום צר בלבד? חלק ממעגל המוות המאושר הוא לראות את הרעיון המופלא בכל מקום – לחשוב, למשל, כיצד

קומוניזם יכול לרפא סרטן אם רק תינתן לו ההזדמנות. כנראה שהסימן המובהק מכולם למנהיג כת הוא טענתו למומחיות לא רק בתחום אחד, או באשכול של תחומים קרובים, אלא *בהכל*. הגורו יודע מה חברי הכת צריכים לאכול וללבוש, במה לעבוד, עם מי עליהם לקיים יחסי מין, אילו יצירות לראות ולאיזו מוזיקה להאזין...

לרוע מזלה של תוכנית זו, רוב האנשים נכשלים כישלון חרוץ כשהם מנסים לתכנן את הקופסה הקטנה והמסודרת שבתוכה על המדע להישאר. התעלול הרגיל, "היי, מדע לא ירפא סרטן," פשוט לא יעבוד. "למדע אין מה לומר על אהבת הורה לילדיו" – סליחה, אך זה פשוט לא נכון. אם תנסו לנתק את המדע מאהבת הורים לילדיהם, למשל, לא רק שתתכחשו למדעים הקוגניטיביים ולפסיכולוגיה האבולוציונית, אלא תתכחשו גם לכך שמרטין רות׳בלָט הקימה את חברת התרופות United Therapeutics כדי לחפש תרופה ליתר לחץ דם ריאתי עבור בתה (בהצלחה, יש לציין). המדע קשור באופן לגיטימי, בדרך זו או אחרת, לכמעט כל היבט חשוב של הקיום האנושי.

ובכן, מהי דוגמה לטענה חיובית *שקרית* שניתן לטעון לגבי המדע?

לעניות דעתי, טענה שקרית אחת היא שהמדע הוא כה מופלא, עד שמדענים אינם צריכים כלל לנסות לקחת אחריות מוסרית על עבודתם, ושהכל יסתדר מעצמו. טענה זו, בעיניי, נובעת מאי־הבנה של האופן שבו המדע מועיל לאנושות. מדענים הם בני אדם, יש להם רצון להיטיב עם החברה בדיוק כמו לרוב האנשים, וזוהי לפחות *חלק* מהסיבה שבגללה בסופו של דבר המדע גורם ליותר תועלת מנזק.

אך נראה שנקודה זו אינה מוסכמת על כולם. לכן, הנה טענה חיובית שקרית ופשוטה יותר: "די בפרסום מספיק מאמרים מדעיים כדי לרפא חולה סרטן." או, "סוציופתים היו יכולים להפוך לאנשים נורמליים לחלוטין, אילו רק היו מתחייבים שלעולם לא יאמינו בשום דבר שאינו מגובה בראיות אמפיריות הניתנות לשחזור ברמת מובהקות של $p < 0.05$."

הדרך להימנע מאמונה בהצהרות כאלו אינה באמצעות הצבת

"תקרה רגשית", כלומר להחליט שהמדע הוא רק *קצת* נחמד. גם לא על ידי חיפוש סיבות להאמין שפרסום מאמרים בכתבי עת *גורם* לסרטן, או שלמדע אין מה לומר על סרטן לכאן או לכאן.

במקום זאת, אם אתם יודעים טוב מספיק כיצד המדע פועל, אזי ברור לכם שאף על פי שייתכן ש"המדע ירפא סרטן", חולה סרטן שכותב מאמרים מדעיים לא יחווה נסיגה פלאית של המחלה. שרשרת הסיבה והתוצאה *הספציפית* שהוצעה פשוט לא תעבוד.

מעגל המוות המאושר הוא בעיה רגשית רק בגלל בעיה תפיסתית – אפקט ההילה – שהופך אותנו לבעלי נטייה גדולה יותר לקבל טענות חיוביות בעתיד מרגע שקיבלנו טענה חיובית ראשונית. איננו יכולים להיפטר מהאפקט הזה רק באמצעות הבעת משאלה. סביר להניח שהוא תמיד ישפיע עלינו במידת מה. אבל אנו כן יכולים לבלום, לעצור, להתייחס לכל טענה חיובית נוספת כאל פרט המכביד על נטל ההוכחה, ולהתמקד בנקודות הספציפיות של הטענה במנותק מהמטען החיובי שלה.

מה אם טענה חיובית ספציפית "אינה ניתנת להפרכה" אבל יש "טיעונים בעד ונגד"? למעשה, מילים אלו צריכות להדליק נורה אדומה באופן כללי, משום שלעיתים קרובות אנשים אומרים זאת כשהם ממחזרים ראיות או מתחמקים מנקודות התורפה האמיתיות. בהתחשב בסכנה של מעגל המוות המאושר, הגיוני לנסות ולהימנע מתחושת התרוממות רוח ביחס לטענות *שטרם הוכחו* – כדי לא להפוך אותן למקור נוסף של רגש חיובי כלפי משהו שכבר חיבבתם מלכתחילה.

מעגל המוות המאושר הוא בעיה רגשית *גדולה* רק בגלל משוב חיובי מוגזם, והיכולת של התהליך לעבור את הסף הקריטי. ייתכן שלא תוכלו להעלים לחלוטין את אפקט ההילה, אבל תוכלו ליישם מספיק חשיבה ביקורתית כדי להשאיר את ההילות במצב תת־קריטי – ולוודא שהתהודה דועכת במקום להתפוצץ.

אפשר אף לומר שהבעיה כולה מתחילה כשאנשים לא טורחים לבחון באופן ביקורתי כל פרט חדש המכביד על נטל ההוכחה – כשהם

אינם דורשים ראיות חזקות מספיק כדי לפצות על תוספת המורכבות, אינם מחפשים פגמים באותה הנחישות שבה הם מחפשים תמיכה, ואינם מעוררים סקרנות ביקורתית – מרגע שקיבלו הנחות יסוד מסוימות. אלמלא כשל החיתוך, אפקט ההילה כנראה עדיין היה קיים, אך לא היה נוצר מעגל מוות מאושר.

אפילו ביחס ל"דברים המופלאים" ביותר ביקום הידוע, רציונליסט מושלם שהיה מחפש בדיוק את הראיות הנדרשות עבור כל טענה (חיובית) נוספת לא היה חווה תהודה רגשית. אינכם מסוגלים לעשות זאת, אך אתם בהחלט יכולים להישאר קרובים מספיק לרציונליות כדי למנוע מהשמחה שלכם לצאת מכלל שליטה.

המקרים המסוכנים באמת הם אלה שבהם *כל ביקורת על איזושהי טענה חיובית כלפי "הדבר המופלא" מעוררת תחושה רעה או נחשבת לבלתי מקובלת חברתית*. טיעונים הם חיילים, כל טענה חיובית היא חייל בצד שלנו, ודקירת החיילים שלכם בגב נחשבת לבגידה. בשלב הזה תגובת השרשרת הופכת לעל־קריטית. ארחיב על כך בהמשך.

סטיוארט ארמסטרונג משיא עצה דומה מאוד:

> פרקו את "הדבר המופלא" שלכם לרעיונות קטנים ועצמאיים, *והתייחסו אליהם ככאלה*.
> למשל, מרקסיסט יפרק את "הדבר המופלא" של מרקס לתאוריית הערך של העבודה, לתיאוריה על היחסים הפוליטיים בין המעמדות, לתאוריית השכר, ולתיאוריה העוסקת במצב הפוליטי האידיאלי של האנושות. לאחר מכן עליו להעריך כל אחת מהן באופן עצמאי, והאמת או השקר שבאחת התיאוריות אינם צריכים להקרין על האחרות. אם נצליח לעשות זאת, נהיה מוגנים מפני מעגל המוות, שכן כל תיאוריה כשלעצמה היא מצומצמת מכדי להתניע מעגל בכוחות עצמה.

באופן מטאפורי, הדבר דומה למניעת התלכדותן של מסות פלוטוניום תת־קריטיות. סביר הרבה פחות ששלושה רעיונות מופלאים יטריפו את דעתכם מאשר רעיון מופלא אחד. העצה של ארמסטרונג מסייעת גם לעודד ספציפיות: ברגע שמישהו אומר "פרסום מספיק

מאמרים יכול לרפא את הסרטן שלך," אתם שואלים: "האם זה יתרון של השיטה הניסויית, ואם כן, באיזה שלב של תהליך הניסוי מחלים הסרטן? או שמא זהו יתרון של המדע כתהליך חברתי, ואם כך, האם הוא נשען על כך שמדענים בודדים רוצים לרפא סרטן, או שמא הם יכולים להיות מונעים מאינטרס אישי?" יש לקוות שהדבר יוביל אתכם הרחק מהתחושה הטובה (או הרעה), אל עבר זיהוי הבלבול וחוסר הביסוס.

לסיכום, *תוכלו* להימנע ממעגל מוות מאושר באמצעות:

- פירוק הרעיון המופלא לחלקיו השונים.
- התייחסות לכל פרט נוסף כאל פרט המכביד על נטל ההוכחה.
- חשיבה על הפרטים הספציפיים של שרשרת הסיבתיות, במקום התמקדות בתחושות הטובות או הרעות.
- הימנעות ממחזור ראיות.
- הימנעות *משאיבת שמחה* מטענות ש"אי אפשר *להוכיח* שהן שגויות."

אך *לא* באמצעות:

- סירוב גורף להעריץ דבר מה אי פעם.
- עריכת חיפוש מוטה אחר נקודות שליליות עד שתרגישו אומללים שוב.
- כליאת רעיון בכוח בתוך קופסה אטומה.

108

על-קריטיות אל-ביקורתית

מדי פעם אנשים מתווכחים בשאלה האם אתאיזם הוא "דת". כפי שציינתי במקום אחר, במאמר "תכלית ופרגמטיזם",* ויכוח על משמעותה של מילה מעיד כמעט תמיד שאיבדתם את הקשר לשאלה המקורית. כיצד בכלל מתחיל ויכוח כזה?

אתאיסט נושא נאום ומאשים את ה"דת" באינקוויזיציה, במסעות הצלב ובסכסוכים שונים עם האסלאם או בתוכו. האדם הדתי עשוי להשיב: "אבל אתאיזם הוא גם כן דת, משום שגם לך יש אמונות בנוגע לאלוהים. אתה מאמין שאלוהים אינו קיים." בתגובה, האתאיסט עונה: "אם אתאיזם הוא דת, הרי שאי־איסוף בולים הוא תחביב" — והוויכוח ניצת.

לחלופין, האדם הדתי עשוי לענות: "אבל זוועות חמורות באותה מידה בוצעו על ידי סטלין, שהיה אתאיסט ודיכא כנסיות בשם האתאיזם. לכן אתה טועה כשאתה מאשים את הדת באלימות." כעת האתאיסט עשוי להתפתות להשיב בכשל "שום סקוטי אמיתי"** ולומר: "הדת של סטלין הייתה הקומוניזם." האדם הדתי עונה: "אם קומוניזם הוא דת, הרי שמועדון המעריצים של 'מלחמת הכוכבים'

* המאמר ("Purpose and Pragmatism") פורסם באתר *LessWrong*.

**מורה: כל סקוטי אוהב הַאגיס.

תלמידה: הדוד שלי שהוא סקוטי לא אוהב האגיס.

מורה: *שום סקוטי אמיתי* לא היה מתנגד להאגיס. (ויקיפדיה)

הוא ממשלה" – והוויכוח מתלקח.

האם אדם "דתי" אמור להיות מוגדר כמי שיש לו דעה נחרצת לגבי קיומו של אל אחד לפחות – למשל, אדם המייחס הסתברות נמוכה מ-10% או גבוהה מ-90% לקיומו של זאוס? או שמא אדם "דתי" צריך להיות מוגדר כמי שמחזיק באמונה חיובית, נניח הסתברות גבוהה מ-90%, בקיומו של אל אחד לפחות? במקרה הראשון, סטלין היה "דתי", ובמקרה השני הוא היה "לא דתי".

אך זוהי בדיוק הדרך השגויה לגשת לבעיה. מה שאתם באמת רוצים לדעת – מה שהוויכוח נסב עליו במקור – הוא מדוע, בנקודות מסוימות בהיסטוריה האנושית, קבוצות גדולות של אנשים נטבחו ועונו, לכאורה בשמו של רעיון. הגדרתה מחדש של מילה לא תשנה את עובדות ההיסטוריה לכאן או לכאן.

הקומוניזם היה קטסטרופה מורכבת, וייתכן שאין "למה" אחד בודד, או חוליה קריטית אחת בשרשרת הסיבתיות. אבל אילו נדרשתי להצביע על טעות שורשית אחת, היא הייתה... ובכן, אניח לאלוהים לנסח זאת במקומי:

כִּי יְסִיתְךָ אָחִיךָ בֶּן־אִמֶּךָ אוֹ־בִנְךָ אוֹ־בִתֶּךָ אוֹ אֵשֶׁת חֵיקֶךָ אוֹ רֵעֲךָ אֲשֶׁר כְּנַפְשְׁךָ בַּסֵּתֶר לֵאמֹר, נֵלְכָה וְנַעַבְדָה אֱלֹהִים אֲחֵרִים אֲשֶׁר לֹא יָדַעְתָּ אַתָּה וַאֲבֹתֶיךָ. מֵאֱלֹהֵי הָעַמִּים אֲשֶׁר סְבִיבֹתֵיכֶם הַקְּרֹבִים אֵלֶיךָ אוֹ הָרְחֹקִים מִמֶּךָּ, מִקְצֵה הָאָרֶץ וְעַד־קְצֵה הָאָרֶץ. לֹא־תֹאבֶה לוֹ **וְלֹא תִשְׁמַע אֵלָיו**, וְלֹא־תָחוֹס עֵינְךָ עָלָיו וְלֹא־תַחְמֹל וְלֹא־תְכַסֶּה עָלָיו. כִּי **הָרֹג תַּהַרְגֶנּוּ**, יָדְךָ תִּהְיֶה־בּוֹ בָרִאשׁוֹנָה לַהֲמִיתוֹ, וְיַד כָּל־הָעָם בָּאַחֲרֹנָה. וּסְקַלְתּוֹ בָאֲבָנִים וָמֵת, כִּי בִקֵּשׁ לְהַדִּיחֲךָ מֵעַל ה' אֱלֹקֶיךָ.

– דברים י"ג, ז'-י"ב

בדומה לכך, זה היה הכלל שהציב סטלין עבור הקומוניזם, והיטלר עבור הנאציזם: אם אחיך מנסה לומר לך מדוע מרקס טועה, אם בנך מנסה לומר לך שהיהודים אינם מתכננים לכבוש את העולם – אל תתווכח איתו ואל תציג בפניו את הראיות שבידך. אל תבצע ניסויים

הניתנים לשחזור ואל תבחן את ההיסטוריה. פשוט הסגר אותו מיד למשטרה החשאית.

ציינתי בפרק הקודם שאחד המפתחות להתנגדות למעגל מוות רגשי הוא עקרון "נטל הפרטים" – פשוט *לזכור* להטיל ספק בפרטיה הספציפיים של כל טענה חיובית נוספת הנוגעת ל"רעיון המופלא". (זו אינה עצה מובנת מאליה. אנשים שוכחים לעיתים קרובות לעשות זאת כשהם מאזינים לעתידן המשרטט תחזיות מפורטות להפליא על נפלאות המחר, וקל וחומר כשהם מהרהרים ברעיון האהוב עליהם מכל.) הדבר לא יעלים את אפקט ההילה, אך יש לקוות שיפחית את התהודה אל מתחת לסף הקריטי, כך שטענה משמחת אחת תעורר בממוצע פחות מ-1.0 טענות משמחות נוספות.

ההפך הגמור מעצה זו, הדבר שמשלח את אפקט ההילה למצב *על-קריטי*, מתרחש כאשר מתעוררת התחושה שאסור לצאת נגד טענה חיובית *כלשהי* הנוגעת ל"רעיון המופלא". פוליטיקה היא אויב התבונה. טיעונים הם חיילים. ברגע שאתם יודעים באיזה צד אתם, עליכם לתמוך בכל הטענות של אותו צד, ולצאת נגד כל טענות האויב. אחרת, הדבר משול להושטת סיוע ותמיכה לאויב, או לנעיצת סכין בגב חבריכם.

אם...

- ...אתם חשים שלסתור אדם שמשמיע טענה חיובית פגומה בזכות האבולוציה משמעו סיוע ותמיכה לבריאתנים;
- ...אתם חשים שאתם צוברים זכויות רוחניות עם כל דבר טוב שאתם אומרים על אלוהים, ושוויכוח על כך יפגע בקשר שלכם איתו;
- ...יש לכם תחושה מובהקת ששאר הנוכחים בחדר לא יראו אתכם בעין יפה משום ש"אינכם תומכים בחיילינו" אם תצאו נגד המלחמה האחרונה;
- ...אמירת דבר-מה נגד הקומוניזם תגרום לכם ~~להיסקל באבנים~~ לעמוד מול כיתת יורים;

...הרי שמעגל המוות הרגשי הפך לעל־קריטי. כעת, זהו מעגל מוות מאושר־על.

הדת כשלעצמה אינה הקטגוריה המרכזית ביחס לשאלה המקורית שלנו: "מה גורם לטֶבח?" ההבחנה הטובה ביותר ששמעתי בין השקפות עולם "על־טבעיות" ל"נטורליסטיות" היא שהשקפת עולם על־טבעית מניחה שישויות תודעתיות, כמו רוחות, הן אבני בניין בסיסיות של המציאות. לעומתה, השקפת עולם נטורליסטית מסבירה תופעות תודעתיות כתוצר של רכיבים שאינם תודעתיים. התמקדות בהבחנה הזו כמקור הבעיה, משמעה לאמץ את התפיסה שלדת יש מעמד חריג. אמנם כדאי לשים לב לטענות על־טבעיות משום שהן תמיד מתבררות כשגויות מסיבות יסודיות למדי, אך זהו עדיין רק סוג אחד של טעות.

מעגל מוות רגשי יכול להתגבש סביב אמונות על־טבעיות. הדבר קורה במיוחד בדתות מונותאיסטיות, שבפסגתן ניצבת "ישות סופר־מאושרת" המוגדרת בעיקר מעצם ההסכמה עם כל טענה נחמדה לגביה. התופעה חריפה במיוחד כאשר מערכות אמונה משתכללות מספיק כדי לאיים בעונשים על־טבעיים על כפירה. אך מעגל המוות יכול להתחיל גם סביב מהפך פוליטי, מנהיג כריזמטי, אמונה בייעוד גזעי, או תיאוריה כלכלית. לקח ההיסטוריה הוא שמעגלי מוות רגשיים הם מסוכנים, בין אם הם במקרה כרוכים בתפיסה על־טבעית ובין אם לאו. דת אינה תופעה ייחודית מספיק, כקטגוריה של טעות, כדי להיות הבעיה המרכזית.

סם האריס התקרב יותר לאמת כשהפנה את האצבע המאשימה כלפי *אמונה עיוורת*. אם אינכם דורשים נטל הוכחה הולם עבור כל טענה נחמדה נוספת באשר היא, התהודה הרגשית מתחילה בקלות *רבה*. הביטו על אנשי הניו־אייג' המסכנים. הנצרות פיתחה הגנות מפני ביקורת באמצעות הטענה בדבר סגולותיה של האמונה. חסידי הניו־אייג' יורשים תרבותית את מחשבת־המטמון שלפיה אמונה היא דבר חיובי, אך חסרים להם כתבי הקודש של הנצרות – ששוללים כל השקפה אחרת – כדי להרחיק רעיונות מתחרים. אנשי הניו־אייג'

מוצאים את עצמם במעגלי מוות מאושרים סביב כוכבים, עצים, מגנטים, דיאטות, לחשים, חדי־קרן...

אך מעגל המוות הרגשי הופך קטלני הרבה יותר לאחר שביקורת מתחילה להיחשב לחטא, לטָעות חברתית חמורה, או לפשע. ישנם דברים בעולם הזה שראויים לשבחים רבים, ואי־אפשר לומר *באופן גורף* ששבח מעבר לנקודה מסוימת הוא אסור. אבל *לעולם* אין רעיון שהוא כה נכון, עד שיהיה זה שגוי להעביר ביקורת על הטיעונים התומכים בו. לעולם לא. לעולם־לעולם לא, בשום אופן ולנצח נצחים. *זהו* כלל גורף. הרוב המכריע של האמונות האפשריות במרחב תשובות שאינו טריוויאלי הן שגויות. באותו אופן, הרוב המכריע של *הטיעונים התומכים* האפשריים, אפילו עבור אמונה נכונה, הם שגויים בעצמם – ואפילו הרעיון המאושר ביותר לא יוכל לשנות זאת.

וחל איסור מוחלט וחמור מאין כמוהו להגיב לביקורת באלימות. ישנם מעט מאוד צווים באמנות הרציונליות האנושית שאין לצידם מילות התניה כמו "אם", "אבל", או סעיפי מילוט. זהו אחד מהם. טיעון גרוע מקבל טיעון־נגדי. הוא לא מקבל כדור. לעולם לא. לעולם־לעולם לא, בשום אופן ולנצח נצחים.

109

קירור באידוי של אמונות קבוצתיות

ראשוני חוקרי הכתות הופתעו לגלות שכאשר כתות סופגות זעזוע משמעותי – נבואה שמתבדה או פגם מוסרי של מנהיג הכת שנחשף – הן לעיתים קרובות שבות חזקות מבעבר, באמונה עזה יותר ובקיצוניות רבה יותר. למשל, 'עֲדֵי יהוה' חזו שיום הדין יגיע בשנת 1975, בהתבסס על חישובים מקראיים. שנת 1975 באה וחלפה. כת האוּנַריוּס, שעדיין פעילה כיום, שרדה את אי־הופעתו של צי חלליות בין־גלקטי ב-27 בספטמבר 1975.

מדוע שאמונה קבוצתית *תתחזק* לאחר שנתקלה בראיות־נגד מוחצות?

הפרשנות המקובלת לתופעה זו נשענת על דיסוננס קוגניטיבי. כשאנשים מבצעים פעולות בלתי הפיכות למען אמונה כלשהי – כמו חלוקה של כל רכושם בציפייה לנחיתתן של צלחות מעופפות – הם פשוט אינם מסוגלים להודות בטעותם. האתגר שניצב בפני אמונתם מעמיד אותם בפני דיסוננס קוגניטיבי עצום. הם מוכרחים לגייס הצדקות כדי להתמודד עם הזעזוע, וכך הם נעשים קיצוניים יותר. לפי פרשנות זו, העלייה בקיצוניות הקבוצתית היא תולדה של עלייה בקיצוניות האישית של החברים.

לפני זמן מה התבוננתי בהדמיה ממוחשבת המדגימה שימוש

ב״קירור באידוי״* ליצירת עיבוי בוֹזֿאיינשטיין,** והכתה בי ההבנה שייתכן כי כוח אחר לחלוטין פועל כאן ליצירת ההקצנה. קירור באידוי מתבצע על ידי ׳מחסום אנרגיה׳ המאפשר רק לאטומים חמים מספיק לעזוב. מטבעה, אנרגיית חום היא סטטיסטית – לא כל האטומים נעים בדיוק באותה המהירות, והאנרגיה הקינטית של כל אטום משתנה כשהם מתנגשים זה בזה. אם תציבו מחסום אנרגיה שהוא גבוה רק במעט מאנרגיית החום הממוצעת, יד המקרה תעניק מדי פעם לאטום מסוים אנרגיה קינטית גבוהה מספיק כדי לברוח מן המלכודת. כשאטום מהיר באופן חריג בורח, הוא לוקח עימו כמות גבוהה באופן חריג של אנרגיה קינטית, והאנרגיה הממוצעת יורדת. הקבוצה נעשית קרה משמעותית ממחסום האנרגיה שמקיף אותה.

בספר הקלאסי *כשהנבואה נכשלת*** מאת פֶסטינגֶר, ריקֶן ושכטר, אחד מחברי הכת קם ויוצא מן הדלת מיד לאחר שהצַלחת המעופפת לא נחתה.[1] מי מתייאש ועוזב *ראשון*? חבר כת *ממוצע*? או אולי חבר ספקן יחסית, שעד לאותו רגע שימש קול מתון וגורם מרסן עבור החברים הקיצוניים יותר?

לאחר שהחברים בעלי ״האנרגיה הקינטית״ הגבוהה ביותר נוטשים את הקבוצה, הדיונים הנותרים יתנהלו בין חברי הכת הקיצוניים ביותר בקצה האחד, לבין הקיצוניים־מעט־פחות בקצה השני, והקונצנזוס הקבוצתי יתייצב אי־שם ב״אמצע״ החדש.

ומה תהיה המקבילה לקריסה היוצרת עיבוי בוזֿאיינשטיין? ובכן, אין צורך אמיתי למתוח את האנלוגיה עד כדי כך. אך אולי אתם

---

* *קירור באידוי* (evaporative cooling) הוא תהליך שבו החלקיקים המהירים והחמים ביותר עוזבים את המערכת, מה שמוריד את האנרגיה הממוצעת ומקרר את החלקיקים שנותרו מאחור, בדומה לאופן שבו זיעה מקררת את הגוף.

** *עיבוי בוֹזֿאיינשטיין* (Bose-Einstein condensate) הוא מצב צבירה של החומר הנוצר בטמפרטורות קרובות לאפס המוחלט, שבו חלקיקים מאבדים את התנהגותם העצמאית ומתמזגים יחד למצב אחיד שבו הם מתנהגים כמעין ״אטום־על״ בודד.

*** הספר *כשהנבואה נכשלת* (When Prophecy Fails) עוקב אחר כת עב״מים קטנה שהאמינה שסוף העולם יגיע בתאריך מסוים. התבדותה של הנבואה לא ריסקה את אמונת חברי הכת אלא דווקא הקצינה אותה. הספר הניח את היסודות לתאוריית הדיסוננס הקוגניטיבי.

זוכרים שהשתמשתי באנלוגיה של תגובת שרשרת גרעינית לתיאור מעגל המוות הרגשי. כאשר קבוצה פולטת החוצה את כל הקולות המתונים שבה, אוסף האנשים שמעודדים זה את זה ומדכאים דעות חולקות עלול להעלות את רמת הקיצוניות בקבוצה. (אין כאן אנלוגיה תרמודינמית, אלא אם מישהו יפַתח נשק גרעיני שמתפוצץ כשהוא מתקרר.)

כאשר הרומן ארוך־השנים של אָיִין רַאנד עם נתנאל בְּרַנדֶן נחשף בפני התנועה האובייקטיביסטית, פלג ניכר מחברי התנועה התפצל והצטרף לברנדן לאימוץ "מערכת פתוחה" של אובייקטיביזם, שכבר לא הייתה כה כבולה לראנד. מי נשאר עם איין ראנד אפילו לאחר התפוצצות השערורייה? אלו *שבאמת ובתמים* האמינו בה – ואולי גם כמה מהמתלבטים, שלאחר עזיבת הקולות המאזנים נחשפו לטיעונים מצד אחד בלבד. הדבר עשוי להסביר מדוע מכון איין ראנד נעשה (לכאורה) קיצוני יותר לאחר הפיצול, בהשוואה לקבוצת הליבה האובייקטיביסטית המקורית תחת ברנדן וראנד.

לפני כמה שנים הייתי חבר ברשימת תפוצה של טרנס־הומניסטים, שבה קבוצה קטנה אשר דגלה ב"טרנס־הומניזם סוציאל־דמוקרטי" גידפה בארסיות כל ליברטריאן ברשימה. רוב הליברטריאנים עזבו את הרשימה, ומרבית האחרים פשוט ויתרו על כתיבת תגובות. כתוצאה מכך, הקבוצה שנותרה נטתה שמאלה באופן משמעותי. האם זה היה מכוון? כנראה שלא, שכן אינני חושב שהמגדפים התמצאו בפסיכולוגיה עד כדי כך. (למעשה, איני זוכר שראיתי את אנלוגיית הקירור באידוי במקום אחר, אם כי אין זה אומר שהיא לא נוסחה בעבר.) לכל היותר, אולי הם חשבו להפוך את עצמם ל"דגים גדולים יותר בבריכה קטנה יותר".

זוהי אחת הסיבות שבגינן חשוב לסַגל נטייה לסובלנות כלפי דעות חולקות. המתינו הרבה *מעבר* לנקודה שבה נראה לכם מוצדק לסלק חבר מהקבוצה, לפני שתסלקו אותו בפועל. אם תיפטרו מנקודות הקיצון הישנות, עמדת הקבוצה תנוע למרכז חדש, ומישהו אחר יהפוך לעוף המוזר התורן. אם תסלקו גם אותו, אתם בדרך הבטוחה

להפוך לעיבוי בוז־איינשטיין – ו, אה, להתפוצץ.

מנגד: תומאס קון האמין שמדע חייב להתגבש לכדי "פרדיגמה" – כזו הנשענת על שפה טכנית משותפת שמדירה אנשים מבחוץ – לפני שיכולה להתבצע עבודה של ממש. בשלבי ההתהוות של דיסציפלינה מדעית כלשהי, טען קון, חסידיה משקיעים מאמצים כבירים בניסיון להנגיש את עבודתם לאקדמאים מבחוץ. ואולם (עדיין לדברי קון), תחום מדעי יכול להשיג התקדמות אמיתית כדיסציפלינה טכנית רק כאשר הוא זונח את הדרישה לנגישות חיצונית, ורק כאשר המדענים הפועלים במסגרת הפרדיגמה מניחים שקיימת ביניהם היכרות מוקדמת עם ליבות ידע טכניות נרחבות. הדבר נשמע ציני בהשוואה לרטוריקות המקובלות על אודות הנגשת המדע לציבור הרחב, אך אני בהחלט יכול לזהות כאן גרעין של אמת.

התיאוריה האישית שלי לניהול קהילות מקוונות היא שעליכם להיות מוכנים להדיר טרולים ומפיצי ספאם כדי לאפשר קיום של דיון פורה. אתם חייבים אפילו להיות מוכנים להדיר אנשים נחמדים אך נטולי רקע טכני מרשימות תפוצה טכניות, אם ברצונכם להגיע לתפוקה ממשית. שיח פתוח באמת באינטרנט מידרדר במהירות. אלו דווקא הטרולים *הרהוטים* שמגירושם עליכם להיזהר, על פי תיאוריה זו – שכן הם ממלאים פונקציה נסתרת בכך שהם מעניקים לגיטימציה לאי־הסכמות פחות קיצוניות. אך אל לכם לאפשר נוכחות של כל כך הרבה טרולים רהוטים עד שהם יתחילו להתווכח זה עם זה, או להשתלט על השיח. אם יש לכם בסביבה אדם אחד שהוא פרקליט השטן הידוע לשמצה בגישתו הלעומתית, אף אדם אחר המביע אי־הסכמה סבירה ומתקבלת יותר על הדעת לא יבלוט כחריג. ייתכן שתאוריית ניהול הקהילות הזו לא שירתה אותי היטב בפועל, אז קחו אותה בעירבון מוגבל.

1. Leon Festinger, Henry W. Riecken, and Stanley Schachter, *When Prophecy Fails: A Social and Psychological Study of a Modern Group That Predicted the Destruction of the World* (Harper-Torchbooks, 1956).

110

כשאיש אינו מעז לקרוא לאיפוק

בוקר אחד קמתי מהמיטה, הדלקתי את המחשב, וקיבלתי את האיימייל עם מקבץ החדשות היומי. ביום זה, דיווחו החדשות ששני מטוסים חטופים התרסקו אל תוך מגדלי התאומים.

סדר המחשבות הראשונות שלי היה כדלקמן:

כנראה אני באמת חי בעתיד.
תודה לאל שלא מדובר בנשק גרעיני.

ואז:

תגובת היתר לכך תהיה גרועה פי עשרה מהאירוע עצמו.

ההערכה שהתגובה תהיה "גרועה פי עשרה" בלבד התבררה כהערכת חסר קיצונית. אפילו אני לא שיערתי עד כמה המצב יידרדר. זהו האתגר שבפסימיות: *קשה מאוד* לכוון נמוך מספיק כדי להיות מופתעים לטובה באותה תדירות ובאותה עוצמה שבהן מופתעים לרעה.*

עם זאת, הבנתי מיד שכולם, בכל מקום, ידברו על עד כמה האירוע הזה נורא ואיום, ושאף אחד לא יעז להיות הקול שקורא לאיפוק, לתגובה מידתית. בתחילה, ב-11 בספטמבר, ההערכה הראשונית

* "אדון האופל חי. ברור שהוא חי [...] למה שאני אחשוב שהוא מת רק כי מישהי אמרה שהגופה שלו נמצאה מפוחמת? אין ספק שיש לי עוד הרבה מה ללמוד על רזי הפסימיות." (הארי ג'יימס פוטר־אוונס־ורס)

הייתה שנהרגו ששת אלפים בני אדם. כל פוליטיקאי שהיה מעז לומר ש-״6,000 הרוגים הם שמינית ממספר ההרוגים השנתי בתאונות דרכים בארצות הברית,״ היה מתבקש להתפטר באותו הרגע.

ה-11 בספטמבר אכן לא היה יום טוב. אבל אם *כולם* מרוויחים נקודות זכות על כך שהם מדגישים את עוצמת הכאב, *ואיש* אינו מעז לקרוא לאיפוק באשר לעוצמת התגובה, הרי שהתגובה תהיה גדולה מהמידה הנדרשת – תהא המידה הנדרשת אשר תהא.

זוהי תמונת המראה האפלה עוד יותר של מעגל המוות המאושר – מעגל השנאה. כל מי שקורא לתקוף את האויב נחשב פטריוט, וכל מי שמנסה להטיל ספק ולו בטענה שלילית אחת לגבי האויב נחשב לבוגד. אבל בדיוק כפי שמרבית ההצהרות המורכבות הן שגויות, כך גם מרבית הדברים השליליים שניתן לומר על מישהו – אפילו על האדם הגרוע בעולם – פשוט אינם נכונים.

לדעתי, ההמחשה הטובה ביותר לכך הייתה הקביעה ש״המחבלים המתאבדים היו פחדנים.״ מעט שכל ישר, ברשותכם? נדרש לא מעט אומץ לרסק מטוס לתוך בניין מתוך בחירה. אפשר להאשים אותם בדברים רבים, אך פחדנות אינה אחד מהם. אבל אני מניח שכל דבר רע שייאמר על מחבל, תהא מידת האיוולת שבו אשר תהא, ייחשב אוטומטית לנכון. האם הייתי זוכה בנקודות זכות רבות עוד יותר אילו האשמתי את אל־קאעידה בהתנקשות בג׳ון קנדי? או בסטליניזם? *פחדנות*, באמת?

כן, ישנה חשיבות לעובדה שחוטפי ה-11 בספטמבר לא היו פחדנים. לא רק לשם הבנה נכונה של פסיכולוגיית האויב. מעגלי שנאה פשוט מחוללים נזק רב מדי. מסוכן מדי שתתקיים בעולם מטרה כלשהי – בין אם מדובר ביהודים ובין אם באדולף היטלר – שכלפיה *השמעת דברים שליליים* מקבלת עדיפות *על פני דיוק בעובדות*.

כאשר כוחות ההגנה כוללים אלפי כלי טיס ומאות אלפי חיילים חמושים היטב, כדאי להביא בחשבון ש״המערכת החיסונית״ עצמה מסוגלת לזרוע הרס רב יותר מתשעה־עשר אנשים וארבעה מטוסים אזרחיים. ארצות הברית שילמה במיליארדי דולרים ובחייהם של

אלפי חיילים, רק כדי לירות לעצמה ברגל באפקטיביות רבה מזו שכל ארגון טרור יכול היה לחלום עליה.

אילו התעלמה ארצות הברית לחלוטין ממתקפת ה-11 בספטמבר – אילו פשוט הייתה מושכת בכתפיה ובונה מחדש את המגדלים – התוצאה הייתה עדיפה על פני המסלול ההיסטורי שהתפתח בפועל. אך זו לא הייתה אפשרות פוליטית. גם אם מישהו שיער בינו לבין עצמו שהתגובה החיסונית תהיה הרסנית יותר מן המחלה, לפוליטיקאים האמריקאים לא הייתה כל ברירה – אם חפצו לשמור על הקריירה שלהם – אלא לצעוד היישר לתוך המלכודת של אל־קאעידה. כל מי שטוען בזכות תגובה חריפה יותר נחשב פטריוט. כל מי בוחן טענה פטריוטית בעין ביקורתית נחשב לבוגד.

בימים הראשונים שלאחר ה-11 בספטמבר, נשמעו גם תגובות חכמות מכפי ששיערתי. ראיתי חבר קונגרס – שכחתי את שמו – אומר מול המצלמות: "שכחנו שתפקידה הראשון של ממשלה אינו הכלכלה, אינו שירותי הבריאות, אלא הגנה על המדינה מפני התקפות." נדהמתי מכך שפוליטיקאי הצליח לומר דבר־מה שאינו סתם נורת מחיאות כפיים. ההלם הרגשי ודאי היה עצום, אם הוא גרם לחבר קונגרס לומר דבר כה... אמיתי.

אך בתוך יומיים התפוגג ההלם האותנטי, והדאגה לתדמית השתלטה מחדש על השיח הפוליטי. בשלב זה, מעגל ההסלמה נטל את המושכות לידיו במלואן. מרגע שהקריאה לאיפוק הופכת לדבר שאסור לבטא, אין זה משנה מאיזו נקודה מתחיל השיח – מפלס הזעם והאיוולת יכול רק לעלות עם הזמן.

111

ניסוי מערת השודדים

האם תהיתם אי פעם בילדותכם שמא מאחורי מחנה הקיץ התמים שלכם מסתתרת מטרה חלופית ומורכבת – למשל, שכל המחנה הוא ניסוי מדעי ושה"מדריכים" הם למעשה חוקרים שעוקבים אחר התנהגותכם?

גם אני לא.

אך ייתכן שהיינו מעט חשדנים יותר אילו קראנו את המחקר "קונפליקט ושיתוף פעולה בין־קבוצתי: ניסוי מערת השודדים", מאת שֶׁריף, הארווי, וייט, הוד ושֶׁריף.[1]* במחקר זה, נבדקי הניסוי – סליחה, "החניכים" – היו 22 בנים, תלמידי כיתות ה' ו-ו', שנבחרו מ-22 בתי ספר שונים באוקלהומה סיטי. כולם הגיעו ממשפחות פרוטסטנטיות יציבות ממעמד הביניים והיו תלמידים טובים, בעלי מנת־משכל חציונית של 112. הם היו נורמטיביים ודומים זה לזה ככל שהיה ביכולתם של החוקרים למצוא.

הניסוי, שנערך באווירת ההלם הכללי שלאחר מלחמת העולם השנייה, נועד לחקור את הגורמים לסכסוכים בין־קבוצתיים – ואת הפתרונות האפשריים להם. כיצד התכוונו החוקרים לעורר סכסוך כזה כדי לחקור אותו? ובכן, 22 הבנים חולקו לשתי קבוצות בנות 11 חניכים, ו...

* הניסוי נקרא כך על שם המקום בו הוא נערך – פארק מערת השודדים ("Robbers Cave State Park") באוקלהומה.

וזה התברר כדַי והותר.

על פי התכנון המקורי, הניסוי נועד לכלול שלושה שלבים. בשלב הראשון, כל קבוצת חניכים הייתה אמורה להתאקלם במחנה, ללא כל מודעות לקיומה של הקבוצה האחרת. לקראת סוף השלב הראשון, הקבוצות היו אמורות להתוודע בהדרגה זו לזו. בשלב השני, סדרה של תחרויות ופרסים נועדה לעמת את שתי הקבוצות זו מול זו.

אך החוקרים לא נדרשו לטרוח עם השלב השני. עוינות ניצתה כמעט מהרגע שכל קבוצה גילתה את דבר קיומה של רעותה: "הם משתמשים בשטח *שלנו*, במגרש הבייסבול *שלנו*." כבר במפגש הראשון ביניהן החלו שתי הקבוצות להטיח עלבונות. הם קראו לעצמם העֲיטים והעַכְסָנים (למרות שלא היו זקוקים לשמות לעצמם כשחשבו שהם היחידים בפארק).

כשהוכרזו התחרויות והפרסים, בהתאם לנוהלי הניסוי שנקבעו מראש, הגיעה היריבות לנקודת רתיחה. במהלך היומיים הראשונים לתחרויות עוד ניכרה רוח ספורטיבית, אך זו התפוררה במהירות.

העיטים גנבו את הדגל של העכסנים ושרפו אותו. בתגובה, העכסנים פשטו על הבקתה של העיטים, גנבו את מכנסי הג'ינס של ראש הקבוצה, צבעו אותם בכתום והניפו אותם למחרת כדגל שעליו נכתב "אחרון העיטים." העיטים יצאו לפשיטת תגמול, הפכו מיטות ופיזרו לכלוך. לאחר מכן הם חזרו לבקתתם, שם התבצרו והכינו כלי נשק (גרביים ממולאות באבנים) למקרה של פשיטת נגד. לאחר שהעיטים ניצחו בתחרות האחרונה שתוכננה לשלב השני, העכסנים פשטו על הבקתה שלהם וגנבו את הפרסים. העימות התפתח לתגרה, והצוות נאלץ להתערב מחשש לפגיעות בגוף. כשהעיטים סיפרו את הסיפור בינם לבין עצמם, הם הפכו את כל העניין לניצחון מפואר — וטענו שרדפו אחרי העכסנים "יותר מחצי הדרך חזרה לבקתה שלהם" (מה שלא קרה מעולם).

כל קבוצה פיתחה סטריאוטיפ שלילי של "הם", וכנגדו סטריאוטיפ חיובי של "אנחנו". העכסנים נהגו לקלל הרבה. העיטים, לאחר שניצחו באחד המשחקים, הגיעו למסקנה שניצחו בזכות התפילות שלהם,

ושלעכסנים "הגיע" להפסיד כי הם משתמשים כל הזמן במילים גסות. העיטים החליטו לחדול משימוש במילים גסות בעצמם. הם גם החליטו שמכיוון שהעכסנים מקללים ללא הרף, מוטב שלא לדבר איתם כלל. העיטים טיפחו לעצמם תדמית "מהוגנת ומוסרית", בעוד העכסנים פיתחו תדמית "קשוחה ומחוספסת".

חברי הקבוצות סתמו את אפם כשחברי הקבוצה השנייה חלפו על פניהם.

בשלב השלישי, ניסו החוקרים להפחית את החיכוך בין שתי הקבוצות. פגישות כשלעצמן (נוכחות באותו חלל ללא תחרות) לא הפחיתו את העוינות. גם חוויות משותפות – למשל, צפייה בזיקוקי יום העצמאות – לא הועילו. במקום זאת, האירוע התדרדר לתגרה שבה השליכו הנערים אוכל זה על זה.

רוצים לנחש מה *כן* עבד?

(זהירות, ספוילר...)

לבָּנים נמסר שייתכן שישנה הפסקת מים בכל המחנה עקב תקלה מסתורית במערכת המים – אולי מעשה ידיהם של וַנדליסטים (אויב חיצוני, אחד הטריקים הישנים ביותר בספר).

נאמר להם שארבע חוליות סריקה נדרשות כדי לבדוק את השטח שבין המחנה לבין מאגר המים. בתחילה, חוליות סריקה אלו הורכבו באופן שווה מחברים משתי הקבוצות. כל החוליות היו אמורות להיפגש על יד מיכל המים אם לא ימָצאו דבר. מכיוון שאכן לא מצאו דבר, הקבוצות נפגשו ונוכחו בעצמן שהמים לא יוצאים מהברז. שתי קבוצות הבנים התייעצו היכן עשויה להיות הבעיה, נקשו על דפנות המיכל, גילו סולם המטפס מעלה, ראו שהמכל מלא, ולבסוף

מצאו את השק שהיה דחוס מאחורי הברז. כל הבנים התגודדו סביב הברז כדי לשחרר את הסתימה. חברים משתי הקבוצות העלו הצעות לפתרון הבעיה, ובנים משני הצדדים ניסו ליישמן.

כשהברז שוחרר לבסוף, העכסנים, שהיו להם מֵימיות, לא התנגדו לכך שהעיטים ישתו ראשונים (לעיטים לא היו מימיות). לא הוטחו כל עלבונות, אפילו לא ה״ליידיז פירסט״ המסורתי.

לא היה זה קיצה של היריבות. למחרת בבוקר התרחש קרב אוכל נוסף, מלווה בעלבונות. אולם כמה משימות משותפות נוספות, שדרשו שיתוף פעולה מצד שתי הקבוצות — כמו התנעת משאית תקועה — עשו את העבודה. בסוף הטיול, העכסנים השתמשו ב-5 דולרים שבהם זכו באחת התחרויות כדי לקנות מילקשייקים לכל הבנים בשתי הקבוצות.

ניסוי מערת השודדים ממחיש את הפסיכולוגיה של חבורות ציידים־לקטים, שמהדהדת עד ימינו, באופן המושלם ביותר שנוצר אי־פעם במדעי החברה.

כל דמיון לפוליטיקה מודרנית הוא פרי דמיונכם בלבד.

(לעיתים אני חושב שהדבר השני שהאנושות זקוקה לו יותר מכל הוא נבל־על. אולי אכנס לתחום עיסוק זה לאחר שאסיים את עבודתי הנוכחית.)

---

1. Muzafer Sherif et al., “Study of Positive and Negative Intergroup Attitudes Between Experimentally Produced Groups: Robbers Cave Study,” Unpublished manuscript (1954).

112

כל מטרה שואפת להפוך לכת

קֵייד מֶץ מכתב העת *The Register* טען לאחרונה כי ברשימת תפוצה סודית של בכירי מנהלי הקהילה בוויקיפדיה התפתחה אובססיה לחסימת כל מבקריו, האמיתיים והפוטנציאליים, של המיזם. בכלל זה נחסם גם משתמש פעיל במיוחד רק משום שאחד המנהלים השתכנע — אך ורק *בשל* אותה פעלתנות — שהמשתמש הוא מרגל שנשלח מטעם *Wikipedia Review*.* עוד נטען כי הבכירים בוויקיפדיה 'צופפו שורות' בכדי להגן על אנשיהם. (טרם חקרתי האשמות אלו בעצמי. תודה לאוּיֶגֶן לַייטֶל על ההפניה.)

האם קיים פגם מוסרי עמוק בניסיון לארגן את הידע העולמי, פגם שמעביר את הדוגלים במטרה זו על דעתם? אולי רק אנשים בעלי נטיות טוטליטריות מולדות ינסו להפוך לסמכות עולמית בכל נושא** —

זהירות: הטיית ייחוס לפניך! (הטיית ייחוס: הסקת מסקנות על תכונות האופי של אדם מתוך התנהגות שניתנת להסבר מלא באמצעות הנסיבות שבהן היא מתרחשת. כשאנו רואים אדם בועט במכונת משקאות אנו סבורים שהוא "טיפוס עצבני", אך כשאנחנו בועטים בה, הרי זה מפני שהאוטובוס איחר, הרכבת הקדימה והמכונה בלעה

---

* *Wikipedia Review* היה פורום שהוקדש לביקורת שלילית על ויקיפדיה ולמעקב אחר התנהלות בעייתית של מפעילי האתר.

** "השתלטות על העולם זה מונח כל־כך מכוער, אני מעדיף לקרוא לזה אופטימיזציית העולם." (הארי ג'יימס פוטר־אוונס־ורס)

את כספנו.) אם ההאשמות כלפי ויקיפדיה נכונות, הרי שהן מוסברות על ידי טבע אנושי רגיל *לחלוטין*, ולא באמצעות טבע אנושי יוצא *דופן*.

הדיכוטומיה של קבוצת פנים מול קבוצת חוץ ("אנחנו" או "הם") היא חלק מן הטבע האנושי הרגיל. כך גם מעגלי מוות מאושרים ומעגלי שנאה. *מטרה נעלה* אינה זקוקה לפגם נסתר ועמוק כדי שחסידיה ייצרו קבוצת פנים סגורה בעלת סממני כת. די בכך שהחסידים הם בני אדם. כל השאר מגיע באורח טבעי, ריקבון המתהווה כברירת מחדל – ממש כפי שמזון מתקלקל במקרר לאחר הפסקת חשמל.

כשם שכל הפרש טמפרטורות נוטה לאזן את עצמו, וכשם שכל תוכנת מחשב נוטה להפוך לערימה של קוד ספגטי, כך כל מטרה *שואפת* להפוך לכת. זהו מצב של אנטרופיה גבוהה שהמערכת נמשכת אליו באופן טבעי. זהו *מושך** בפסיכולוגיה האנושית. ייתכן שאין לכך כל קשר לשאלה האם המטרה אכן נעלה. ניתן היה לשער שמטרה טובה תקרין משהו מטוּבה על כל היבט בחייהם של האנשים הקשורים בה – ושחסידיה של אותה מטרה יהיו רגישים פחות למשחקי כבוד, להטיות קבוצתיות, למעגלים רגשיים ולהערצת מנהיגים. ואולם, אמונה ברעיון נכון אחד לא מכבה את אפקט ההילה. מטרה נעלה לא תהפוך את חסידיה לעל־אנושיים. יש אמנם שפע של רעיונות גרועים שמסוגלים לחולל נזק רב, אך זה לא בהכרח מה שמתרחש כאן.

כל קבוצת אנשים בעלת מטרה יוצאת דופן – טובה, רעה או סתמית – תיטה אל מושך־הכת אלא אם כן תשקיע מאמץ מתמיד להימנע מכך. אפשר לשמור על בית קריר יותר מסביבתו החיצונית, אך לשם כך יש להפעיל את המזגן ללא הרף. ברגע שמנתקים את זרם החשמל – כלומר, מוותרים על המאבק באנטרופיה – המצב ישוב ל"נורמליות".

באחד המקרים הבולטים, קבוצה מסוימת הפכה למעין כת, למרות שקריאת הקרב שלה הייתה "רציונליות! היגיון! מציאות

* ראו הערה בסוף פרק 106.

אובייקטיבית!" (ארחיב על כך בהמשך.) הדבקת התווית של "רציונליות" על הרעיון הגדול לא תגן עליכם, ממש כפי שתליית שלט המכריז "קר!" על ביתכם לא תקרר אותו. עליכם עדיין להפעיל את המזגן – להשקיע את האנרגיה הנדרשת כדי להתנגד להידרדרות הטבעית לעבר הכת. סגידה לרציונליות לא תהפוך אתכם לשפויים יותר, כשם שסגידה לכוח המשיכה לא תאפשר לכם לעוף. אי אפשר לדבר עם התרמודינמיקה, ואי אפשר להתפלל לתורת ההסתברות. ניתן *להשתמש* בהן, אך לא להצטרף אליהן כאל מועדון חברתי.

הנטייה לעבר הכת היא עניין כמותי, לא איכותי. השאלה אינה "כת – כן או לא?" אלא "באיזו מידה זו כת, ובאיזה מובן?" אפילו במדע, שהוא אב הטיפוס למטרה נעלה אמיתית, אפשר להצביע בקלות על חזיתות המאבק העכשווי נגד אנטרופיית הכת, שבהן קו החזית נע קדימה ואחורה. האם סביר יותר שכתבי עת יקבלו מאמרים מאת כותב מוכר, או מאת כותב לא־מוכר ממוסד מוכר, בהשוואה לכותב לא־מוכר ממוסד לא־מוכר? איזה חלק מהאמון במאמר נובע מן הסמכות, ואיזה חלק נובע מהניסוי עצמו? אילו כתבי עת מיישמים ביקורת עמיתים עיוורת, ועד כמה היא אכן אפקטיבית?

אני מביא את הדוגמה הזו – במקום אחת ההאשמות העמומות והשגורות בסגנון "מדענים אינם פתוחים לרעיונות חדשים" – משום שהיא מצביעה על *קו חזית* ברור. זהו מקום שבו הפסיכולוגיה האנושית נהדפת לאחור באופן פעיל, וממנו נשאבת החוצה אנטרופיית־הכת המצטברת. (כמובן, תהליך זה מחייב פליטה של מעט "חום שיורי".)

פרק זה אינו מהווה קטלוג של טכניקות למאבק אקטיבי נגד הנטייה להפוך לכַּת. את חלקן הזכרתי בפרקים קודמים, ואחרות אזכיר בהמשך. *בשלב זה* ברצוני רק להבהיר כי ערכה וחשיבותה של המטרה *אינם* פוטרים אתכם מהשקעת המאמץ הנדרש להתנגדות לאותו *מושך־כת*. יתרה מזאת, עצם היכולת להצביע על קווי חזית עכשוויים אינה בבחינת הודאה בכך שהמטרה הנעלה שלכם אינה ראויה. ייתכן שתסברו שאילו השאלה הייתה "כת – כן או לא?" הייתם מחויבים להשיב "לא" פן תבגדו במטרה האהובה עליכם. ואולם,

גישה זו שקולה לסיווג מנועים כ"יעילים לחלוטין" או "לא יעילים", במקום למדוד את הפחת והחום שהם פולטים.

מנגד, אם תאמינו שמה שהכשיל את אותן מטרות *אחרות ואווילית* הינו פגם כלשהו שהיה טבוע בהן מלכתחילה, אם תלעגו לאיוולתם של "קורבנות הכת", אם תחשבו שכתות מונהגות ומאוכלסות בידי מוטנטים – אזי לעולם לא תשקיעו את המאמץ הנדרש כדי להילחם באנטרופיה. כלומר, לא תשקיעו את המאמץ הנדרש כדי להתנגד לעצם היותכם בני אדם.

113

שומרי האמת

לעיתים קרובות מוטחת ברציונליסטים הביקורת הבאה: "גם האינקוויזיציה חשבה שכל האמת נמצאת אצלה! ברור שכל העסק הזה של 'אמת' הוא מסוכן."

קיים שפע של תשובות מתבקשות לכך, למשל: "אם אתם סבורים שהחזקה באמת מעניקה לכם רישיון לענות ולהרוג, הרי שהטעות אתם שוגים בה אינה קשורה כלל לאפיסטמולוגיה." או: "ובכן, ההצהרה ההיסטורית הזו שהשמעתם כרגע על אודות האינקוויזיציה – האם היא אמיתית?"

טיפשות הפוכה איננה חוכמה: "אם המחשב שלכם מפסיק לעבוד, אינכם יכולים להסיק מכך שכל הרכיבים שלו פגומים ושאתם חייבים מחשב חדש לגמרי בלי מעבד של AMD, בלי כרטיס מסך של ATI [...] למרות שבמחשב הישן שלכם היו את כל הדברים האלה. יכול להיות שאתם פשוט צריכים כבל חשמל חדש."* כדי להגיע למסקנה גרועה נדרש רק צעד שגוי אחד. אין הכרח שכל צעד בדרך יהיה לא נכון. האינקוויזיטורים האמינו ש- 2+2=4, אך לא זה היה מקור הטירוף שלהם. האם ייתכן שגם ריאליזם אפיסטמולוגי לא היה הבעיה המרכזית?

אכן נראה סביר שאילו האינקוויזיציה הייתה מורכבת

* ראו פרק 62.

מרלטיביסטים – שהיו מצהירים שדבר אינו אמיתי ודבר אינו משנה – היה להם קשה יותר לגייס התלהבות למלאכת העינויים שלהם. עם זאת, סביר שהם היו מגייסים התלהבות פחותה גם אילו היו עוברים לובוטומיה. לדעתי זוהי אנלוגיה הוגנת.

ובכל זאת... אני סבור שיחסה של האינקוויזיציה לאמת אכן מילא תפקיד מסוים. האינקוויזיציה האמינה שיש דבר כזה אמת, ושהיא חשובה. ובכן, כך חשב גם ריצ׳רד פיינמן. אך האינקוויזיטורים לא היו *מבקשי* האמת, הם היו *שומרי* האמת.

קראתי פעם טענה (שאיני מצליח למצוא את מקורה) לפיה מרכיב מרכזי *בצַייטגַייסט** הוא השאלה האם התרבות רואה את האידיאלים שלה בעתיד או בעבר. כמעט כל התרבויות שקדמו לעידן הנאורות האמינו ב״ירידת הדורות״ – האמונה שפעם, בעבר הרחוק, הכל היה מושלם, אך מאז שהכתה קטסטרופה הכל רק הולך ומידרדר:

> בעידן שבו החיים על פני האדמה היו מלאים... הם אהבו זה את זה ולא ידעו שזוהי ׳אהבת רֵעַ׳. הם לא רימו איש, ובכל זאת לא ידעו שהם ׳אנשים שניתן לסמוך עליהם׳. הם היו אמינים ולא ידעו שזהו ׳תום לב׳. הם חיו יחד בחופשיות, נותנים ולוקחים, ולא ידעו שהם נדיבים. משום כך עלילותיהם לא סופרו. הם לא עשו היסטוריה.
>
> *– דרכו של ג׳ֶאנג דְזֶה*[1]

העבר המושלם, על פי מיטב העדויות האנתרופולוגיות שבידינו, לא התקיים מעולם. אולם תרבות שרואה את החיים כצועדים באופן בלתי נמנע לקראת שקיעה, שונה בתכלית מתרבות הסבורה שניתן להעפיל לפסגות חסרות תקדים.

(אני משתמש במונח ״תרבות״ ולא ״חברה״, משום שבחברה יכולה להתקיים יותר מתת־תרבות אחת.)

אולי אפשר לטעון שההבדל בין ריצ׳רד פיינמן לאינקוויזיציה היה

---

* *צַייטגַייסט* (מגרמנית: zeitgeist, ״רוח הזמן״) הוא מונח המתאר את הלך הרוח, הרעיונות, והאקלים התרבותי והאינטלקטואלי המאפיינים תקופה מסוימת.

שהאינקוויזיציה האמינה שהאמת *כבר מצויה* בידיה, בעוד שפיינמן *חיפש* אחר האמת. אך טיעון זה אינו מחזיק מים, שהרי אין ספק שהיו מספר אמיתות שריצ׳רד פיינמן סבר *שמצויות* גם בידיו. ״השמיים כחולים״, לדוגמה, או ״2+2=4״.

כן, אכן יש למדע אמיתות ודאיות. תורת היחסות הכללית אולי תופרך על ידי פיזיקה עתידית כלשהי, אך בשום אופן לא בצורה שתנבא שהשמש מקיפה את צדק. על התיאוריה החדשה לאמץ את הניבויים המוצלחים של התיאוריה הישנה, לא לסתור אותם. בנוסף, תורת האבולוציה מתקיימת ברמת ארגון גבוהה מזו של אטומים, ואף גילוי עתידי על קוואַרקים לא ישליך לפח את הדרוויניזם, את תאוריית התא בביולוגיה, את התיאוריה האטומית בכימיה, או מאה תגליות מבריקות אחרות שאמיתותן מבוססת כיום מעבר לכל ספק *סביר*.

האם אלו ״אמיתות מוחלטות״? לא במובן שיש להן הסתברות של 1.0 מוחלט. אך אלו מקרים שבהם המדע אכן חושב שהאמת בידיו. ובכל זאת, מדענים אינם מענים אנשים שמטילים ספק בתיאוריה האטומית של הכימיה. מדוע לא? משום שהם אינם מאמינים שוודאות מהווה רישיון לעינויים? ובכן, כן, זהו ההבדל *על פני השטח*, אך *מדוע* מדענים אינם מאמינים בכך?

משום שהכימיה אינה מאיימת בעונש על־טבעי של עינויי נצח על מי שכופר בתיאוריה האטומית? אך כאן אנו נכנסים שוב לרקורסיה ושואלים: למה? *מדוע* כימאים לא מאמינים שתלכו לגיהנום אם לא תאמינו בתיאוריה האטומית?*

משום שכתבי עת לא יפרסמו את המאמר שלכם עד שתשיגו תצפית אמפירית מוצקה של הגיהנום? הרי מדענים רבים מסוגלים לדכא את הרפלקס הספקני שלהם כרצונם. מדוע לכימאים אין איזושהי כת סודית שטוענת שמי שאינו כימאי ילך לגיהנום, בהתחשב בכך שרבים מהם ממילא נוצרים?

לשאלות מעין אלו אין תשובות אלגנטיות הנשענות על גורם יחיד.

---

* הערת המתרגם: בתור כימאי אני לגמרי מאמין שמי שלא מאמין בתיאוריה האטומית הולך לגיהנום.

אולם אטען *שאחד* הגורמים קשור לאימוץ עמדה *יוצרת* כלפי האמת, בניגוד לעמדה *מתגוננת*.

כאשר אתם *שומרי האמת*, אין לכם שום דבר מועיל לתרום לאמת *מלבד* עצם השמירה עליה. לעומת זאת, כשאתם מנסים לזכות בפרס נובל לכימיה באמצעות גילוי הבֶּנזֶן או הפוֹלֶרֶן הבא, אדם המערער על התיאוריה האטומית מהווה פחות איום על תפיסת עולמכם ויותר מטרד המבזבז את זמנכם.

כשאתם *שומרי האמת*, כל שביכולתכם לעשות הוא לנסות לבלום את ההידרדרות הבלתי נמנעת לאנטרופיה באמצעות יירוט של כל מה שסוטה מהאמת. אילו הייתה דרך כלשהי לפעול להקטנת האנטרופיה – לייצר אמונות אמיתיות חדשות יחד עם מעט חום שיורי – מנגנון שכזה היה יכול להותיר את האמת בחיים ללא משטרה חשאית. בכימיה, תוכלו לשחזר ניסויים ולחזות בהם במו עיניכם, וכך נשמרת האמת היקרה בחיים ללא צורך באלימות.

אין זה אֵיום נורא כל כך אם אנו שוגים בטעות אחת פה ושם ומוצאים עצמנו מאמינים בטעות קטנה למשך זמן קצר, משום *שמחר* נוכל להדביק את הפער.

אך כל זה עובד רק משום שהשיטה הניסויית היא קריטריון של איכות, ולא קריטריון של השוואה למקור סמכות. מכיוון שניסויים יכולים לשחזר את האמת ללא צורך בסמכות, הם יכולים גם *לעקוף* את הסמכות ולייצר אמונות אמיתיות חדשות במקום שבו לא היו כאלה קודם לכן.

במקום שבו קיימים קריטריונים של איכות שאינם מתמצים בקריטריונים של השוואה, יכולים להתקיים *שינויים* המהווים *שיפור* ולא *איום*. במקום שבו ישנם *רק* קריטריונים של השוואה – שבו אין כל דרך לעקוף את הסמכות – אין גם דרך ליישב חילוקי דעות בין מקורות סמכות שונים, פרט להשמדה. הטנקים יכריעו.

אני לא מתכוון לתאר כאן איזושהי תפיסת־על היסטורית גורפת המבוססת על גורם יחיד. אולם בכוונתי להצביע על הבדל פסיכולוגי עמוק בין תפיסת המטרה הגדולה של חייכם *כהגנה, שמירה ושימור*,

לבין תפיסתה *כגילוי, יצירה ושיפור*. האם "למעלה" על חץ הזמן מצביע אל העבר או אל העתיד? זוהי הבחנה שמקרינה על הכל ושולחת זרועות לכל עבר.

זו הסיבה שתמיד התעקשתי, למשל, שאם בכוונתכם להתחיל לדבר על "אתיקה של בינה מלאכותית", כדאי שתדברו על האופן שבו אתם עתידים *לשפר* את המצב הקיים באמצעות בינה מלאכותית, ולא רק על הדרכים למנוע מדברים מסוימים להשתבש. ברגע שאתם מאמצים קריטריונים של השוואה גרידא, אתם מתחילים לאבד קשר עם האידיאלים שלכם – אתם מאבדים את ההבחנה בין טוב לרע, ומתחילים לראות רק "שונה" או "זהה".

אני גם טוען שההבדל הפסיכולוגי הבסיסי הזה הוא אחת הסיבות לכך שתחום אקדמי שחדל מלהתקדם באופן פעיל נוטה להפוך *למרושע*. (לפחות בסטנדרטים המעודנים של המדע. רצח *אופי* הוא משחק ילדים בסטנדרטים היסטוריים, שהרי רוב מערכות האמונה בעלות העמדה המתגוננת הלכו על הדבר האמיתי.) אם זעזועים משמעותיים אינם מתרחשים לעיתים קרובות מספיק כדי לקדם חוקרים צעירים על בסיס כישרון ולא על בסיס קונפורמיזם, התחום חדל להתנגד להידרדרות הסטנדרטית אל עבר סמכותנות. כשלא מתגלות תגליות רבות, לא נותר מה לעשות כל היום מלבד לנהל ציד מכשפות נגד הכופרים.

כדי להפיק את המרב מהיתרונות לבריאות הנפש שמציעה עמדת ה"גילוי/יצירה/שיפור", עליכם לייצר *התקדמות בפועל*, ולא רק לקוות שהיא תגיע.

---

1. Zhuangzi and Thomas Merton, *The Way of Chuang Tzu* (New Directions Publishing, 1965).

114

שומרי הגנום

כמו כל אדם משכיל בן המאה ה-21, ודאי שמעתם על מלחמת העולם השנייה. אולי אתם זוכרים שהיטלר והנאצים תכננו לקדם תפיסה רומנטית של אבולוציה, במטרה להשביח גזע אדונים חדש – אדם־על חזק וחכם יותר מכל מה שהיה קיים אי־פעם.

למעשה, זוהי טעות נפוצה. היטלר האמין שאדם־העל האָרי *כבר התקיים בעבר* – הסטריאוטיפ הנורדי, חיית הטרף זהובת השיער ותכולת העיניים – אך דמו *זוהם* בעקבות התבוללות עם גזעים טהורים פחות. מבחינתם של הנאצים, התחוללה כאן "ירידת הדורות" של הגזע.

העובדה שברגע שאנו שומעים על תוכנית השבחת־הגזע הנאצית עולה בנו מיד המחשבה "הם ניסו ליצור אדם־על," מלמדת משהו על המידה שבה תפיסת *הקידמה* חלחלה לתרבות המערבית. אתם, קוראים יקרים – אילו איבדתם את המצפן המוסרי שלכם עד כדי כך שתתמכו בהשבחת־גזע כפויה, הייתם מנסים ליצור אדם־על. זאת משום שאתם ממקמים את האידיאלים שלכם בעתיד, ולא בעבר. משום שאתם אנשי *יצירה*. המחשבה לבצע "הרבעה לאחור" אל עבר איזה ארכיטיפ נורדי מלפני אלף שנה כלל לא הייתה עולה על דעתכם בתור אפשרות – מה, רק *הוויקינגים*? זה *הכל*? אילו הידרדרתם לשפל שבו אתם מוכנים להרוג למען המטרה, הייתם מנסים, לעזאזל, להגיע לפסגות שמעולם לא נכבשו. אחרת, איזה בזבוז זה יהיה, לא? ובכן,

זו אחת הסיבות לכך שאינכם נאצים, קוראים יקרים.

העובדה שכאשר מסַפרים לנו על הנאצים אנו נוטים לעוות את הסיפור כדי לשוות להם דמות של טרנס־הומניסטים פגומים, ממחישה עד כמה קשה לאנשים בריאים יחסית להיכנס לנעליהם של אנשים חולים יחסית.

הקומוניסטים הם אלה שהיו הטרנס־הומניסטים הפגומים — "האדם הסובייטי החדש" וכדומה. הנאצים היו, ללא ספק, הביו־שמרנים בסיפור הזה.

115

שומרי איין ראנד

״עבור ספקנים, הרעיון שרציונליות עשויה להוביל לכת הוא אבסורדי. מאפייניה של כת הפוכים ב-180 מעלות לאלה של רציונליות. אך כפי שאַראה, לא רק שדבר כזה יכול לקרות, הוא אכן קרה – ודווקא לקבוצה שמן הראוי היה שתיחשב לכת הכי פחות סבירה בהיסטוריה. זהו שיעור על מה שמתרחש כשהאמת הופכת להיות חשובה יותר מהחיפוש אחר האמת...״

– מייקל שֶׁרמֶר, ״הכת הכי פחות סבירה בהיסטוריה״[1]

לדעתי, מייקל שרמר מסבך יתר על המידה את ההסבר שלו לאובייקטיביזם. ארחיב על כך בהמשך.

הרומנים של איין ראנד מעלים על נס את הטכנולוגיה, את הקפיטליזם, את התרסתו של היחיד נגד המערכת, וכן אידיאלים של ממשלה קטנה, קניין פרטי ואנוכיות. הגיבור הבדיוני האולטימטיבי שלה, ג׳ון גאלְט, היה (אזהרת ספוילר!) מדען שהמציא סוג חדש של אנרגיה מתחדשת וזולה, אולם הוא סירב להעניק אותה לעולם משום שהרווחים פשוט ייגנבו כדי לתחזק ממשלות מושחתות.

ואז – באורח פלא – הכל הפך ל״מערכת סגורה״ מוסרית ופילוסופית שאיין ראנד ניצבת במרכזה. המונח ״מערכת סגורה״ אינו האשמה מצידי: זהו המונח שמכון איין ראנד עצמו משתמש בו לתיאור האובייקטיביזם.

האובייקטיביזם מוגדר על ידי כתביה של איין ראנד. כעת, משראנד אינה בין החיים, האובייקטיביזם סגור. אם אתם חולקים על כתביה של ראנד מכל בחינה שהיא, אינכם יכולים להיחשב לאובייקטיביסטים.

מקס גַלַקמן אמר פעם: "מדע הוא כל תחום שבו הטיפש של הדור הנוכחי יכול להגיע רחוק יותר מהגאון של הדור הקודם." המדע מתקדם באמצעות שחיטת פרות קדושות, בדיוק כפי שאיינשטיין הדיח את ניוטון. כל פיזיקאי צעיר חולם להיות האלוף החדש שפיזיקאים עתידיים יחלמו להדיח מכיסאו.

האליל הפילוסופי של איין ראנד היה אריסטו. ייתכן שאריסטו היה כישרון מתמטי צעיר ומבטיח לפני 2,350 שנה, אך המתמטיקה רשמה התקדמות מסוימת מאז ימיו. תורת ההסתברות הבייסיאנית היא הלוגיקה הכמותית שהלוגיקה האיכותית של אריסטו מהווה רק מקרה פרטי שלה.* אולם, אין שום סימן לכך שאיין ראנד הכירה את תורת ההסתברות הבייסיאנית כשכתבה את יצירת המופת שלה, *מרד הנפילים*. ראנד כתבה על "רציונליות", אך לא טרחה להכיר את המחקר המודרני בתחום ההיוריסטיקות וההטיות הקוגניטיביות. כיצד יכול אדם לטעון שהוא רציונליסט מומחה, ועם זאת לא לדעת דבר על נושאים בסיסיים כל כך?

"רגע אחד," ימחה הקורא, "זה לא הוגן! *מרד הנפילים* פורסם ב-1957! כמעט אף אחד לא ידע על משפט בייס באותה עת." שטויות. תכף תאמרו לי שאיין ראנד מתה ב-1982, ולכן לא הייתה לה הזדמנות לקרוא את *שיפוט בתנאי אי־ודאות: היוריסטיקות והטיות*, שפורסם באותה שנה.

המדע אינו הוגן. זו בדיוק הנקודה. לרציונליסט מתחיל ב-2007 יש יתרון עצום על פני רציונליסט מתחיל ב-1957. כך אנו יודעים שהתרחשה התקדמות.

בעיניי, המחשבה לאמץ מרצון מערכת הכרוכה במפורש

* דהיינו, בעוד שהלוגיקה הקלאסית מכירה רק ב"אמת" או "שקר" מוחלטים (הסתברויות 1 או 0), ההסתברות הבייסיאנית עוסקת ברצף של דרגות ודאות בסביבה של אי-ודאות.

באמונותיו של אדם יחיד שכבר *הלך לעולמו*, נופלת אי־שם בין מגוחכת להתאבדותית. הרי מחשב הופך למיושן עוד בטרם מלאו לו חמש שנים.

החיוניות שראנד העריצה במדע, במסחר, בכל מסילת רכבת שהחליפה את נתיב הכרכרות, בכל גורד שחקים שנבנה בארכיטקטורה *חדשה* – כולה נובעת מהעיקרון של *התעלות על גדולי העבר*. כיצד יכול להתקיים מדע, אם המדען הבקי ביותר שיהיה אי־פעם כבר חי בעבר? מי היה מרים את קו הרקיע של ניו יורק שראנד כה העריצה, אם הבניין הגבוה ביותר שיתקיים אי פעם כבר נבנה?

ואף על פי כן, איין ראנד לא הכירה באף אדם שעלה עליה, בעבר או בעתיד לבוא. ראנד, שהחלה את דרכה בהערצת הרציונליות והאינדיבידואליות, סיימה בהחרמת כל מי שהעז לחלוק עליה. שרמר מצטט:

> [ברברה] ברנדן נזכרה בערב שבו חבר של ראנד ציין שהוא נהנה מהמוזיקה של ריכרד שטראוס. "כשהוא עזב בסוף הערב, איין אמרה, בתגובה שהפכה לאופיינית יותר ויותר, 'עכשיו אני מבינה למה הוא ואני לעולם לא נוכל להיות נפשות תאומות באמת. הפער בתחושת החיים שלנו גדול מדי.'" לעיתים קרובות היא לא המתינה עד שחבר יעזוב כדי להשמיע הערות כגון אלו.

איין ראנד השתנתה עם הזמן, יש לשער.

ראנד גדלה ברוסיה, וראתה במו עיניה את המהפכה הבולשביקית.

היא קיבלה אשרה לביקור קרובי משפחה באמריקה בגיל 21, ומעולם לא שבה. קל לשנוא סמכותנות כשאתה הקורבן. קל להיות אביר חירותו של היחיד כשאתה בעצמך מדוכא.

דרוש הרבה יותר חוסן נפשי כדי לחשוש מסמכות כשהכוח *נתון בידיים שלך*. כשאנשים נושאים אליך עיניים בציפייה לתשובות, קשה יותר לומר: "מה לכל הרוחות אני מבינה במוזיקה? אני סופרת, לא מלחינה," או "קשה לראות כיצד אהבת יצירה מוזיקלית יכולה להיחשב לדבר פסול."

כשאתה זה שרומס את אלה שמעזים לפגוע בך, איכשהו הפעלת הכוח נראית *מוצדקת* הרבה יותר מאשר כשאתה זה שנרמס. כל מיני תירוצים מצוינים קופצים לפתע לראש.

מייקל שרמר יורד לפרטים בנוגע לאופן שבו, לדעתו, הידרדרה בסופו של דבר הפילוסופיה של ראנד לסממנים של כת. ספציפית, שרמר טוען (כך נראה) שהאובייקטיביזם נכשל משום שראנד סברה שוודאות מוחלטת היא אפשרית, בעוד שהמדע לעולם אינו ודאי. איני יכול להסכים עם שרמר בנקודה זו. התיאוריה האטומית בכימיה היא ודאית למדי, ובכל זאת כימאים לא הפכו לכת.

למעשה, אני סבור ששרמר נופל קורבן להטיית הייחוס כשהוא מניח את קיומה של קורלציה ספציפית כלשהי בין הפילוסופיה של ראנד לבין האופן שבו חסידיה התגבשו לכדי כת. כל תנועה אידיאולוגית שואפת להיות כת.

איין ראנד נמלטה מברית המועצות, כתבה ספר על אינדיבידואליזם שאנשים רבים אהבו, זכתה לשפע של מחמאות, וגיבשה סביבה חוג מעריצים. מעריציה מצאו דברים מחמיאים יותר ויותר לומר עליה (מעגל המוות המאושר), והיא נהנתה מכך יותר מדי מכדי להשתיק אותם. היא מצאה עצמה אוחזת בכוח לרמוס את מי שלא נשא חן בעיניה, ולא עמדה בפיתוי.

איין ראנד ונתנאל ברֶנדֶן ניהלו רומן סודי מחוץ לנישואים. (ברשות בני הזוג של שניהם, עובדה שנושאת משקל רב בעיניי. אם ברצונכם להפוך זאת ל"בעיה", עליכם לציין שבני הזוג היו *אומללים* – וגם אז, הדבר אינו מעניינם של זרים.) כשהתגלה שברנדן "בגד" בראנד עם אישה נוספת, ראנד חטפה התקף זעם ונידתה אותו. אובייקטיביסטים רבים פרשו כשהחדשות על הרומן הפכו לנחלת הכלל.

מי נשאר עם ראנד, במקום ללכת אחרי ברנדן או לעזוב את האובייקטיביזם לחלוטין? תומכיה *האדוקים* ביותר. מי עזב? הקולות המתונים של העבר (בריחת המתונים, או "קירור באידוי"). ומאז, אחיזתה של ראנד בחוג שנותר סביבה הייתה מוחלטת, ולא הותרה שום הטלת ספק.

הדבר יוצא הדופן היחיד בכל הסיפור הזה הינו עד כמה הוא היה שגרתי.

אולי הייתם מצפים שמערכת אמונות המעלה על נס "תבונה", "רציונליות" ו"אינדיבידואליזם" תזכה לאיזושהי חסינות מיוחדת...? ובכן, היא לא.

הדבר הועיל בערך כפי שמועיל לתלות שלט "קר" על מקרר שאינו מחובר לחשמל.

המאמץ האקטיבי הנדרש כדי לפעול נגד האנטרופיה לא היה קיים שם, והריקבון פשה באופן בלתי נמנע.

אם אתם מכנים זאת "הכת הכי פחות סבירה בהיסטוריה," הרי שאתם פשוט מדביקים שמות גנאי למציאות.

מי ייתן ודבר זה יהווה לקח לכולנו: מהללי ה"רציונליות" אינם חסינים. אפילו אמירה כגון "עליכם להצדיק את אמונותיכם באמצעות התבונה, ולא באמצעות הסכמה עם המנהיג הדגול," פשוט מריצה תוכנה אוטומטית קטנה שלוקחת את כל דברי המנהיג הדגול ומייצרת הצדקה שחסידיכם האחרים יראו כהגיונית.

אם כן, היכן ניתן למצוא את אומנות הרציונליות האמיתית? בלימוד המתמטיקה של תורת ההסתברות ותורת ההחלטות. בשאיבת ידע מהמדעים הקוגניטיביים, כמו פסיכולוגיה אבולוציונית או חקר היוריסטיקות והטיות. בקריאת ספרי היסטוריה...

"למדו מדע, לא רק אותי!" — זו הייתה, ככל הנראה, העצה החשובה ביותר שאיין ראנד הייתה אמורה להשיא לחסידיה, ולא עשתה כן. אין אף אדם שחי אי פעם שכתפיו היו רחבות מספיק לשאת את מלוא משקלו של מדע אמיתי, שהוא פרי עמלם של רבים.

ראוי לציון, לדעתי, שגיבוריה הבדיוניים של איין ראנד היו אדריכלים ומהנדסים. ג'ון גאלט, הדמות האולטימטיבית שלה, היה פיזיקאי. ואולם, איין ראנד עצמה לא הייתה מדענית דגולה. עד כמה שידוע לי, היא גם לא הצטיינה במיוחד במתמטיקה. היא לא יכלה לשאוף להתחרות בגיבורים שלה עצמה. ייתכן שמשום כך החלה לאבד את הרצון להמשיך ולשפר את עצמה.

אני, כידוע לכם, מעריץ את התעוזה של פרנסיס בייקון, אך אני שומר על היכולת להודות בבושה מסוימת: "אילו יכולתי לחזור אחורה בזמן, ואיכשהו לגרום לפרנסיס בייקון להבין את הבעיה שאני עובד עליה כעת, עיניו היו קופצות מחוריהן כפקקי שמפניה ומתפוצצות."

אני מעריץ את הישגיו של ניוטון. אולם, גישתי כלפי זכות הבחירה לנשים מונעת ממני לקבל את ניוטון כמופת מוסרי. בדיוק כפי שהידע שלי בהסתברות בייסיאנית מונע ממני לראות בניוטון את המקור האולטימטיבי והחד־משמעי של ידע מתמטי. והידע שלי בתורת היחסות הפרטית – זעום ובלתי מנוצל ככל שיהיה – מונע ממני לראות בניוטון את הסמכות האולטימטיבית לפיזיקה.

באופן ריאלי, ניוטון לא היה מסוגל לגלות אף לא אחד מהרעיונות שבזכותם אני מתנשא עליו – *אך ההתקדמות אינה הוגנת! זוהי בדיוק הנקודה!*

למדע יש גיבורים, אך אין לו אלים. השמות הגדולים אינם נעלים עלינו, או אפילו יריבים שלנו. הם אבני דרך שחלפנו על פניהן במסלולנו. ואבן הדרך החשובה ביותר היא הגיבור שעוד יבוא.

להיות אבן דרך נוספת במסלולה של האנושות, זהו הדבר הטוב ביותר שניתן לומר על כל אדם. אולם נראה שעבור איין ראנד, תפקיד זה היה נחות מכדי לספק אותה. וכך היא הפכה לסתם עוד נביאה.

---

1. Michael Shermer, "The Unlikeliest Cult in History," *Skeptic* 2, no. 2 (1993): 74–81, http://www.2think.org/02_2_she.shtml.

116

שני קואנים על כת

רציונליסט מתלמד, חניכו של מאסטר אוּגִי, ננזף בידי חברו: "אתה מבלה את כל זמנך בהאזנה למאסטר שלך, ומדבר על 'רציונליות' פה ו'רציונליות' שם – נפלת לכת!"

המתלמד הוטרד מכך מאוד. המילים "נפלת לכת!" הדהדו באוזניו כששכב על יצועו באותו לילה, ואף הופיעו בחלומותיו.

למחרת ניגש המתלמד לאוגי, סיפר לו את שאירע ואמר: "מאסטר, אינני מפסיק להתייסר מהחשש שמא זו באמת כת, ושתורתך אינה אלא אמונה עיוורת."

אוגי השיב: "אם תמצא פטיש מונח בדרך ותבקש למכור אותו, תוכל לדרוש בעבורו מחיר נמוך או גבוה. אך אם תשמור את הפטיש ותשתמש בו כדי לקבוע מסמרים, מי יוכל להטיל ספק בערכו?"

המתלמד אמר: "אתה מבין, זה בדיוק מה שמדאיג אותי – תשובות הזֵן המסתוריות האלה."

אוגי אמר: "בסדר גמור. אם כך אדבר בצורה ברורה יותר, ואפרוש בפניך טיעונים הגיוניים לחלוטין שיוכיחו שלא נשאבת לכת. אבל קודם לכן עליך לחבוש את הכובע המטופש הזה."

אוגי הושיט למתלמד כובע בוקרים חום וענק.

"אה, מאסטר..." אמר המתלמד.

"כשאסביר לך הכל," אמר אוגי, "תבין מדוע זה היה נחוץ. לחלופין, תוכל להמשיך לשכב ער בלילות ולתהות האם זו כת."

המתלמד חבש את כובע הבוקרים.

אוגי אמר: "עד מתי תחזור על מילותיי ותתעלם ממשמעותן? מחשבות משובשות נוצרות כשאנו נקשרים רגשית למסקנות המועדפות עלינו. אתה חרד יתר על המידה לדימוי העצמי שלך כרציונליסט. באת אליי רק כדי שארגיע אותך. אילו היית סקרן באמת, ללא נטייה לכאן או לכאן, היית מחפש דרכים ליישב את ספקותיך. רק משום שהיית כה להוט ליישב את הדיסוננס הקוגניטיבי שלך, היית מוכן לחבוש כובע מטופש. אילו הייתי אדם רשע, יכולתי לגבות ממך מאה מטבעות כסף. כשאתה מתרכז בשאלה הנוגעת לעולם האמיתי, ערכה או חוסר ערכה של הבנתך מתבררים במהרה. אתה דומה ללוחם שלא מפסיק להגניב מבטים הצידה כדי לראות אם מישהו צוחק עליו —"

"*בסדר*," אמר המתלמד.

"אתה זה שביקש את הגרסה הארוכה," אמר אוגי.

מתלמד זה ירש לימים את אוגי, ונודע בשם נִי נוֹ טָאצִ'י. מאז ואילך לא התיר לתלמידיו לצטט את דבריו בוויכוחיהם, והיה אומר: "אל תדברו על טכניקות. השתמשו בהן."

רציונליסט מתלמד ניגש למאסטר אוגי ואמר: "מאסטר, אני חושש שדוג'ו הרציונליות שלנו... ובכן... קצת מזכיר כת."

"זוהי דאגה כבדת משקל," אמר אוגי.

המתלמד המתין זמן מה, אך אוגי לא הוסיף דבר.

אז פצה המתלמד שוב את פיו: "כלומר, סליחה, אבל החובה ללבוש את הגלימות האלה ואת הברדס — זה פשוט נראה כאילו אנחנו מסדר הבונים החופשיים או משהו."

"אה," אמר אוגי, "הגלימות והסממנים הטקסיים."

"ובכן, *כן*, הגלימות והסממנים," אמר המתלמד. "זה פשוט נראה

כל כך לא רציונלי.״

״אתייחס לכל החששות שלך,״ אמר המאסטר, ״אבל קודם עליך לחבוש את הכובע המטופש הזה.״ ואוגי שלף כובע קוסמים, רקום סהרונים וכוכבים.

המתלמד לקח את הכובע, הביט בו, ואז התפרץ בתסכול: ״*איך בדיוק זה יכול לעזור?*״

״מכיוון שאתה כל כך מוטרד מהקשר שבין ביגוד לתורת ההסתברות,״ אמר אוגי, ״אתה לא אמור להיות מופתע שעליך לחבוש כובע מיוחד כדי להבין.״

כשהמתלמד הגיע לדרגת סטודנט לתואר מתקדם, הוא אימץ את השם בּוֹזוֹ, והסכים לדון ברציונליות רק כשהוא לבוש בחליפת ליצן.

117

ניסוי הקונפורמיות של אָש

בסדרת ניסויים שנערכו במקור בשנות החמישים ושוחזרו מאז בהצלחה רבה, הפנה סולומון אָש את הזרקור לתופעה המוכרת כיום כ"קונפורמיות". בניסוי הקלאסי, מתבקש הנבדק להתבונן בחידה דוגמת זו שבתרשים המצורף: איזה מהקווים – A, B או C – זהה באורכו לקו X? קחו לכם רגע להחליט מהי התשובה שלכם...



העוקץ הוא שהנבדק יושב לצד כמה אנשים נוספים שמתבוננים בתרשים – אנשים המוצגים בפניו כנבדקים אחרים, אך בפועל הם משתפי פעולה של הנסיין. ה"נבדקים" האחרים בניסוי מצהירים, בזה אחר זה, שקו C נראה זהה באורכו לקו X. הנבדק האמיתי יושב אחד־לפני־אחרון בתור. לוּ הועמדו במצב זה, כמה אנשים לדעתכם היו אומרים "C" – ומספקים תשובה שגויה בעליל, אך כזו שמתיישרת עם תמימות הדעים של שאר הנבדקים? מה לדעתכם יהיה האחוז?

שלושה רבעים מהנבדקים בניסוי של אש ענו תשובה קונפורמית לפחות פעם אחת. שליש מהנבדקים הפגינו קונפורמיות ביותר

ממחצית מהפעמים.

ראיונות שנערכו לאחר הניסוי העלו שבעוד שרוב הנבדקים טענו שלא באמת האמינו לתשובות הקונפורמיות שלהם, חלקם אמרו שבאמת ובתמים חשבו שהאופציה הקונפורמית היא הנכונה.

אש היה מוטרד מן התוצאות הללו:

> "העובדה שמצאנו נטייה כה חזקה לקונפורמיות בחברה שלנו... מדאיגה. היא מעלה שאלות באשר לדרכי החינוך שלנו ולגבי הערכים המנחים את התנהגותנו."[1]

השאלה האם הנבדקים בניסויים של אש התנהגו בצורה *לא רציונלית*, אינה שאלה טריוויאלית. "משפט ההסכמה" של ישראל אומן מראה שבייסיאנים כנים אינם יכולים "להסכים שלא להסכים" – אם יש להם ידע משותף לגבי הערכות ההסתברות שלהם, הם יחזיקו באותה הערכת הסתברות. משפט ההסכמה של אומן הוכח למעלה מעשרים שנה לאחר הניסויים של אש, אך הוא רק מנסח ומחזק נקודה שמובנת מאליה מבחינה אינטואיטיבית: האמונות של אנשים אחרים מהוות לעיתים קרובות ראיה לגיטימית.

לו התבוננתם בתרשים כמו זה שלמעלה, והייתם יודעים *בוודאות* ששאר האנשים בניסוי כנים ורואים את אותו התרשים כמוכם, ושלושה אנשים אחרים היו אומרים ש-C זהה באורכו ל-X, מה היו הסיכויים *שדווקא אתם* אלה שצודקים? איני מייחס לעצמי שום יתרון בחשיבה *חזותית* – איני חושב שאני טוב יותר מהאדם הממוצע בהשוואת אורכם של שני קווים. במונחים של רציונליות אישית, אני מקווה שהייתי מבחין בבלבול החמור שלי עצמי, ואז מעניק הסתברות של יותר מ-50% להצבעת הרוב.

במונחים של רציונליות קבוצתית, נראה לי שהדבר הראוי שרציונליסט כֵּן צריך לומר הוא: "כמה מפתיע, לי נראה ש-B זהה באורכו ל-X. אבל אם כולנו מתבוננים באותו תרשים ואומרים את דעתנו בכנות, אין לי שום סיבה להניח שההערכה שלי טובה משלכם." המשפט האחרון הוא קריטי – זוהי טענת אי־הסכמה חלשה בהרבה

מאשר, "הו, *אני* מבין את האשליה האופטית. אני רואה מדוע אתם חושבים שזה C, כמובן, אבל התשובה האמיתית היא B."

לפיכך, הנבדקים הקונפורמיים בניסויים הללו אינם מורשעים *אוטומטית* בחוסר רציונליות, לפחות על בסיס מה שתיארתי עד כה. אולם כפי שניתן לצפות, השטן טמון בפרטים הקטנים של תוצאות הניסוי. על פי מטא־אנליזה של למעלה ממאה שחזורים, שבוצעה על ידי סמית׳ ובונד:[2]

הקונפורמיות עולה בחדות בנוכחות של עד 3 משתפי פעולה, אך אינה ממשיכה לעלות עד ל-10-15 משתפי פעולה. אם אנשים מפגינים קונפורמיות מתוך רציונליות, הרי שדעתם של 15 נבדקים אחרים אמורה להוות ראיה חזקה משמעותית מדעתם של 3 נבדקים.

הוספת מתנגד יחיד – רק אדם אחד נוסף שעונה את התשובה הנכונה, או אפילו תשובה שגויה ששונה מזו של הקבוצה – מפחיתה את הקונפורמיות בצורה חדה *מאוד*, לשיעור של 5-10% מהנבדקים. אם אתם מיישמים גרסה אינטואיטיבית כלשהי של משפט אומן – ומניחים שכאשר אדם אחד חולק על שלושה, השלושה כנראה צודקים – הרי שבאותה המידה אתם יכולים לצפות ששני אנשים יחלקו על שישה. (זה אינו נכון אוטומטית, אבל נכון בהינתן ששאר התנאים שווים.) מנגד, אם יש לכם עסק עם אנשים שחוששים להיות "חריגים", קל לראות כיצד אדם אחד ויחיד שמסכים עימם – או אפילו אדם אחד ויחיד שאינו מסכים עם הקבוצה – יפיג משמעותית את המתח שלהם.

כצפוי, נבדקים בגרסת הניסוי שכללה גם מתנגד אחד לקבוצה, לא סברו שאי־הקונפורמיות שלהם הושפעה מאותו מתנגד או התאפשרה בזכותו. בדומה ל-90% מהנהגים הסבורים שרמת הנהיגה שלהם היא ~~מעל הממוצע~~ בחציון העליון, ייתכן שחלקם צודקים בעניין זה, אך לא כולם. אנשים אינם מודעים לעצמם ביחס לגורמים המביאים אותם להסכים או שלא להסכים עם קבוצה, עובדה המכבידה על הניסיון לטעון שאלו הם ביטויים של רציונליות. למשל, בהנחה שאנשים בוחרים לשקר מתוך "רציונליות חברתית" רק כדי שלא להתבלט,

נראה שלפחות חלק מהנבדקים בתנאי 'המתנגד היחיד' אינם מודעים מראש לאסטרטגיה שהיו מפעילים אילו ניצבו בפני תמימות דעים.

כשהמתנגד היחיד חזר לפתע *להתיישר עם הקבוצה*, שיעורי הקונפורמיות של הנבדקים זינקו חזרה בדיוק לאותה רמה כמו בגרסת הניסוי עם תמימות הדעים. להיות המתנגד הראשון זהו שירות חברתי בעל ערך (ויקר !), אך חובה להתמיד בו.

באופן עקבי, הן במסגרת כל ניסוי בנפרד והן בהשוואה בין ניסויים שונים, קבוצות הכוללות נשים בלבד (נבדקת אישה לצד משתפות פעולה) הפגינו קונפורמיות לעיתים קרובות יותר באופן משמעותי לעומת קבוצות הכוללות גברים בלבד. כמחצית מהנשים הפגינו קונפורמיות ביותר ממחצית מהפעמים, לעומת שליש מהגברים. אם תטענו שהנבדק הממוצע הוא רציונלי, המשמעות היא שנשים נוטות לרַצות יותר מדי וגברים נוטים להתנגד יותר מדי, כך ששניהם אינם *רציונליים* באמת...

מניפולציות של השתייכות קבוצתית (למשל, נבדק בעל מוגבלות לצד משתפי פעולה בעלי מוגבלויות) מראות, בדומה לכך, שהקונפורמיות גבוהה משמעותית בקרב חברים השייכים לאותה קבוצת השתייכות.

הקונפורמיות נמוכה יותר במקרה של תרשימים מובהקים, כמו זה שבתחילת הפרק, לעומת תרשימים שבהם השגיאות מעודנות יותר. קשה להסביר זאת אם (כל) הנבדקים מקבלים החלטה רציונלית־חברתית להימנע מלהתבלט.

פול קרולי מזכיר לי לציין שכאשר הנבדקים יכולים לענות באופן שאינו גלוי לחברי הקבוצה, הקונפורמיות צונחת גם כן, מה שמהווה טיעון נוסף נגד הסברת תוצאות הניסוי על פי משפט אומן.

---

1. Solomon E. Asch, "Studies of Independence and Conformity: A Minority of One Against a Unanimous Majority," *Psychological Monographs* 70 (1956).

2. Rod Bond and Peter B. Smith, "Culture and Conformity: A Meta-Analysis of Studies Using Asch's (1952b, 1956) Line Judgment Task," *Psychological Bulletin* 119 (1996): 111–137.

118

על הבעת חששות

הדבר המפחיד בניסויי הקונפורמיות של אָש הוא שניתן לגרום לאנשים רבים לומר ששחור הוא לבן, רק מעצם הושבתם בחדר עם אנשים אחרים שאומרים זאת. הדבר המעודד בניסויים אלו הוא שמתנגד יחיד מפחית דרמטית את שיעור הקונפורמיות, אפילו אם אותו מתנגד רק סיפק תשובה שגויה אחרת. ואילו הדבר *המתסכל* הוא שאי־קונפורמיות אינה *נרכשת* במהלך הניסוי – כשהמתנגד היחיד חזר ליישר קו עם הקבוצה, שיעורי הקונפורמיות זינקו בחזרה.

להשמיע קול מתנגד עשוי להביא תועלת אמיתית לקבוצה, אך כפי שאנו יודעים, יש לכך גם מחיר. נוסף על כך, עליכם להתמיד בהתנגדות הזו. ויתרה מזאת, ייתכן שאתם בכלל טועים.

חוויתי לאחרונה מקרה מעניין כששוחחתי על פרויקט עם שני אנשים שכבר הספיקו לגבש לגביו תכנון ראשוני בכוחות עצמם. סברתי שהתוכנית שלהם אופטימית מדי, והעליתי כמה הצעות בנוגע למרווחי ביטחון. עד מהרה חלף במקום אדם רביעי, שהיה אמור לתת טרמפ הביתה לאחד משני האחרים, והחל להעלות הצעות אף הוא. בשלב זה הכתה בי הבנה פתאומית באשר לאופן שבו קבוצות מפתחות ביטחון מופרז: בכל פעם שהעליתי בעיה אפשרית, האדם הרביעי מיהר לומר, "אל תדאגו, אני בטוח שנוכל להתמודד עם זה!" או משפט מרגיע דומה.

אדם שעובד לבדו יחווה ספקות טבעיים. הוא יחשוב לעצמו,

"האם אני באמת יכול לעשות את א' ב' ג'?" משום שאין כל חוסר נימוס בהטלת ספק בכישוריך *שלך*. אולם, כששני אנשים חסרי ביטחון מרכיבים קבוצה, נחשב למנומס לומר דברים חיוביים ומרגיעים, וזהו חוסר נימוס לפקפק בכישוריו של האחר. יחד הם נעשים אופטימיים יותר מכפי שכל אחד מהם היה לבדו. ספקותיו של כל אחד מהם שוככים על ידי מילות ההרגעה של רעהו, מלאות־הביטחון לכאורה, מבלי שיבינו כי גם לאדם השני היו בתחילה בדיוק את אותן הספקות.

האפשרות המפחידה ביותר שעולה מניסויי הקונפורמיות של אש היא הסכנה הטמונה במצב שבו כולם מסכימים עם הקבוצה, מושפעים מביטחונם של האחרים ונזהרים שלא לחשוף את ספקותיהם – מבלי להבין שגם שאר החברים מדחיקים חששות דומים. תופעה זו מוכרת בפסיכולוגיה כ"בורות פלורליסטית".

רובין הנסון ואני מנהלים ויכוח ארוך־שנים בשאלה מתי בדיוק רציונליסטים מתחילים צריכים להעז שלא להסכים. אני נוטה לעמדה הרווחת לפיה אין לך למעשה ברירה אלא לגבש את דעותיך בעצמך. רובין הנסון מצדד בעמדה חתרנית יותר, שלפיה *עליך* – ולא רק על אנשים אחרים – להביא בחשבון שייתכן שאחרים חכמים ממך. חרף המחלוקות השונות בינינו, על דבר אחד אנחנו מסכימים: לפי משפט ההסכמה של אומן, ידע משותף על מחלוקת עובדתית מוכיח *שמישהו* פה אינו רציונלי. על אף המבטים התמהים שאנו סופגים, אנו דבקים בעמדתנו ביחס לצניעות: תשכחו ממה שכולם אומרים לכם על אינדיבידואליזם, עליכם לייחס חשיבות למה שאנשים אחרים חושבים.

אהמ. הנקודה היא, שעבור רציונליסטים, אי־הסכמה עם הקבוצה היא עסק רציני. אי־אפשר פשוט לפטור זאת באמירה "לכל אדם יש זכות לדעה משלו."

אני סבור שהלקח החשוב ביותר שניתן להפיק מניסוייו של אש הוא ההבחנה בין "הבעת חששות" לבין "אי־הסכמה". העלאת נקודה שאחרים טרם השמיעו אינה בגדר הבטחה לחלוק על הקבוצה בתום הדיון שלה.

כדי שבייסיאנים אידיאליים יתכנסו לאותה המסקנה, על הצדדים

לשתף ראיות שהצד השני לא יכול היה לִצפות מראש. תוצאת ההסכמה של אומן חלה רק במצב של *ידע משותף* – כלומר, מצב שבו אתה יודע, אני יודע, אתה יודע שאני יודע, וכן הלאה. מאמרו של הנסון "We Can't Foresee to Disagree" ("איננו יכולים לחזות אי־הסכמה") ממחיש עד כמה מוזר יהיה לִצפות ברציונליסטים אידיאליים מתכנסים לעבר הערכת הסתברות אחידה. זה כלל לא נראה כמו שני סוחרים בשוק המתכנסים לעבר מחיר משותף.

לרוע המזל, מבחינה *חברתית* אין הבדל ניכר בין הבעת חששות לבין אי־הסכמה. קבוצה של רציונליסטים עשויה להסכים להעמיד פנים שקיים הבדל, אך בני אדם אינם בנויים כך באמת. ברגע שפתחתם את הפה, ביצעתם פעולה חברתית בלתי הפיכה: הפכתם למסמר הבולט, לצרימה בהרמוניה הקבוצתית הנוחה, ולא ניתן להשיב את הגלגל לאחור. אדם שנעלב מכך שהטלתם ספק ביכולתו לעמוד במשימה א׳ ב׳ ג׳, כנראה ימשיך לנטור לכם טינה גם אם לבסוף תיישרו קו עם הקבוצה.

הניסוי של אש מראה שכוחה של התנגדות לעורר השראה באחרים הוא כוח אמיתי, אך הוא מראה גם שכוחה של הקונפורמיות אמיתי באותה מידה. אם כולם יימנעו מהשמעת ספקותיהם האישיים, הדבר אכן עלול להביא קבוצות לכדי ניתוק מוחלט מהמציאות. אולם ההיסטוריה שופעת לקחים על המחיר הכרוך בלהיות הראשון, או אפילו השני, שאומר שהמלך הוא עירום. בני אדם אינם בנויים להבחין בין "הבעת חשש" לבין "אי־הסכמה אפילו בתנאים של ידע משותף". ההבחנה הזו אינה אלא התחכמות של רציונליסטים. אם תקראו את הז׳אנר הציני יותר של ספרי העזרה העצמית (כגון *הנסיך* של מַקיָאוֶולי), תגלו שהם מייעצים לכם להסוות לחלוטין את חוסר הקונפורמיות שלכם, *לא* להשמיע את חששותיכם בתחילה ולהסכים בסוף. אם תיקחו על עצמכם את תפקיד המתריע בשער ותצביעו על הבעיות הגלויות לכל, אל תצַפו שהקבוצה תודה לכם על כך.

אלו הם המחירים והיתרונות של אי־הסכמה, בין שאתם מתכוונים "לחלוק" על הקבוצה ובין שאתם רק רוצים "להביע חשש" – וההחלטה נתונה בידיכם.

119

התנגדות בודדת

ניסוי הקונפורמיות של אָש הראה שנוכחותו של מתנגד יחיד מפחיתה דרמטית את שכיחות מתן התשובות השגויות ה"קונפורמיות". הניסוי מלמד שאינדיבידואליזם הוא קל כל עוד יש לכם שותפים להתרסה. אם כל נבדק אחר בחדר, למעט אחד, אומר ששחור הוא לבן, ואתם הופכים לאדם השני שאומר ששחור הוא שחור – הרי שזו הרגשה נהדרת: שניכם, בודדים מתריסים ומורדים נגד העולם! (ראיונות המשך הראו שנבדקים בגרסת הניסוי שכללה מתנגד אחד בקבוצה, הביעו תחושות אחווה עזות כלפי אותו מתנגד – אף שברור שהם לא חשבו שנוכחותו היא שהשפיעה על חוסר־הקונפורמיות שלהם עצמם.)

אך אפשר *להצטרף* למרד רק אחרי שמישהו, אי־שם, הפך *לראשון* למרוד. מישהו צריך לומר ששחור הוא שחור אחרי ששמע את *כולם*, בזה אחר זה, אומרים ששחור הוא לבן. וזה – כפי שמראה הניסוי – קשה *הרבה יותר*.

התנגדות בודדת לא מרגישה כמו ללכת לבית הספר לבושים בשחור. היא מרגישה כמו ללכת לבית הספר לבושים בחליפת ליצן.

זהו ההבדל בין *הצטרפות למרד* לבין *עזיבת הלהקה*.

אם יש דבר אחד שאיני מסוגל לסבול, הרי זה זיוף ושקר־עצמי – אולי שמתם לב לכך. ובכן, התנגדות בודדת היא לבטח אחת התכונות המזויפות ביותר, ובאופן המוחצן ביותר שיש. כולם רוצים להיות

שוברי מוסכמות.

אין בכוונתי לזלזל בעצם ההצטרפות למרד. ישנן מרידות ששווה להצטרף אליהן. אכן נדרש אומץ כדי להתמודד עם גינוי מצד קבוצת השווים שלכם, או גרוע מכך, עם משיכות הכתפיים שלהם. מיותר לציין שהליכה להופעת רוק אינה מרד. אך צמחונות, למשל, כן נחשבת לכזה. אינני צמחוני בעצמי, אך אני רוחש כבוד לאנשים שכן, משום שאני מניח שנדרשת כמות בלתי מבוטלת של אומץ שקט כדי להגיד לאנשים שהמבורגר לא בא בחשבון לארוחת ערב. (אף שבאזור מפרץ סן־פרנסיסקו, אנשים שואלים זאת כדבר שבשגרה.)

ובכל זאת, אם תסַפרו לאנשים שאתם צמחונים, הם יחשבו שהם מבינים את מניעיכם (גם אם הם לא). הם עשויים שלא להסכים עמכם. הם עשויים להיעלב אם תכריזו על כך בגאווה רבה מדי, או לצורך העניין, להיעלב פשוט משום שהם נעלבים בקלות. אבל הם יודעים כיצד להתייחס אליכם.

כשמישהו מגיע לבית הספר לבוש בשחור, המורים והילדים האחרים מבינים את התפקיד החברתי שהוא נוטל על עצמו. הדבר נחשב כהצהרה שהוא נמצא "מחוץ למערכת" – בדרך הסטנדרטית שכולם מכירים ומבינים. כלומר, לא *באמת* מחוץ למערכת. זהו "אתגר כנגד החשיבה המקובלת," מהסוג המקובל – כך שאנשים יכולים לומר בזעף "אני לא מצליח להבין למה אתה –" אבל לא באמת נדרשים לחשוב שום דבר שלא חשבו קודם. כפי שאומרים: "האם 'ספרות חתרנית' כלשהי שקראת גרמה לך לשנות אף במקצת את דעותיך הפוליטיות?"

מה שדורש אומץ *אמיתי* הוא להתמודד עם *חוסר הבנה מוחלט* מצד האנשים שסביבכם, כשאתם עושים משהו *שלא* נכנס למשבצת של "מרד סטנדרטי מספר 37" – משהו שאין להם תסריט מוכן מראש עבורו. הם לא שונאים אתכם על עצם היותכם מורדים. הם פשוט חושבים שאתם ממש מוזרים, ומפנים לכם עורף. האפשרות הזו מעוררת פחד עמוק הרבה יותר. זהו ההבדל בין להסביר צמחונות

לבין להסביר קריוניקה.* אכן, אי־שם בעולם קיימים תומכי קריוניקה נוספים, אבל הם אינם שם לצידכם באותו רגע. עליכם להסביר זאת, לבדכם, לאנשים שפשוט סבורים שזה *מוזר*. לא אסור, אלא מחוץ לגבולות הדברים שאנשים בכלל מקדישים להם מחשבה. "אתה הולך להקפיא את הראש שלך? אתה חושב שזה ימנע ממך למות? למה אתה מתכוון, 'מידע מוחי'? הא? מה? אתה *משוגע*?"

אני מתפתה לנסח הסבר בדיעבד הלקוח מהפסיכולוגיה האבולוציונית: בימי אבותינו יכולתם לחבור לקבוצה קטנה של חברים ולעזוב יחד את שבט הציידים־לקטים שלכם, אך להיאלץ לצעוד *לבדכם* ביערות היה כנראה גזר דין מוות – לפחות מבחינה רבייתית. איננו חושבים על כך בצורה מודעת, אך לא זה טיבה של פסיכולוגיה אבולוציונית. להצטרף למרד מקובל זה מפחיד, אבל זה לא מתקרב לרמת הפחד שכרוכה בעשיית משהו שונה באמת. משהו שבימי אבותינו עלול היה להסתיים לא בפיצולו של השבט, אלא בגירושכם אל מחוץ למחנה, לבדכם.

כפי שממחישה הדוגמה של הקריוניקה, הפחד לחשוב בצורה שונה *באמת* חזק מהפחד מפני המוות. ציידים־לקטים נדרשו להיות מוכנים להתמודד עם מוות על בסיס יומי, כשיצאו לצוד יונקים גדולים, או סתם כששוטטו בעולם רוחש טורפים. הם נזקקו לאומץ הזה כדי לשרוד. אולם, אומץ להתריס כנגד דפוסי החשיבה הסטנדרטיים של השבט, ולהשתעשע במחשבות שנראות מוזרות באמת – ובכן, דפוס שכזה כנראה לא שירת את בעליו באותה המידה. איננו חושבים על כך בצורה מפורשת. לא כך עובדת הפסיכולוגיה האבולוציונית. אנו, בני האדם, פשוט בנויים בצורה כזו שרבים מאיתנו יעדיפו לצנוח צניחה חופשית מאשר להירשם לקריוניקה.

וזה אפילו אינו האומץ הגדול ביותר. יש יותר מתומך־קריוניקה אחד בעולם. רק רוברט אֶטינגֶר נאלץ להיות *הראשון* שאומר זאת.

---

* *קריוניקה* (cryonics) היא שימור בהקפאה עמוקה של בני אדם ובעלי חיים שקצרה ידה של הרפואה המודרנית מלהושיעם, בתקווה שריפויים והשבתם לחיים תתאפשר בעתיד. (ויקיפדיה)

כדי להיות מהפכן *מדעי*, עליך להיות האדם הראשון שסותר את מה שכל שאר האנשים שאתה מכיר חושבים. זהו אינו המסלול היחיד לגדולה מדעית, והדבר נדיר אפילו בקרב הגדולים מכולם. איש אינו יכול להפוך למהפכן מדעי מעצם הניסיון לחקות מהפכנות. ניתן להגיע לכך רק באמצעות חתירה לתשובה הנכונה בכל דבר, בין אם התשובה הנכונה מהפכנית ובין אם לאו. אך אם, בחלוף הזמן – אם, לאחר שספגתם את כל העוצמה והחוכמה של הידע שכבר נצבר – אם, אחרי כל זה ובתוספת מנה של מזל טהור, אתם מגלים שחתירתכם העיקשת לתשובות נכונות מביאה אתכם לטריטוריה חדשה... *אז* תינתן לאומץ שלכם ההזדמנות להכזיב.

זהו האומץ האמיתי של התנגדות בודדת, שכל להקת רוק בעולם מנסה לזייף.

כמובן שלא כל דבר שדורש אומץ הוא רעיון טוב. נדרש אומץ כדי לצעוד מעבר לקצה הצוק, אבל אז פשוט תתרסקו מטה.

הפחד מהתנגדות בודדת מהווה מכשול לרעיונות טובים, אך לא כל רעיון מהפכני הוא טוב – עיינו גם בחיבורו של רובין הנסון *Against Free Thinkers* ("נגד הוגים חופשיים"). רוב הקושי בניסוחה של מחשבה מדעית חדשה ואמיתית טמון בחלק של ה"אמיתית".

באמת שאין *צורך* להיות שונים אך ורק לשם השונות. אם תעשו דברים אחרת רק כאשר תראו לכך סיבה טובה בצורה מכרעת, עדיין יהיו לכם די והותר צרות שיספיקו לכל חייכם.

ישנן סביבנו כמה חבורות אמיתיות של שוברי מוסכמות. *כנסיית הסָאבּ־גֶ׳ינְיוּס*, למשל, באמת נראית כאילו מטרתה היא *לבלבל* אנשים מן השורה, ולא סתם לפגוע בהם.* וישנם איים של סובלנות אמיתית בעולם, כגון כנסי מדע בדיוני. *ישנם* אנשים מסוימים שאינם חוששים כלל מעזיבת העדר. מספרם של אנשים אלה בפועל קטן בהרבה

* כנסיית הסָאבּ־גֶ׳ינְיוּס (the Church of the SubGenius) היא דת פארודית שנוסדה בארצות הברית בסוף שנות ה־70, ולועגת לדתות מאורגנות ולמערכות אמונה פופולריות.

ממספר האנשים המדמיינים את עצמם כמורדים, אך הם קיימים. ועם זאת, מהפכנים מדעיים נדירים לאין שיעור. הקדישו לכך רגע של מחשבה.

עכשיו, באשר אליי, אתם יודעים, אני *באמת* שובר מוסכמות. כולם חושבים שהם כאלה, אבל במקרה שלי זה *נכון*, אתם מבינים. אני *לגמרי* הייתי מוכן ללבוש חליפת ליצן לבית הספר. השיחות הרציניות שלי נוהלו עם ספרים, לא עם ילדים אחרים.*

אך אם נדמה לכם שגם אתם הייתם, ללא ספק, לובשים את חליפת הליצן הזו, אל תהיו גאים בכך יותר מדי ! פירוש הדבר הוא שעליכם להתאמץ בכיוון ההפוך כדי להימנע מלהתנגד בקלות רבה מדי. זה מה שאני מוכרח לעשות כדי לתקן את נטייתי הטבעית. לאנשים אחרים יש סיבות לחשוב את מה שהם חושבים, ולהתעלם מהן לחלוטין זה גרוע בדיוק כמו לפחד לסתור אותם. לא הייתם רוצים לסיים בתור "הוגים חופשיים". זו אינה מעלה, אתם מבינים – זו פשוט הטיה, לכאן או לכאן.

* "אין לי מה לומר לילדים רגילים. מבוגרים לא מכבדים אותי מספיק כדי לדבר איתי באמת, ואם לומר את האמת, אפילו אם הם היו מכבדים אותי מספיק, הם לא היו נשמעים חכמים כמו ריצ'רד פיינמן, אז עדיף לי כבר לקרוא משהו שריצ'רד פיינמן כתב." (הארי ג'יימס פוטר־אוונס־ורס)

120

כת ה-"אנחנו לא כת"

בעולם המודרני, הצטרפות לכת היא ככל הנראה אחד הדברים הגרועים ביותר שיכולים לקרות לכם. בתרחיש הטוב ביותר, תמצאו את עצמכם בחברת אנשים כנים אך שוגים באשליות, שטועים בתום לב אך מלבד זאת מתנהגים כהלכה. תשקיעו זמן וכסף רבים, ולבסוף תצאו בידיים ריקות. למעשה, תיאור זה יכול להתאים לכל חברת סטרט־אפ כושלת בעמק הסיליקון. וזו לרוב חוויה קשה ומייסרת, אם חושבים על כך – אז כן, זה מפחיד מאוד.

כתות של ממש הן גרועות בהרבה. "הפצצת אהבה" כטכניקת גיוס, המכוונת לאנשים המצויים במשבר אישי. מניעת שינה. יצירת תשישות יזומה באמצעות עבודת כפיים מפרכת. קומונות מרוחקות במטרה לבודד את המגויס מחבריו וממשפחתו. פגישות יומיות לווידוי על מחשבות טמאות. אין זה נדיר שכתות שואבות את כל כספו של המגויס – את כל חסכונותיו בתוספת משכורתו החודשית – מה שהופך אותו לתלוי בכת למזון וביגוד. הרעבה כעונש על חוסר ציות. שטיפת מוח קיצונית ופגיעות חמורות.

בהתחשב בכל זאת, כנראה שהייתי אמור לגלות אהדה רבה יותר כלפי אנשים שחוששים מאוד מפני *הצטרפות אפשרית לכת*, כשהם נכנסים לאיזה מיזם שנראה קצת מוזר. הדבר לא אמור היה לעלות לי על העצבים. אך הוא כן.

נקודה ראשונה: "כתות" ו"לא־כתות" אינן סוגים טבעיים נפרדים

כמו כלבים וחתולים. אם תעיינו ברשימה כלשהי של מאפייני כתות, תמצאו סעיפים שיכולים בקלות לתאר מפלגות פוליטיות ותאגידים – ״עידוד חברי הקבוצה להתייחס בחשדנות לביקורת מבחוץ כאילו היא נובעת ממניעים נסתרים,״ או ״מבנה סמכות היררכי.״ כתבתי בפרקים הקודמים על כשלים התנהגותיים קבוצתיים כגון קיטוב קבוצתי, מעגלי מוות מאושרים, היעדר ביקורתיות ו״קירור באידוי״, שכולם נראים כניזונים אלו מאלו. כשהכשלים הללו מתערבלים ומשתלבים, הם חוברים יחדיו ליצירת *כשל־על* טיפשי יותר מכל אחד מרכיביו בנפרד (בדומה למגה־זורד). אך זו אינה *מהות* של כת – זהו *מושך** של כת.

כלבים נולדים עם דנ״א של כלב, וחתולים נולדים עם דנ״א של חתול. בעולם הנוכחי, אין מצבי ביניים. (אפילו בהנדסה גנטית, לא מדובר בעניין פשוט כמו יצירת אורגניזם שחצי מהגנים שלו הם של כלב וחצי מהם של חתול.) אין בנמצא אוסף של ״מאפייני כלב״ שמחזקים זה את זה, כך שחתול מסוים יכול לאמץ אותם עד כדי הפיכתו לחצי־כלב.

נראה שהמוח האנושי, בחשיבתו הקטגוריאלית, מעדיף *מהויות* על פני *מושכים*. בני אדם מבקשים לומר ״זוהי כת״ או ״זו אינה כת,״ ובכך לסיים את משימת הסיווג. אם אתם מבחינים שלסוקרטס יש עשר אצבעות, שהוא עוטה בגדים ומדבר יוונית שוטפת, אתם יכולים לומר ״סוקרטס הוא אדם,״ ומכאן להסיק ש״סוקרטס פגיע לרוש״** – מבלי שתידרשו לערוך לו בדיקות דם כדי לאמת זאת. הכרעתם בנוגע לאנושיותו של סוקרטס, וזה כל מה שדרוש.

אך אם אתם מבחינים שקבוצת אנשים מסוימת מפגינה קיטוב של ״אנחנו או הם״, ואתם מזהים אפקט הילה חיובי סביב הדבר האהוב עליהם מכל – בין אם מדובר באובייקטיביזם, בצמחונות או ברשתות

---

* ראו הערה בסוף פרק 106.

** *רוש עקוד* (poison hemlock) הוא צמח עשבוני רעיל ממשפחת הסוֹכְכִיים. ביוון העתיקה שימש להוצאה להורג באמצעות שיקוי שהופק ממנו. המקרה הידוע ביותר הוא מותו של סוקרטס.

נוירונים מלאכותיות – אינכם יכולים להסיק, *מן הראיות שנאספו עד כה*, האם הם הגיעו למצב של אי־ביקורתיות. אינכם יכולים להסיק אם הרעיון המרכזי שלהם נכון, שגוי, או אכן שימושי, גם אם לא במידה שהם סבורים. *מהמידע שנאסף עד כה* אינכם יכולים להסיק אם מלבד זאת הם אנשים מנומסים, או שמא הם יגררו אתכם לבידוד וימנעו מכם שינה ומזון. מאפייני כתות אינם בגדר הכל או לא כלום – הם אינם נוכחים כולם או נעדרים כולם.

אם תתבוננו בוויכוחים ברשת בסגנון "האם תופעה זו היא כת?" תראו שצד אחד עובר על רשימה אינטרנטית של מאפייני כת, מוצא מאפיין אחד שמתאים וקובע: "לכן זוהי כת!" מנגד, הצד המגן מוצא מאפיין שאינו מתאים ופוסק: "לכן זו אינה כת!"

אי אפשר לבנות תמונה מדויקת של הדינמיקה המחשבתית בקבוצה באמצעות מהותנות מסוג זה. עליכם לבחון את המאפיינים הבודדים, בזה אחר זה.

יתרה מזאת, טיפשות הפוכה איננה חוכמה. אם אתם מתעניינים *ברעיון* העיקרי, ולא רק בקבוצה המיישמת אותו, דעו כי גם לרעיונות חכמים עשויים להיות חסידים טיפשים. אנשי ניו־אייג' רבים מדברים על "פיזיקה קוונטית", אך הדבר אינו עומד לחובתה של פיזיקת הקוונטים. כמובן שגם לרעיונות טיפשיים עשויים להיות חסידים טיפשים. המהותנות הבינארית מביאה עימה את המחשבה שאם הסקתם שקבוצה מסוימת היא "כת", הרי שאמונותיה חייבות להיות שקריות – שהרי אמונות שקריות הן מאפיין של כתות, ממש כפי שפרווה היא מאפיין של חתולים. אם הרעיון הוא זה שמעניין אתכם, בחנו את הרעיון, לא את האנשים. 'כת' היא מאפיין של *קבוצות* יותר משל *השערות*.

הטעות השנייה היא שכאשר אנשים שואלים במתח: "זו אינה כת, נכון?" נשמע לי שהם מחפשים *אישור מחודש לרציונליות שלהם*. העיקרון שלפיו רציונליסט אינו אמור להיקשר יתר על המידה לדימויו העצמי כרציונליסט, ראוי לחיבור בפני עצמו (כתבתי עוד על כך

באינטרלוד *שתים־עשרה המעלות של הרציונליות*[*] ובפרקים 3 ו-116). אולם גם מבלי להיכנס לפרטים, ודאי שניתן להבחין בכך *שחיפוש מתוח אחר אישור חיצוני* איננו הלך הרוח המיטבי להערכת שאלות של רציונליות. אינכם סקרנים באמת, ואינכם חושבים על דרכים לפשוט את ספקותיכם. תחת זאת, תמצאו דף אינטרנט כלשהו המציין שכתות משתמשות במניעת שינה כדי לשלוט באנשים, תבחינו בכך שהקבוצה האהובה עליכם אינה משתמשת במניעת שינה, ותסיקו: "זו אינה כת. פיו!" אם אין לזה פרווה, זה בוודאי לא חתול. מרגיע מאוד.

ואולם, כל מטרה שואפת להפוך לכת – בין אם היא חכמה ובין אם היא טיפשית. הדיכוטומיה של "אנחנו או הם" וכיוצא בזה היא חלק מהטבע האנושי ולא קללה ייחודית של מוטנטים. הרציונליות היא היוצאת מן הכלל, לא הכלל. עליכם להשקיע מאמץ מתמיד כדי לשמר את הרציונליות נוכח הגלישה הטבעית לאנטרופיה. אם תחליטו ש"זו אינה כת!" ותנשמו לרווחה, הרי שלא תשקיעו את המאמץ המתמשך הדרוש להדיפת הנטיות *הרגילות* להפיכה לכת. תחליטו שמהות הכת נעדרת, ותחדלו מלפעול נגד האנטרופיה של *מושך־הכת*.

אם אתם חוששים מאוד מפני כיתתיות, תבקשו להכחיש כל רמז ולו הקל ביותר למאפיין המזכיר כת. אלא *שכל* קבוצה בעלת מטרה הנתפסת באור חיובי חשופה לסיכון של אפקט ההילה, וצריכה לפעול נגד האנטרופיה כדי להימנע ממעגל מוות מאושר. הדבר נכון אפילו לגבי מוסדות שגרתיים כמו מפלגות פוליטיות – למשל, אנשים הסבורים ש"ערכים ליברליים" או "ערכים שמרניים" מסוגלים לרפא סרטן. זה נכון לחברות סטרט־אפ בעמק הסיליקון, כושלות ומצליחות כאחת. זה נכון למשתמשי מק ולמשתמשי לינוקס. אפקט ההילה לא הופך ללגיטימי רק משום שכולם לוקים בו. לו כולם היו קופצים מצוק, לא הייתם קופצים גם אתם. יש להילחם בכשל המחשבתי, ולא להשלים עימו. אך אם אתם חוששים מדי ושואלים "האם אתם

* האינטרלוד מופיע בכרך 5 של *רציונליות: מ-א ועד ת*.

בטוחים שזו לא כת?", תסתייגו מזיהוי סימן *כלשהו* המאפיין כתות, שכן הדבר ירמז על כך שאתם חברים בכת – *והרי זו אינה כת!!* כפועל יוצא מכך, לא תבחינו בחזיתות המאבק הנוכחיות שבהן הנטיות *הרגילות* להפיכה לכת מזדחלות קדימה או נהדפות לאחור.

הטעות השלישית הגלומה בשאלה מלאת החשש "זו לא כת, נכון?" נעוצה בכך *שיש* לי חשד כבד *שהחשש* נובע מסיבות שגויות לחלוטין.

מדוע קבוצות שמהללות את "הדבר המשמח" שלהן עד השמיים, מעודדות את חבריהן לתרום את כל כספם ולעבוד בשעבוד מרצון, ומנהלות מתחמים פרטיים שבהם החברים מוחזקים בקפידה בבידוד, זוכות לכינוי "דתות" ולא "כתות" ברגע שקיומן נמשך כמה מאות שנים?

מדוע רוב האנשים השואלים בחשש על אודות קרִיוֹניקָה* "זו לא כת, נכון?" אינם חוששים באותה המידה להשתתף בעצרות פוליטיות של הימין או של השמאל? דיכוטומיות של "אנחנו או הם", כמו גם מעגלי מוות מאושרים, עלולות להתהוות בשיח הפוליטי, בדתות המרכזיות ובהערצת קבוצות ספורט. אילו *החשש* נבע מפחד מפני *שגיאות רציונליות*, אנשים היו שואלים "אין פה דיכוטומיה של 'אנחנו או הם', נכון?" גם לגבי עצרות פוליטיות של הימין או של השמאל – ובאותו טון מפוחד בדיוק.

ישנה סיבה לגיטימית לחשוש פחות מפני ליברטריאניזם מאשר מפני כת המאמינה בצלחות מעופפות, שכן לליברטריאנים לא יצא מוניטין של שימוש במניעת שינה כאמצעי השפעה. אולם גם לתומכי הקריוניקה אין מוניטין כזה. אם כן, מדוע יש לדאוג יותר מהקפאת הראש לאחר שהנשימה פוסקת?

אני חושד *שהחשש* לא נובע מהפחד להחזיק באמונות שקריות, או מהפחד מפגיעה פיזית. זהו הפחד מפני התנגדות בודדת. זו אותה תחושת חשש שחווים הנבדקים בניסוי הקונפורמיות של אָש כאשר

* ראו הערה בפרק 119.

כל שאר הנבדקים (למעשה משתפי פעולה) מצהירים בזה אחר זה שקו C זהה באורכו לקו X, בעוד שלנבדק נראה שדווקא קו B הוא בעל האורך הזהה.* זהו הפחד מעזיבת הלהקה.

זו הסיבה לכך שקבוצות שאמונותיהן שורדות זמן רב מספיק כדי להיראות "נורמליות" אינן מעוררות את אותה עצבנות שמעוררות "כתות" – אף שגם חלק מהדתות המרכזיות עשויות לקחת את כל כספכם ולשלוח אתכם למנזר. זו הסיבה שקבוצות המועדות מאוד לשגיאות רציונליות, כגון מפלגות פוליטיות, אינן מעוררות את אותו החשש כפי שמעוררות "כתות". המילה "כת" אינה משמשת לסימון שגיאות רציונליות, אלא כתווית לדבר־מה *הנראה מוזר*.

לא כל שינוי הוא שיפור, אך כל שיפור הוא בהכרח שינוי. אם ברצונכם לעשות דבר־מה טוב יותר, לא תהיה לכם ברירה אלא לעשותו אחרת. החוכמה המקובלת אכן מגלמת כמות לא מבוטלת של... ובכן, חוכמה אמיתית. וכן, הגיוני לדרוש נטל הוכחה כבד יותר עבור דברים מוזרים. אך *החשש* הזה אינו בבחינת שיקול דעת מכוון ורציונלי מסוג זה. זהו הפחד מפני החזקה באמונה שתגרום לחבריכם לתת בכם מבטים מוזרים במיוחד. מסיבה זו אנשים שואלים "זו לא *כת*, נכון?" בטון שלעולם לא היו משתמשים בו בהקשר של הגעה לעצרת פוליטית או הקמת מיצג חג־מולד ענקי.

זה החלק שמעצבן אותי.

כאילו ברגע שבו אתם מאמינים במשהו שאבותיכם לא האמינו בו, צונחת מן השמיים *פיית הכתות* ומחדירה בכם את *מהות הכת*, ולפני שתספיקו למצמץ, כולכם תעטו גלימות ותשירו מזמורים. כאילו אמונות "מוזרות" הן *מחולל הבעיות הישיר*, בלי קשר למניעת השינה ולמכות. כאילו הנזק שמחוללות כתות, כדוגמת ההתאבדות ההמונית של כת "שער גן־עדן" וכדומה, פשוט מוכיח שכל אדם המחזיק באמונה משונה הוא משוגע – המאפיין הראשון והעיקרי של "חברי כת" הוא היותם *חריגים בעלי מנהגים מוזרים*.

* ראו פרק 117.

אכן, אמונה החורגת מהנורמה החברתית מעמידה את הקבוצה בסיכון לצורת חשיבה של "אנחנו או הם", ל"קירור באידוי" ולבעיות נוספות. אולם החריגות היא גורם סיכון, ולא המחלה בפני עצמה. הוא הדין באשר למטרה שאתם סבורים שראוי להשיגה. בין אם האמונה נכונה ובין אם לאו, קיומה של מטרה נחמדה חושף אתכם תמיד לסיכון של מעגל מוות מאושר. הדבר הופך מטרות נעלות לגורם סיכון, ולא למחלה. וישנן מטרות שאכן ראוי לרדוף אחריהן.

מנגד, אינני רואה שום הצדקה לגיטימית למניעת שינה או לאיום במכות על מתנגדים. נקודה. כאשר קבוצה נוקטת פעולות מעין אלו, הרי שבין אם תכנו אותה "כת" ובין אם לאו, התשובה הפרגמטית לשאלה האם להצטרף אליה ברורה לחלוטין.

בעיה רביעית: הפחד מלהיות מתנגד בודד הוא כלי שאותו מנצלות *הכתות עצמן*. *החשש שחבריכם יבחנו אתכם בעין ביקורתית הוא בדיוק האפקט שבו משתמשות כתות אמיתיות לשם גיוס ושימור חברים* – על ידי הקפתו של המגויס בהסכמה מקיר לקיר מצד מאמיני הכת.

הפחד מפני רעיונות מוזרים והדחף לקונפורמיות הרחיקו ללא ספק קורבנות פוטנציאליים רבים מכתות של צלחות מעופפות. בהיותכם מחוץ לכת, הם משאירים אתכם בחוץ. אך בהיותכם *בפנים*, הם משאירים אתכם *בפנים*. קונפורמיות פשוט מדביקה אתכם למקומכם הנוכחי, יהא זה מקום טוב או רע.

האדם מייחל לקיומה של דרך שבאמצעותה יוכל להיות *סמוך ובטוח* שהוא אינו נתון ב"כת". לתשובה ניצחת ומוחלטת לאלה המישירים אליו מבטים מוזרים. לדרך שבה יוכל לדעת אחת ולתמיד שהוא עושה את הדבר הנכון, חף מכל אותם ספקות מציקים. נדמה לי שהדבר מכונה "הצורך בסגירת מעגל". וכמובן, הכתות מנצלות גם אותו.

מכאן הביטוי "כת ה-'אנחנו לא כת'".

חיים בצל הספק אינם מהווים מעלה – תכליתו של כל ספק היא להשמיד את עצמו, בהצלחה או בכישלון, וספק שמוסיף לרחף סתם

כך אינו משיג דבר. ואולם לעיתים דרוש לַסָּפֵק זמן לא מבוטל כדי להשמיד את עצמו. החיים עם ערימת ספקות שטרם באו על פתרונם הם עובדה בלתי נמנעת עבור רציונליסטים. אל לַספק לעורר פחד. שאם לא כן, תיאלצו לבחור בין חיים רדופים להחריד לבין חיים אוויליים להחריד.

אם באמת ובתמים אינכם מצליחים להכריע בשאלה האם קבוצה מסוימת היא ״כת״, תיאלצו פשוט לבצע בחירה בתנאי חוסר ודאות. בזה בדיוק עוסקת תורת ההחלטות.

בעיה חמישית: היעדר חשיבה אסטרטגית.

אני מכיר אנשים שנזהרים בסביבת תומכי הסינגולריות ונוקטים את אותה מידת זהירות בסביבת מפלגות פוליטיות ודתות מרכזיות. *נזהרים* – לא לחוצים או מתגוננים. אנשים אלה יכולים להבחין בן־רגע שהתנועה הסינגולרית איננה כת מן המניין המפעילה מניעת שינה וכדומה. אולם הם חוששים שמא *תהפוך* לכזו, בגין גורמי סיכון כגון הפיכת המושג ״בינה מלאכותית רבת־עוצמה״ ל״סוכן מאושר־מאוד״ (סוכן המוגדר בראש ובראשונה על ידי עצם ההסכמה על כל דבר חיובי שנאמר עליו). עצם העובדה שתנועה איננה כת בהווה, איננה ערובה לכך שלא תהפוך לכת בעתיד. תכונת ה׳כת׳ היא *מושך*, ולא מהות.

האם זהירות מסוג *זה* מעצבנת אותי? כלל וכלל לא. אני עצמי מקדיש זמן רב לדאגה מפני תרחיש זה. אני משתדל למקם את אבני הגו שלי מראש כדי לחסום כל תנועה בכיוון זה. מכאן, למשל, נולדה סדרת הפרקים העוסקת בכשלי חשיבה המאפיינים כתות.

אנשים המדברים על ״רציונליות״ מתמודדים עם גורם סיכון נוסף: לייעץ לאנשים איך לחשוב זהו עסק מסוכן מיסודו. אולם זהו *גורם סיכון*, ולא *מחלה*.

שתי המטרות האהובות עליי חשופות לסיכון התיאורטי של נפילה לכת. אף על פי כן, משום מה שואלים אותי ״האם אתה בטוח שזו לא כת?״ בתדירות גבוהה הרבה יותר כשאני מדבר על בינה מלאכותית רבת־עוצמה מאשר כשאני מדבר על תורת ההסתברות או על מדעי

הקוגניציה. אינני יודע אם רמת הסיכון של האחת גבוהה מזו של האחרת, אך אני יודע מי משתיהן *נשמעת מוזרה יותר*...

בעיה שישית עם השאלה "זו אינה כת, נכון?"...

עצם העלאת השאלה מציבה אותי במעין *מלכוד 22* מעצבן במיוחד. מנהיג כת מרושע באמת היה לבטח מנצל את חששו של השואל ורוקם טיעון משכנע ומפותל המסביר *מדוע זו אינה כת* – טיעון שהאדם היה להוט לאמץ. לעיתים מתקבל אצלי הרושם שזה מה שאנשים *מצפים* שאעשה! בכל פעם שאני מנסה לכתוב על סכנת הכת ועל הדרכים להימנע ממנה, מנקרת בי התחושה שאני נכנע לאותה כמיהה פגומה ושבסופו של דבר אני מספק לאנשים *אישור חיצוני*. אפילו כאשר אני מבהיר שנדרש מאבק מתמיד באנטרופיה.

תחושתי היא שאני הופך עצמי למתנגד הראשון בניסוי הקונפורמיות של אש, ואומר לאנשים: "כן, קו X אכן זהה לקו B, וזה בסדר גמור שגם אתם תאמרו זאת." אלא שהם כלל לא אמורים להזדקק לאישור שלי! או, גרוע מכך, אני מרגיש כאילו אני מציג טיעון מפותל המסביר *מדוע אין מדובר בכת*. זוהי שאלה *שגויה מיסודה*.

פשוט בחנו בעצמכם את תהליכי החשיבה של הקבוצה והחליטו אם זהו דבר־מה שברצונכם לקחת בו חלק, לאחר שתשתחררו מן הפחד מפני מוזרות. האחריות למנוע מעצמכם לגלוש לחשיבת־כת מוטלת עליכם, ואין זה משנה באיזו קבוצה אתם פעילים כרגע.

ברגע שאדם שואל "זו לא כת, נכון?" הרי שלא משנה כיצד אשיב, תמיד תתלווה לכך התחושה שאני מגן על משהו. אינני מחבב את ההרגשה הזו. תפקידו של מאסטר בייסיאני אינו לספק אישור חיצוני, כשם שאין זה מתפקידם של רציונליסטים להתגונן.

כתות ניזונות מחשיבת עדר, מחשש ומהשתוקקות לאישור. אי אפשר להפיג חשש באמצעות משאלות לב, וביטחון עצמי מזויף הוא גרוע שבעתיים. אך כל עוד אדם זקוק לאישור חיצוני – אפילו אישור בדבר היותו רציונליסט – זה תמיד יהיה סדק בשריון שלו. סייף מיומן ממוקד כל כולו במטרה ואינו מסיט את מבטו הצידה כדי

לבדוק שמא מישהו לועג לו. כאשר תדעו מה אתם מנסים להשיג ומדוע, תדעו גם האם אתם נוחלים בכך הצלחה, והאם קבוצה מסוימת מסייעת לכם או מפריעה בדרככם.

(נ.ב: אם אדם ניגש אליכם ושואל, "האם אתם *בטוחים* שזו לא כת?", אל תנסו להסביר את כל המושגים הללו בנשימה אחת. אתם מזלזלים במרחקי היסק.* אותו אדם יאמר, "אהא! אז אתה *מודה* שזו כת!" או "רגע, אתה טוען שאין לי סיבה לחשוש מהצטרפות לכתות?" או "אז... הפחד מכתות הוא עצמו כת? זה נשמע לי לגמרי כמו כת." מכאן נובעת הנקודה המעצבנת האחרונה – השביעית, אם אתם סופרים – והיא שלהסביר את כל הדברים הללו דורש אריכות כה רבה.)

* ראו פרק 9.

חלק י"א

לשחרר

הכרך *איך לשנות את דעתך הלכה למעשה* נחתם ברצף "לשחרר".

— **רוב בנסינגר,**
רציונליות: מבוא

121

חשיבותה של המילה "אופס"

זה עתה סיימתי לקרוא ספר על אודות קריסתה של חברת אֶנרוֹן, "האנשים הכי החכמים בחדר" (The Smartest Guys in the Room), שזוכה בזאת מטעמי בפרס "כותרת הספר הכי פחות הולמת".

מאפיין בלתי מפתיע של הדעיכה האיטית והקריסה הפתאומית של אנרון היה שחברי ההנהלה מעולם לא הודו באף טעות *גדולה* שעשו. כאשר אסון מספר 247 תפח לממדים שחייבו שינוי מדיניות ממשי, הם נהגו לומר: "חבל שזה לא עבד – זה היה רעיון כל כך טוב – איך נסתיר את הבעיה במאזן שלנו?" זאת בניגוד לאמירות כגון "במבט לאחור, ברור כעת שזו הייתה טעות מלכתחילה," או "הייתי טיפש." מעולם לא התרחש שם קו פרשת מים, רגע של תובנה מעוררת ענווה או הכרה בבעיית *עומק*. לאחר פשיטת הרגל דחה ג'פרי סקילינג, לשעבר סמנכ"ל התפעול ולזמן קצר גם מנכ"ל החברה, את עצת עורכי דינו לשמור על זכות השתיקה. במקום זאת, הוא העיד בפני הקונגרס שאנרון הייתה חברה *נהדרת*.

לא כל שינוי הוא שיפור, אך כל שיפור הוא בהכרח שינוי. אם נודה אך ורק בשגיאות מקומיות וקטנות, נבצע רק שינויים מקומיים וקטנים. המוטיבציה לשינוי *יסודי* נובעת מהכרה בטעות *יסודית*.

בילדותי גדלתי במידה שווה על ברכי המדע והמדע הבדיוני. מהַיינלַיין ועד פיינמן, ספגתי את מוסכמותיה של הרציונליות המסורתית: תיאוריות חייבות להיות נועזות ולהעמיד עצמן

למבחן הפרכה, הֱיו מוכנים להקרבה ההרואית הכרוכה בוויתור על רעיונותיכם כשאתם ניצבים מול ראיות סותרות, התווכחו בהגינות, נסו שלא להונות את עצמכם, ועוד שלל ניסוחים מילוליים מעורפלים.

חינוך רציונליסטי מסורתי חותר להצמיח מתדיינים שייכנעו, *בסופו של דבר*, בפני ראיות סותרות – כלומר, צריך שיהיה הר *כלשהו* של ראיות שיהיה גדול מספיק על מנת להזיז אתכם מעמדתכם. אין זה עניין של מה בכך. זהו הקו המבדיל בין מדע לדת. ואולם, חינוך כזה שם פחות דגש על *מהירות*, על נטישת המאבק *בשלב מוקדם ככל האפשר*, ועל שילוב *יעיל* של ראיות כך שיידרש רק *מינימום* של ראיות סותרות כדי למוטט את האמונה היקרה לליבכם.

התחנכתי על מורשת הרציונליות המסורתית, והחזקתי מעצמי רציונליסט לא קטן. עברתי ל"אמנות ההסקה הבייסיאנית" (לָפְלָס/ג'יינס/טֶבֶרסקי/כהנמן) בעקבות... ובכן, זהו סיפור ארוך. בקווים כלליים, עשיתי זאת משום שהבנתי כי המוסכמות המילוליות העמומות של הרציונליות המסורתית לא הצליחו למנוע ממני לעשות טעות אדירה.

רק לאחר שהודיתי סוף־סוף בטעותי במלואה, הבטתי לאחור על הנתיב שהוביל אותי לאותה "תובנה נוראה". נוכחתי לדעת שעשיתי סדרה של ויתורים קטנים ומינימליים – ויתרתי בחירוק שיניים על כל שעל, הכרתי בטעותי מעט ככל האפשר בכל הזדמנות, והודיתי בכישלון רק במנות קטנות ונסבלות. הבנתי שיכולתי להתקדם הרבה יותר מהר אילו רק הייתי צועק *"אופס!"*

אז עלתה בי המחשבה: *אני צריך להעלות את רמת המשחק שלי*.*

ישנו *יתרון עצום* בהודאה בכך שעשיתם טעות *גדולה*. זה כואב. אך זה גם יכול לשנות את חייכם מן הקצה אל הקצה.

---

* "זה לא שהארי הלך בדרך הלא נכונה, זה לא שהדרך לשפיות נחה מחוץ למדע. אבל קריאת מאמרים מדעים לא הייתה מספיקה. כל מאמרי הפסיכולוגיה הקוגנטיבית על הכשלים הידועים במוח האדם ושאר הדברים מהסוג הזה אכן עזרו, אבל זה לא היה מספיק. הוא נכשל בלהגיע למה שהארי התחיל להבין שהיה רמה גבוהה בצורה *מזעזעת*, רמת רציונליות שלא תאמן שאיתה אתה ממש יכול *לעשות* דברים נכון, בניגוד לשפה שימושית בה ניתן לתאר לאחר מעשה את כל מה שבו טעית."

חשוב לחוות את קו פרשת המים הזה, את רגע התובנה מעורר הענווה. להכיר בבעיה *מהותית*, ולא לפרק אותה לטעויות זעירות שקל לעכל.

אל תתמסרו לדרמה ותתגאו בעצם ההודאה בטעויות. מובן שעדיף לעשות את הדבר הנכון בפעם הראשונה, אך אם כבר טעיתם, מוטב בהרבה להכיר בכך בבת אחת. אפילו מבחינה נהנתנית, עדיף לספוג הפסד אחד גדול מאשר הפסדים קטנים רבים. החלופה היא למתוח את המאבק הפנימי על פני שנים. החלופה היא חברת אנרון.

מאז הזדמן לי לצפות באחרים שעושים סדרת ויתורים מינימליים משלהם, נסוגים בחירוק שיניים מכל שעל, לעולם אינם מודים בטעות גורפת כאשר ניתן להסתפק בהודאה בשגיאה מקומית, ותמיד מפיקים את המינימום האפשרי מכל שגיאה. את מה שיכלו לתקן באבחה אחת ומרצונם החופשי, הם הופכים לאוסף טלאים מקומיים זעירים שצריך לאלץ אותם לקבל תוך ויכוח. גם כשהם מודים בטעות, לעולם לא תשמעו אותם אומרים משהו כמו: "הייתי טיפש." הם עושים כל שביכולתם כדי למזער את מבוכתם באמצעות קביעות כגון "צדקתי ברמה העקרונית," "זה היה יכול לעבוד," או "אני עדיין רוצה להיאחז במהות האמיתית של ה[מה־שזה־לא־יהיה שאליו אני קשור]." עצם ההגנה על גאוותם באותו רגע חולף מבטיחה שיחזרו על אותה טעות בדיוק, ושוב ייאלצו להגן על גאוותם.

מוטב לבלוע את הגלולה המרה במלואה, בלגימה איומה אחת.

122

ההצעה להפוך לטרחן

כשהייתי צעיר מאוד – בן שלוש־עשרה או ארבע־עשרה, כמדומני – סברתי שמצאתי הפרכה לטיעון האלכסון של קַנטוֹר, משפט מפורסם המוכיח שיש יותר מספרים ממשיים ממספרים רציונליים.* אח, חלומות התהילה והגדולה שריצדו בראשי!

הרעיון שלי היה שמכיוון שכל מספר שלם ניתן לפירוק לחזקות של 2, אפשר למפות את המספרים השלמים אל קבוצת תת־הקבוצות של המספרים השלמים, פשוט באמצעות כתיבת ייצוגם הבינארי. המספר 13, לדוגמה, שייצוגו 1101, ימופה לקבוצה $\{3, 2, 0\}$. נדרש לי שבוע שלם עד שעלה בדעתי שאולי כדאי לי *להחיל* את טיעון האלכסון של קנטור על הבנייה המתוחכמת שלי. כצפוי, הניסיון הניב דוגמה נגדית – המספר הבינארי (...1111), שאינו תואם לאף מספר שלם סופי.

וכך, מצאתי את הדוגמה הנגדית הזו, הבנתי שנסיון ההפרכה שלי שגוי, ויחד איתו התפוגגו גם חלומות התהילה והגדולה שלי.

בתחילה הייתי מאוכזב מעט.

מחשבה חלפה בראשי: "אני עוד אכניע את המשפט הזה בסוף! *יום אחד* אפריך את טיעון האלכסון של קנטור, גם אם ניסיוני הראשון

* בניסוח מדויק יותר, המשפט מוכיח *שעוצמת* קבוצת המספרים הממשיים ($\mathbb{R}$) גדולה מזו של קבוצת המספרים הרציונליים ($\mathbb{Q}$). *עוצמה* הוא המונח המתמטי לגודל של קבוצה, גם כשמדובר בקבוצות אינסופיות.

כשל!" נטרתי טינה למשפט על כך שהתעקש להיות נכון, על שגזל ממני את תהילתי ועושרי, והתחלתי לחפש הפרכות אחרות.

ואז הבנתי משהו. הבנתי שטעיתי. מרגע שעמדתי על טעותי, לא הייתה לי כל סיבה לחשוד באיתנותו של טיעון האלכסון של קנטור יותר מאשר בכל משפט מרכזי אחר במתמטיקה.

ראיתי אז בבהירות רבה שהונחה לפתחי ההצעה להפוך לטרחן מתמטי, ולבלות את שארית חיי בכתיבת מכתבים זועמים בדיו ירוקה לפרופסורים למתמטיקה. (קראתי פעם ספר על טרחנים מתמטיים.)

לא רציתי שכך ייראה העתיד שלי, אז צחקתי קלות והנחתי לעניין. נפרדתי לשלום מטיעון האלכסון של קנטור במיטב האיחולים, ולא פקפקתי בו שוב.

ואינני זוכר כעת אם חשבתי על כך כבר אז, או בשלב מאוחר יותר... אבל איזה מבחן נורא ובלתי הוגן להעמיד בו ילד בן שלוש־עשרה. העובדה שנדרשתי כבר בגיל צעיר כל כך להיות כה רציונלי, או להיכשל.

ככל שאתם חכמים יותר, כך סביר להניח שתהיו צעירים יותר בפעם הראשונה שבה יעלה בדעתכם רעיון שייראה לכם מהפכני באמת. שיחק לי מזלי והבחנתי בטעות בעצמי. לא נדרש מתמטיקאי אחר כדי להצביע עליה, ובכך אולי לספק לי גורם חיצוני שאוכל להאשים. התמזל מזלי וההפרכה הייתה פשוטה דיה כך שאוכל להבין אותה. ייתכן שהייתי מתאושש בסופו של דבר, כך או כך. התאוששתי מדברים גרועים מזה בהרבה כאדם בוגר. אולם, אילו הייתי פונה בפניה הלא הנכונה בשלב מוקדם כל כך, האם הייתי אי־פעם מפתח את המיומנות להודות בטעות?

אני תוהה כמה מהאנשים שכותבים מכתבים זועמים בדיו ירוקה היו בני שלוש־עשרה כשעשו את אותו צעד ראשון שגוי וגורלי. אני תוהה כמה מהם היו מוחות מבטיחים קודם לכן.

עשיתי טעות. זה הכל. לא *צדקתי מבחינה מהותית, עמוק בפנים*; לא רשמתי ניצחון מוסרי; לא הפגנתי שאפתנות, ספקנות או כל מעלה פלאית אחרת; לא הייתה זו שגיאה סבירה; לא צדקתי באופן חלקי,

ואפילו לא בחלקיק. חשבתי מחשבה שלעולם לא הייתי חושב אילו הייתי חכם יותר, וזה כל הסיפור.

אילו לא הייתי מסוגל להודות בכך בפני עצמי, אילו הייתי מפרש את הטעות שלי מחדש כמעלה, אילו הייתי מתעקש להיות לפחות *קצת* צודק למען גאוותי, הרי שלא הייתי מרפה. הייתי ממשיך לחפש פגם בטיעון האלכסון. ובמוקדם או במאוחר, אולי אף הייתי מוצא אחד.

עד שלא תודו בטעות, לא תוכלו להמשיך בחייכם: הדימוי העצמי שלכם יישאר כבול לטעות הישנה.

בכל פעם שאתם מתפתים להיאחז בְּמחשבה שלעולם לא הייתם חושבים אילו הייתם חכמים יותר, מונחת לפתחכם ההצעה להפוך לטרחנים כפייתיים – גם אם לעולם לא תכתבו מכתבים זועמים בדיו ירוקה. גם אם איש אינו טורח להתווכח עמכם, או אם לעולם לא תשתפו איש ברעיון שלכם, ייתכן שאתם עדיין טרחנים. *ההיאחזות בטעות* היא שמגדירה זאת.

זה לא נכון. עמוק בפנים זה לא נכון. זה לא חצי נכון או אפילו מעט נכון. אין זו אלא מחשבה שלעולם לא הייתם צריכים לחשוב. לא בכל רע טמון גם טוב. בני אדם עושים טעויות, ולא כולן הן הצלחות במסווה. בני אדם טועים. זה קורה, וזה הכל. אמרו "אופס," והמשיכו הלאה בחייכם.

123

תתייאשו כבר

קייסי סֶרין, מתכנת אתרים בן 24 ללא כל ניסיון קודם בנדל״ן, חייב לבנקים 2.2 מיליון דולר לאחר ששיקר בבקשות למשכנתא כדי לרכוש במקביל שמונה בתים במדינות שונות בארצות הברית. הוא משך מזומנים מתוך המשכנתאות (באמצעות הגשת בקשות לסכומים גבוהים ממחירי הבתים בפועל) והשתמש בכסף להוצאות מחיה וסמינרים בנושא נדל״ן. ככל הנראה, הוא ציפה שהשוק יעלה.

וזהו אפילו לא החלק העצוב. החלק העצוב באמת הוא שסרין *עדיין לא אמר נואש*. הוא לא השלים עם תבוסתו. הוא סירב להכריז על פשיטת רגל או למצוא עבודה. הוא עדיין סבור שביכולתו להתעשר מנדל״ן. הוא המשיך לבזבז כסף על סמינרים, ואף ניסה לקחת משכנתא על בית תשיעי. הוא הרי לא נכשל – הוא פשוט עבר *חוויה מלמדת*.

זה מה שקורה כשמסרבים לאבד תקווה.

אף שהתנהגות זו עשויה להיראות כטיפשות גרידא, היא בנוסף מזכירה לי שני כלכלנים זוכי פרס נובל...

...הלוא הם מרטון ושולס מקרן הגידור LTCM.

אף ש-LTCM גרפה רווחי ענק בשלוש שנותיה הראשונות, בשנת 1998 כשלי השוק שהקרן ניצלה לטובתה החלו להיעלם – אנשים אחרים פיצחו את השיטה, ולכן היא חדלה לפעול.

LTCM סירבה לאבד תקווה. תוך התמכרות לתשואות שנתיות של

40%, החברה לקחה עוד ועוד הלוואות כדי לנצל שולי רווח שהלכו והצטמצמו. כאשר הכל החל להשתבש עבור LTCM, עמד לרשותה הון עצמי של 4.72 מיליארד דולר, מינוף של 124.5 מיליארד דולר, ופוזיציות בנגזרים בסך 1.25 טריליון דולר.

כל מקצוע דורש סוג אחר של חוכמה – מיומנויות מגוונות שיש לרכוש וכללים שונים שיש לפעול לפיהם. לכן, אולי תחשבו שלחקר ה"רציונליות", כדיסציפלינה כללית, לא תהיה תרומה רבה להצלחה בחיים האמיתיים. ובכל זאת, נדמה לי שהידיעה *כיצד לא להיות טיפש* מתאפיינת בעקרונות משותפים החלים על שלל מקצועות. אם תבקשו ללמד מישהו *כיצד לא להפוך טעויות קטנות לטעויות גדולות*, תגלו שזו כמעט אותה אמנות, בין אם מדובר בקרנות גידור ובין אם מדובר ברומנטיקה. ואחד המפתחות המרכזיים לכך הוא זה: היו מוכנים להודות שהפסדתם.*

* "אני מציע לך כוח," אמרה הדמות האפלולית, "ואספר לך על הכוח הזה ועל מחירו. הכוח מגיע מהכרת מבנה המציאות ועל ידי כך השגת שליטה עליה [...] מחירו של כוח זה הוא שעליך ללמוד לשאול את הטבע שאלות, ומה שקשה הרבה יותר, ללמוד לקבל את תשובותיו. אתה תערוך ניסויים, תבצע בדיקות ותצפה בתוצאות. ואתה תהיה חייב לקבל את משמעותן של התוצאות האלו כשהן יאמרו לך שאתה טועה. אתה תהיה חייב ללמוד להפסיד – לא לי, אלא לטבע. כשתמצא את עצמך מתווכח עם המציאות, תהיה חייב לתת למציאות לנצח." (הארי ג'יימס פוטר־אוונס־ורס)

124

השימוש הנכון בספק

פעם, בהרצאה על *הדרך*, ציינתי שרוב מערכות האמונה המאורגנות נועדו למעשה לאפשר *הסתלקות מן הספק*. אחד המאזינים השיב לי שהישועים* בוודאי חסינים מפני ביקורת זו, שכן הם מתַרגלים ספק מאורגן. הוא טען שלמתלמדים שלהם נאמר להטיל ספק בנצרות, בקיום האל, באמיתות הייעוד שלהם, ובמידת התאמתם לנדרי פרישות ועוני נצחיים. ואני שאלתי: "אה, אבל הם אמורים להתגבר על הספקות האלה, נכון?" הוא ענה: "לא, הם אמורים להטיל ספק באפשרות שהם יצליחו להתגבר על הספקות."

חיפוש מהיר לא העלה אישוש או הפרכה לטענות הללו. אולם, התרחיש הזה מרתק בעיניי וראוי לדיון, בין שהוא נכון מבחינה היסטורית ובין שלא. אילו אכן תרגלו הישועים ספק מכוון כפי שתואר, האם *פירוש הדבר* שפעלו באופן רציונלי?

אני נאלץ להודות שלא יהיה נכון לתאר את הישועים בתרחיש (הספק־היפותטי) הזה כמי ש"מסתלקים מן הספק". ואולם, התנהגות (ספק־היפותטית) זו עדיין מעוררת בי חשד כבד. עבור רציונליסט אמיתי, ספק לא אמור להיות משהו מאיים. ההתנהגות שתוארה נשמעת לי כמו תוכנית להקהיית־רגישות לגורם מפחיד, בדומה

* *ישועים* (jesuits), או בשמם הרשמי *אגודת ישוע* (society of jesus), הם מסדר דתי קתולי שנוסד במאה ה-16. הם ידועים במערכת חינוך שיטתית המדגישה לוגיקה, רטוריקה ומדעים לצד האמונה.

לחשיפת אדם הסובל מאֲרַכְנוֹפוֹבְיָה* לעכבישים בתנאים מבוקרים בקפידה.

ובכל זאת, הם מעודדים את המתלמדים שלהם להטיל ספק – הלא כן? האם זה משנה שמניעיהם פגומים? האם הדבר אינו נחשב בעיני הרציונליסט למעשה ראוי?

כל סקרנות שואפת להשמיד את עצמה. אין סקרנות שאינה משתוקקת לתשובה. אך אם השגתם תשובה, אם סיפקתם את סקרנותכם, הרי שהתעלומה המפוארת חדלה להיות מסתורית.

באותו אופן, כל ספק קיים כדי להשמיד אמונה מסוימת. אם ספק נכשל בהשמדת מטרתו, הוא גווע מבלי שהתממש – אך זוהי עדיין הכרעה, נקודת סיום, גם אם משמחת פחות. ספק שלא משמיד את עצמו ולא משמיד את מטרתו, כמוהו כספק שלא היה קיים מעולם. *הכרעת* הספקות, ולא עצם פעולת הטלת הספק, היא המניעה את גלגלי הרציונליות קדימה.

כל שיפור הוא שינוי, אך לא כל שינוי הוא שיפור. כל רציונליסט מטיל ספק, אך לא כל ספק הוא רציונלי. עטיית אצטלה של ספק לא הופכת אתכם לרציונליסטים, בדיוק כפי שלבישת חלוק לבן לא הופכת אתכם לרופאים.

ספק רציונלי מתעורר מסיבה ספציפית – קיימת הצדקה מסוימת לחשד שהאמונה שגויה. סיבה זו, בתורה, מתווה כיוון חקירה שישמיד את האמונה שעל הכוונת, או שישמיד את הספק עצמו. הדבר נכון אפילו לגבי ספקות מופשטים ביותר, כגון: "אני תוהה אם קיימת השערה פשוטה יותר המסבירה גם היא את הנתונים הללו." במקרה כזה, החקירה מתבצעת באמצעות ניסיון להגות השערות פשוטות יותר. ככל שהחיפוש נמשך ואינו נושא פרי, כך ההסתברות שהניסיון הבא ינחל הצלחה הולכת ופוחתת. בסופו של דבר, עלות החיפוש תעלה על התועלת הצפויה ממנו, ותחדלו לחפש. בשלב זה אינכם יכולים לטעון עוד שאתם מטילים ספק *באופן מועיל*. ספק שלא

* פוביה ספציפית מעכבישים.

נחקר, כמוהו כספק שאינו קיים. כל ספק קיים כדי להשמיד את עצמו, בדרך זו או אחרת. ספק לא־פתור הוא פעולת סרק. הוא לא מסובב את הגלגל, לא קדימה ולא אחורה.

אם אתם באמת ובתמים מאמינים בדת שלכם (ולא רק "מאמינים באמונה" בה), מדוע שתורו למתלמדים שלכם לשקול ספקות שנגזר עליהם לגווע מבלי להתממש? הדבר משול להוראה לסטודנטים לפיזיקה להתאמץ ולהטיל ספק במהפכות המדעיות של המאה העשרים, שמא בעצם המכניקה הניוטונית הייתה נכונה לכל אורך הדרך. אם אינכם מטילים ספק *באמת*, מדוע *להעמיד פנים* שכן?

משום שכולנו רוצים להצטייר כרציונליים, והטלת ספק *נתפסת לרוב* כתכונה רציונלית. אולם מעטים מבינים שנדרשת סיבה מסוימת כדי להטיל ספק, או שספק לא־פתור הוא פעולת סרק. במקום זאת, אנשים סבורים שהדבר קשור *לצניעות*, להתנהגות כנועה, או לשמירה על היררכיית המעמדות השבטית — בעיה כמעט זהה לזו הקיימת ביחס למושג הענווה, עליה כתבתי בעבר.* מפגן פומבי וראוותני של ספק, שנועד לשכנע את עצמכם שאתם רציונליסטים, יועיל לכם בערך כמו לבישת חלוק מעבדה.

כדי להימנע מהעמדת פנים של ספקנות, זכרו:

- ספק רציונלי קיים כדי להשמיד את אמונת־המטרה שלו, ואם אינו משמיד את מטרתו, הוא גווע מבלי להתממש.
- ספק רציונלי נובע מסיבה ספציפית כלשהי לחשוד שהאמונה עשויה להיות שגויה.
- ספק לא־פתור הוא פעולת סרק.
- ספק שלא נחקר, כמוהו כספק שלא היה קיים מעולם.
- אל לכם להתגאות בעצם הטלת הספק, אף שזכותכם המלאה להתגאות אם רגע לפני כן *קדעתם לגזרים* אמונה שהייתה יקרה ללִבכם.

* ראו פרק 46.

- אף שייתכן כי דרוש אומץ כדי להתמודד עם ספקותיכם, לעולם אל תשכחו *שעבור מוח אידיאלי*, ספק לא אמור להיות גורם מאיים מלכתחילה.

125

אתה יכול להתמודד עם המציאות

מה שאמיתי הוא כבר האמת.
להכיר בו לא ירע את המצב.
להתעלם ממנו לא יגרום לו להיעלם.
ומפני שהוא אמיתי, הוא הדבר היחיד עמו אנו באים במגע.
כל מה שלא אמיתי אינו שם כדי לחיות אותו.
בני אדם יכולים לשאת את האמת,
שהרי הם מתמודדים איתה ממילא.

— יוג׳ין גנדלין

126

מדיטציית סקרנות

המעלה הראשונה היא סקרנות.

— שתים־עשרה המעלות של הרציונליות

כרציונליסטים, אנו מחויבים למתוח ביקורת על עצמנו ולהטיל ספק באמונותינו... הלא כן?

תארו לעצמכם מה קורה לכם, ברמה הפסיכולוגית, אם אתם פותחים באמירה: "זוהי חובתי לבקר את אמונותיי שלי." רוג'ר זילאזני הבחין פעם בין "לרצות להיות סופר" לבין "לרצות לכתוב". מארק טוויין אמר: "קלאסיקה היא יצירה שכולם רוצים להימנות עם מי שכבר קרא אותה, ואיש אינו רוצה לקרוא אותה בפועל." באותו אופן, העברת ביקורת עצמית מתוך תחושת חובה מותירה אתכם עם הרצון *להיות במצב שבו כבר חקרתם*, על מנת שתוכלו לומר שאמונתכם אינה עיוורת. זה לא אותו הדבר כמו *לרצות לחקור*.

דבר זה עלול להוביל לעצירה מוקדמת של החקירה שלכם. אתם שוקלים השגה כלשהי, מיד לאחר מכן מעלים טיעון־נגד לאותה השגה, *ושם* אתם עוצרים. אתם חוזרים על הפעולה עם כמה השגות, עד שאתם חשים שמילאתם את חובתכם לחקור, *ושם אתם עוצרים*. השגתם את היעד הפסיכולוגי הסמוי שלכם: סילוק הדיסוננס הקוגניטיבי שהיה נוצר מכך שאתם מחשיבים את עצמכם

לרציונליסטים, ויחד עם זאת יודעים שלא ניסיתם למתוח ביקורת על אמונתכם. ניתן לכנות זאת רכישת סיפוק רציונליסטי – הניסיון ליצור את אותה "תחושה חמימה" של יציאה־ידי־חובה.

לאחר מכן, רמת ההסתברות המוצהרת שלכם תהיה גבוהה דיה כדי להצדיק את שימור התוכניות והאמונות שאיתן יצאתם לדרך, אך לא גבוהה עד כדי כך שתעורר חוסר אמון מצידכם או מצד רציונליסטים אחרים.

כאשר אתם סקרנים באמת, תימשכו באופן טבעי לחקירות שמסתמנות כבעלות הפוטנציאל הגבוה ביותר לחולל תזוזות באמונה, או לחקירות השונות ככל האפשר מאלו שניסיתם בעבר. לאחר מכן, סביר להניח שהתפלגות ההסתברות שלכם *לא* תיראה כפי שהיתה כשהתחלתם – אמורות להתרחש תזוזות, כלפי מעלה או כלפי מטה. כל אחד מהכיוונים אמור להיות מקובל עליכם באותה המידה, אם אתם אכן סקרנים באמת.

השוו זאת למניע התת־מודע המבקש להותיר את חקירתכם בטריטוריה מוכרת, כדי שתוכלו לסיים איתה מהר, כדי שתוכלו להיות *אחרי החקירה*, ולהשיב את האיזון עליו מושתתות תוכניותיכם ואמונותיכם המוכרות.

באשר לאופן שבו לדעתי צריכה להיראות סקרנות אמיתית, ולכוח הטמון בה, אני מפנה אתכם לפרק 12 – "משל על מדע ופוליטיקה". כל אחת מהדמויות בו נועדה להמחיש לקחים שונים. הדמות האחרונה, פְּריס, מגלמת את כוחה של סקרנות תמימה: קלילות והושטת יד נלהבת לעבר ראיות.

אורסוּלָה ק. לֶה גְוִוין כתבה: "בתמימות אין כוח נגד הרוע. אבל יש בה כוח למען הטוב."[1] סקרנות תמימה עלולה להשתבש בתום לב. לכן עלינו להעז ולקחת על עצמנו, כמעין סיכון, את הכשרת הרציונליסט ואת התחכום הנלווה אליה, אם ברצוננו להיות חזקים. אף על פי כן, עלינו לנסות ולשמר את הקלילות ואת הושטת־היד הנלהבת של אותה סקרנות תמימה.

כפי שכתוב ב*שתים־עשרה המעלות*:

> אם בליבך אתה מאמין שאתה כבר יודע, או אם בליבך אינך חפץ לדעת, אזי חקירתך תהא חסרת תכלית וכישוריך נטולי כיוון. הסקרנות מבקשת להשמיד את עצמה. אין סקרנות שאינה משתוקקת לתשובה.

פשוט אין בנמצא תחליף ראוי לסקרנות אמיתית. "תשוקה יוקדת לדעת נעלָה יותר מנדר חגיגי לרדוף אחר האמת." אולם, אי אפשר לייצר סקרנות בכוח הרצון בלבד, בדיוק כפי שאינכם יכולים להורות בכוח הרצון לכף רגלכם להרגיש חום בשעה שהיא חשה קור. לעיתים, כל שנותר בידינו הם נדרינו החגיגיים ותו לא.

אם כן, מה ניתן לעשות עם תחושת החובה? בתור התחלה, נוכל לנסות למצוא עניין בחקירות שאנו עורכים מתוך חובה – לפקוח עין בחיפוש אחר ניצוצות של עניין אותנטי, או אפילו אחר בורות אמיתית והרצון לפתור אותה. הדבר משתלב היטב עם מודעות מיוחדת כלפי אפשרויות מכאיבות שאתם נרתעים מפניהן – אין זו סתם "חשיבה שלילית".

ייתכן שיעזור גם לעשות מדיטציה על *חוק שימור הראיות הצפויות*.* עבור כל נקודת חקירה *חדשה*, או עבור כל פיסת ראיה *שטרם נצפתה* ואשר אתם מתבוננים בה לפתע, ההסתברות הפוסטריורית הצפויה אמורה להיות שווה להסתברות הפריורית שלכם. בתהליך־המיקרו של החקירה, אמונתכם צריכה להיות תמיד מאוזנת ומוכנה לנוע במידה שווה לכל אחד מהכיוונים. לא כל נקודה עשויה להספיק כדי לשנות את התמונה לחלוטין – למשל, להזיז את האמונה מהסתברות של 70% ל-30% – אך אם אמונתכם הנוכחית עומדת על 70%, עליכם להיות מוכנים להוריד אותה ל-69% באותה המידה שבה אתם מוכנים להעלותה ל-71%. אל לכם להניח שאתם יודעים לאיזה כיוון היא תנוע (בממוצע), שכן לפי חוקי תורת ההסתברות, אם אתם יודעים

* ראו פרק 28.

מהו היעד שלכם – הרי שאתם כבר נמצאים בו. אם אתם מסוגלים לחקור בכנות, כך שלכל נקודה *חדשה* באמת יהיה פוטנציאל שווה להזיז את האמונה כלפי מעלה או כלפי מטה, הדבר עשוי לסייע לכם לשמור על עניין ואפילו על סקרנות במהלך תהליך־המיקרו של החקירה.

אם הטיעון שאתם שוקלים אינו חדש, מדוע תשומת ליבכם מופנית אליו? האם זהו המקום שבו הייתם מחפשים אילו הייתם סקרנים באמת? האם בתת־המודע שלכם אתם מבקרים את אמונתכם בנקודות החוזק שלה, ולא בנקודות התורפה שלה? האם אתם עושים שימוש חוזר בראיות קיימות?

אם תצליחו להימנע מחזרה על ראיות תומכות שכבר ידועות לכם, ואם תצליחו להפחית את רמת האמונה שלכם בנגיסה זעירה אחת בכל פעם בעקבות ראיות חדשות, ייתכן שאף תוכלו לזנוח את האמונה לחלוטין – להבין מאיזה כיוון נושבות נגדכם רוחות הראיות.

אמצעי משקם נוסף לסקרנות הוא מה שהתחלתי לכנות *המנטרה של טרסקי*. זוהי למעשה מטא־מנטרה הניתנת להתאמה אישית לכל מקרה ומקרה (כפי שאכן ראוי). לדוגמה, אם אני תוהה במתח האם קופסה נעולה מכילה יהלום, הרי שבמקום לחשוב על כל ההשלכות הנפלאות שיהיו אילו הקופסה אכן מכילה יהלום, אני יכול לחזור על המנטרה של טרסקי:

> אם הקופסה מכילה יהלום,
> אני רוצה להאמין שהקופסה מכילה יהלום.
> אם הקופסה אינה מכילה יהלום,
> אני רוצה להאמין שהקופסה אינה מכילה יהלום.
> מי ייתן ולא איקשר לאמונות שאולי לא ארצה בהן.

לאחר מכן, עליכם לעשות מדיטציה על האפשרות שאין יהלום, על היתרון שיצמח לכם מכך שתאמינו שאין יהלום, ועל החיסרון שייגרם לכם אם תאמינו שיש יהלום.

אם תוכלו למצוא בתוככם ולו שמץ קל שבקלים של אי־ודאות אמיתית, שמרו עליו כפי שנווד מגונן על גחלת לוחשת. אם תוכלו לגרום לה להתלקח לכדי להבת סקרנות בוערת, הדבר ימלא אתכם קלילות ותשוקה, ויעניק תכלית לחקירתכם ונתיב לכישוריכם.

---

1. Ursula K. Le Guin, *The Farthest Shore* (Saga Press, 2001).

127

אין מי שיפטור אתכם מחוקי הרציונליות

הרציונליות המסורתית מנוסחת במונחים של *כללים חברתיים*, שהפרתָם מתפרשת כרמאות – כסטייה מנורמות של שיתוף פעולה. אם ברצונכם לשכנע אותי באמונה שלכם, חלה עליכם החובה לספק לי כמות מסוימת של ראיות. אם תנסו להתחמק מכך, כולנו נדע שאתם מועלים בחובתכם. תאוריה מחויבת לספק תחזיות נועזות משל עצמה, ולא רק לגנוב תחזיות שתאוריות אחרות עמלו קשות כדי לייצר. תאוריה מחויבת לחשוף את עצמה למבחן הפרכה – אם היא מנסה להשתמט מכך, זה כמו לנסות להשתמט מטקס חניכה אכזרי. עליכם לשלם את המחיר.

הרציונליות המסורתית מנוסחת בדומה למנהגים השולטים בחברות אנושיות, מה שמקל על העברתה מפה לאוזן. בני אדם מזהים רמאות חברתית במהימנות גבוהה הרבה יותר מאשר הפרות של כללי לוגיקה מופשטים, אפילו כשהמבנה שלהן זהה לחלוטין. ואולם, תפיסת הרציונליות כחובה חברתית מביאה לכמה רעיונות משונים.

כך למשל, ניתן למצוא אנשים דתיים המגִנים על אמונתם וטוענים: "ובכן, גם *אתם* לא יכולים להצדיק את האמונה שלכם במדע!" במילים אחרות: "איך אתה מעז לבקר אותי על כך שאני מחזיק באמונות חסרות ביסוס, צבוע שכמוך! הרי גם אתה עושה זאת!"

עבור בייסיאנים, המוח הוא מכונה שהפלט שלה הוא התאמה־למציאות: הוא מעבד ומזקק ראיות שזורות לכדי מפה המשקפת את השטח. עקרונות הרציונליות הם חוקים בדיוק באותו מובן שבו החוק השני של התרמודינמיקה הוא חוק: דרושה כמות ניתנת לחישוב של ראיות שזורות כדי לגבש אמונה מהימנה, בדיוק כשם שקירור אמין של תכולת מקרר דורש כמות מינימלית, הניתנת אף היא לחישוב, של אנרגיה חופשית.

באופן עקרוני, חוקי הפיזיקה הפיכים בזמן, כך שישנה הסתברות זעירה לאין שיעור – כזו שאינה שונה בפועל מאפס עבור כל אדם שאיננו מתמטיקאי – שמקרר יקרר את עצמו באופן ספונטני תוך ייצור חשמל. קיים סיכוי זעיר לאין שיעור, גדול מזה אך במעט, שתוכלו לשרטט מפת רחובות מפורטת ומדויקת של ניו יורק מבלי לבקר בה אפילו פעם אחת, בעודכם יושבים בסלון מוגף תריסים וללא חיבור לאינטרנט. אבל לא הייתי בונה על זה.

לפני שתנסו למפות שטח שמעולם לא ראיתם, מזגו מעט מים לכוס בטמפרטורת החדר והמתינו עד שיקפאו מעצמם בטרם תמשיכו. כך תוכלו לוודא שהשיטה הכללית – הסתמכות על סיכויי הצלחה אפסיים – אכן עובדת. ייתכן שלא תבינו מיד שהמפה שלכם שגויה, בייחוד אם לעולם לא תבקרו בניו יורק, אך תוכלו לראות במו עיניכם שמים אינם קופאים מעצמם.

אם כללי הרציונליות הם נימוסים חברתיים, אזי נראה שניתן לתרץ התנהגות מסוימת בטענה שגם אחרים נוהגים כך. לא יהיה זה *הוגן* לדרוש מכם ראיות, אם איננו מסוגלים לספק אותן עבור עצמנו. אנו נבין שאיש מאיתנו אינו טוב מרעהו, נתפייס, ונפטור אתכם ברוב־חסד מחובתכם החברתית לספק ראיות לאמונתכם. וכולנו נחיה באושר ובעושר, חירות, שוויון ואחווה.

אך אם כללי הרציונליות הם חוקים מתמטיים, אזי ניסיון להצדיק אמונה נטולת ראיות כיוון שגם אחרים נוהגים כך יהיה יעיל בערך כמו לְמנות שלושים סיבות מדוע אינכם אמורים ליפול מצוק. גם אם כולנו נחליט פה אחד שזהו חוסר צדק שמקרר זקוק לחשמל, הוא

עדיין לא יעבוד (בהסתברות של 1~). גם אם כולנו נחליט פה אחד שאינכם צריכים לבקר בניו יורק, המפה עדיין תהיה שגויה. אמא טבע מפורסמת באדישותה לתחנונים מסוג זה, וכמותה גם גבירת המתמטיקה.

לכן – בחזרה לשפה החברתית של הרציונליות המסורתית – אל תחשבו שתוכלו *לצאת ידי חובה* בטענה שזה בסדר להחזיק באמונות שרירותיות לגבי משהו אך ורק משום שגם לאנשים אחרים יש אמונות שרירותיות. אם שני צדדים לחוזה מתנהגים בצורה גרועה באותה המידה, שופט אנושי עשוי להחליט שלא להטיל עונש על אף אחד מהם. אבל אם שני מהנדסים מתכננים את המנועים שלהם בצורה גרועה באותה המידה, אף לא אחד מהמנועים יעבוד. שגיאת תכנון אחת אינה יכולה לתרץ שגיאה אחרת. גם אם *אני* עושה משהו בצורה שגויה, זה לא עוזר לכם, או פוטר אתכם מהכללים. זה פשוט אומר ששנינו בצרות צרורות.

מבחינת החוק האנושי בדמוקרטיות ליברליות, לכל אדם שמורה הזכות להחזיק באילו אמונות שיחפוץ. מבחינת חוקי הטבע, אין לכם זכות לדיוק. איננו עוצרים אנשים על כך שהם מאמינים בדברים משונים, לפחות לא במדינות הנאורות יותר. אך איש אינו יכול לבטל את החוק שלפיו אתם זקוקים לראיות כדי לגבש אמונות *מדויקות*. אפילו הצבעה פה אחד של המין האנושי כולו לא תזַכה אתכם ברחמים בבית הדין של הטבע.

פיזיקאים לא קובעים מהם חוקי הפיזיקה, הם רק מנחשים אותם. רציונליסטים לא קובעים מהם חוקי הרציונליות, אנו רק מנחשים אותם. אינכם יכולים להפוך דבר שאינו רציונלי מלכתחילה לרציונלי. אם, בזכות יכולת שכנוע יוצאת דופן, תשכנעו את כל הפיזיקאים בעולם שאתם פטורים מחוק המשיכה, ותצעדו מעבר לקצה צוק – אתם תיפלו. אפילו לומר "זה לא ביידיים *שלנו*" זהו משגה של האנשת־יתר. אין אף סמכות עליונה שיכולה לפטור אתכם. יש רק סיבה ותוצאה.

זכרו זאת כאשר אתם מתחננים לקבל פטור, רק הפעם. *איננו* יכולים לפטור אתכם. זה לא ביידיים שלנו.

128

השאירו נתיב מילוט

כשאתם מכתרים את האויב
השאירו לו תמיד נתיב מילוט.
עליו לראות שישנה
חלופה למוות.

— סוּן דְזֶה, *אמנות המלחמה*[1]

אל תעלו את הלחץ, הנמיכו את החומה.

— לוֹיס מקמָסטֶר בוּז׳וֹלד, *קוֹמאר*[2]

יצא לי פעם לשוחח עם בחורה — ״לא־רציונליסטית״ — שנקלעה במקרה למפגש רציונליסטים מקומי. היא הצהירה (א) על אמונתה בנשמות, ו-(ב) על כך שאינה מאמינה בקריוניקה,* משום שלשיטתה הנשמה לא תישאר עם הגוף הקפוא. שאלתי אותה: ״אבל כיצד את יודעת זאת?״ מבע הבלבול שהבזיק על פניה הבהיר שהשאלה הזו מעולם לא חלפה בדעתה. אינני אומר זאת לגנאי — היא הצטיירה כאדם חביב ללא כל הכשרה בחשיבה רציונלית, ממש כמו רוב רובו של המין האנושי. אני באמת חייב לכתוב את הספר הזה.

* ראו הערה בפרק 119.

מרבית השיחה שהתפתחה בינינו נסבה על נושאים שכבר כוסו באתר *Overcoming Bias*: אם אתם סקרנים *באמת* לגבי דבר־מה, סביר להניח *שתוכלו* למצוא דרך טובה לבחון אותו; נסו קודם כל לגבש אמונות מדויקות ואז תנו לרגשותיכם לנבוע מהן; וכיוצא באלה. אולם השיחה הזכירה לי מושג אחד שטרם עסקתי בו כאן.

"תודאי," הצעתי לה, "שאת מדמיינת כיצד היה נראה העולם אילו לא היו נשמות, ומה היית עושה במקרה הזה. אל תחשבי על כל הסיבות שבגללן הדבר אינו ייתכן. פשוט קבלי זאת כהנחת מוצא, ואז דמייני את ההשלכות. חשבי לעצמך כך: 'ובכן, אם אין נשמות, אני פשוט יכולה להירשם לקריוניקה,' או 'אם אין אלוהים, אני פשוט יכולה להמשיך להיות אדם מוסרי בכל זאת,' במקום לראות בכך תרחיש מחריד מכדי להתמודד עימו. מתוך כבוד עצמי, עלייך לנסות להאמין בָּאמת, תהא בלתי נוחה ככל שתהא. אולם, בהתחשב בטבע האנושי, זה עוזר להפוך אמונה למעט פחות מאיימת, לפני שמנסים להעריך את הראיות התומכות בה."

העיקרון הניצב בבסיס הטכניקה הזו פשוט: כפי שסוּן דְזֶה מייעץ לכם לעשות עם אויביכם, כך עליכם לנהוג עם עצמכם – השאירו לעצמכם נתיב מילוט, כדי שנסיגה תהיה פחות קשה. כך, למשל, התרחיש של אובדן מקום העבודה עשוי להיראות מפחיד הרבה יותר כשאינכם מסוגלים לשאת אפילו את המחשבה עליו, מאשר לאחר שחישבתם בדיוק לכמה זמן יספיקו חסכונותיכם, בחנתם את שוק העבודה באזורכם, ותכננתם במדויק מה תעשו הלאה. רק אז תהיו מוכנים להעריך בצורה *הוגנת* מהי ההסתברות שתשמרו על עבודתכם בגל הפיטורים המתוכנן בחודש הבא. הֱיו פחדנים אמיתיים, ותכננו את נתיב המילוט שלכם לפרטי פרטים – דמיינו כל שלב – ורצוי בטרם תגיעו לראשונה לשדה הקרב.

התקווה היא שיידרש פחות אומץ כדי לדמיין מצב עניינים לא נוח *בתור ניסוי מחשבתי*, מאשר לבחון *עד כמה סביר* שהוא נכון. ולאחר שתעשו את הראשון, יהיה קל יותר לעשות את השני.

זכרו שהסקה בייסיאנית דורשת דיוק: גם אם טענה מפחידה אמורה

באמת להיראות בלתי סבירה, חשוב עדיין לשקלל את כל הראיות, בעד ונגד, בהגינות מוחלטת, במטרה להגיע להסתברות כמותית רציונלית. ההסכמה לדמיין אמונה מפחידה *אינה* בבחינת הודאה בכך שאתם חושבים, עמוק בפנים, שהיא ככל הנראה נכונה. אתם יכולים לדמיין אמונה מפחידה מתוך עקרונות כלליים של היגיינה מחשבתית תקינה. "מחשבות שאינכם מעזים לחשוב שולטות בכם יותר מהמחשבות שאתם מבטאים בקול" – והדבר קורה גם אם המחשבה שלא ניתן להעלות על הדעת היא שקרית לחלוטין!

טכניקת השאירו־נתיב־מילוט אכן דורשת מינימום מסוים של כנות עצמית כדי לעשות בה שימוש ראוי.

בתור התחלה, עליכם לפחות להיות מסוגלים להודות בפני עצמיכם *אילו* רעיונות מפחידים אתכם, ולאילו רעיונות אתם קשורים רגשית. אולם זהו מבחן קל בהרבה משקלול הוגן של ראיות עבור רעיון המעורר בכם פחד. האם יעזור אם אומר שיוצא לי להשתמש בטכניקה הזו בעצמי? אחרי ככלות הכל, רציונליסט איננו אדם חסר רגשות. ישנם רעיונות שמפחידים אותי, ואף על פי כן אני עדיין מאמין שהם שקריים. ישנם רעיונות שאני יודע שאני קשור אליהם, ובכל זאת אני עדיין מאמין שהם נכונים. אולם אני עדיין מתכנן את נתיבי המילוט שלי, לא משום שאני *מתכנן* לסגת, אלא מפני שתכנון הנסיגה מראש עוזר לי לחשוב על הבעיה מבלי להיקשר אליה.

ואולם, המבחן הגדול יותר לכנות עצמית הוא לאמץ *באמת ובתמים* את הטענה הלא־נוחה כהנחת מוצא, ולהבין כיצד הייתם מתמודדים עימה בפועל. כאשר אנו ניצבים מול רעיון לא נוח, הדחף הראשון שלנו הוא, באופן טבעי, לחשוב על כל הסיבות לכך שזה פשוט *בלתי אפשרי*. מסיבה זו, תיתקלו במידה מסוימת של התנגדות פסיכולוגית פנימית אם תנסו לדמיין בדיוק כיצד העולם היה נראה, ומה הייתם עושים בנידון, אילו האמונה היקרה לכם מכל הייתה שקרית, או אם האמונה המפחידה אתכם מכל הייתה נכונה.

חשבו על כל אותם אנשים הטוענים כי ללא אלוהים, לא ייתכן מוסר (וכן, הנושא אכן עלה בשיחה, כך שלא מדובר באיש קש). לו

יכלו אנשים דתיים לדמיין את תגובתם *האמיתית* להכרה העובדתית שאלוהים אינו קיים, ייתכן שהיו מבינים שבעצם לא – הם לא היו יוצאים לרחובות וטובחים בתינוקות. הם היו עשויים להבין שאתאיסטים מגיבים לאי־קיומו של אלוהים פחות או יותר באותה דרך שבה הם עצמם היו מגיבים אילו היו מגיעים לאמונה כזו. אני מציין זאת כדי להראות שזהו אתגר לא מבוטל לדמיין את האופן שבו הייתם מגיבים *באמת*, אילו האמנתם בהיפוכה של אמונה שבה אתם נאחזים בחוזקה.

יתרה מזאת, תמיד מפתיע לגלות עד כמה אנשים מסתגלים למציאות חדשה. למשל, אנשים שהפכו לאחרונה למשותקים בארבע גפיים אינם עצובים כפי שהם ציפו שיהיו כעבור חצי שנה, וכן הלאה. ההבנה שאילו האמונה המפחידה הייתה מתבררת כנכונה *הייתם* משלימים איתה איכשהו, עשויה להיות מפתיעה באותה מידה. משותקים מתמודדים עם מצבם, וכך הייתם עושים גם אתם.

מה שאמיתי הוא כבר האמת. להכיר בו לא ירע את המצב. אל לכם לחשוש מעצם הניסיון *לדמיין* עולם שמפחיד אתכם. אם עולם זה הוא כבר המציאות, לדמיין אותו לא יעשה אותו לגרוע יותר. ואם הוא *איננו* המציאות, הדמיון לא יסב כל נזק. וזכרו בעודכם מדמיינים: אם הדברים המפחידים שאתם מְשווים בנפשכם אכן נכונים – והם עשויים שלא להיות! – הרי שאז הייתם, ללא ספק, רוצים להאמין בכך, ועליכם לדמיין גם את זה. אי־אמונה לא הייתה מועילה לכם במאום.

כמה אנשים דתיים היו דבקים באמונתם באלוהים, לו יכלו לדמיין *במדויק* את אותו עולם היפותטי שבו אין אלוהים והם עצמם הפכו לאתאיסטים?

השארת נתיב מילוט היא טכניקה רבת־עוצמה, אך איננה קלה ליישום. דמיון *כֵּן* אינו דורש מאמץ רב כפי שדורשת ההודאה *הגלויה* שאלוהים אינו קיים, אך הוא עדיין דורש מאמץ.

1. Sun Tzu, *The Art of War* (Cloud Hands, Inc., 2004).

2. Lois McMaster Bujold, *Komarr*, Miles Vorkosigan Adventures (Baen, 1999).

129

משבר אמונה

זה לא משבר אמונה אמיתי אלא אם הדברים יכלו להסתיים כך או אחרת באותו הקלות.

– תור שֶׁנקֶל

רבים בעולם הזה מחזיקים באמונות שגם ילד בן עשר היה מזהה את פגמיהן מיד, לו היה שומע אותן לראשונה. אינני מדבר על טעויות דקות מן הדק. עבור תודעה שאינה מעורבת רגשית, הוויתור על אמונות אלו היה משחק ילדים, אילו רק ספקנותו של בן עשר הייתה מיושמת ללא מורא. כפי שניסח זאת המשתמש Premise Checker: "לו רעיון האלוהים לא היה מופיע עד לעידן המדעי, רק אדם מוזר באופן חריג היה ממציא רעיון כזה ומעמיד פנים שהוא מסביר משהו."

ובכל זאת, מומחים מדעיים מיומנים, אפילו חלוצי המחקר המובילים בתחומם, ואפילו בימינו אנו, אינם מיישמים את הספקנות הזו בהצלחה. חתן פרס נובל ישראל אומן, הוגה *משפט ההסכמה של אומן*, הוא יהודי אורתודוקסי. אני חש ביטחון מסוים בהערכתי שאומן בחן את אמונתו והטיל בה ספק בשלב כזה או אחר. ועם זאת, הוא לא עשה זאת בהצלחה. אנו משנים את דעתנו פחות מכפי שנדמה לנו.

עובדה זו אמורה להחריד אתכם עד עמקי נשמתכם. משמעותה היא שאתם יכולים להיות מדענים ברמה עולמית, בקיאים ברזי המתמטיקה הבייסיאנית, *ועדיין* להיכשל בדחיית אמונה שכל בן עשר בעל מבט רענן היה מבחין באבסורד שבה. עובדה זו ממחישה את העמדה המתגוננת והבלתי מנוצחת שאמונה מסוגלת לבנות לעצמה, אם רק קיננה במוחכם למשך זמן רב מספיק.

מה נדרש כדי להביס טעות שבנתה לעצמה מבצר?

העניין הוא, שמרגע שאתם *יודעים* שזו טעות, היא למעשה כבר הובסה. השאלה איננה "כיצד אוכל לדחות את האמונה השגויה הזו, שבה החזקתי זמן רב?" אלא "כיצד אדע אם אמונה זו שבה החזקתי זמן רב, באמת שגויה?" הכנות העצמית הופכת לשברירית ביותר דווקא כשאיננו בטוחים איזו מן הדרכים היא הנכונה. לפיכך, השאלה הופכת להיות:

כיצד נוכל לייצר בעצמנו משבר אמונה אמיתי, כזה שיכול באותה הקלות לנטות לכאן או לכאן?

דת היא מקרה הבוחן שכולנו מסוגלים לדמיין. (קוראים שנולדו להורים אתאיסטים הפסידו ניסיון חיים מהותי, וייאלצו להסתפק בתחליף הדל של דימוי חבריהם הדתיים.) אך אם איבדתם כל סימפטיה, וכעת אתם תופסים מאמינים כמוטנטים מרושעים, לא תוכלו לדמיין את הניסיונות הפנימיים האמיתיים שהם חווים. לא תוכלו לשאול את השאלה:

"איזו אסטרטגיה כללית צריך אדם דתי לאמץ כדי להימלט מדתו?"

אני בטוח שישנם כאלו המתבוננים באתגר הזה וכבר מדקלמים רשימה של טיעונים אתאיסטיים שגורים: "הם יצטרכו להודות שאין ראיות בייסיאניות לקיומו של אלוהים," "הם יצטרכו להבחין בהתחמקויות המוסריות שהם מבצעים כדי לתרץ את התנהגותו של אלוהים בתנ"ך," "הם צריכים ללמוד להשתמש בתערו של אוקאם—"

טעות! טעות, טעות! החזרה השבלונית הזו, שבה אתם פשוט שולפים נקודות *שכבר חשבתם עליהן מזמן*, היא *בדיוק* דפוס

החשיבה המותיר אנשים כלואים בתוך הדת שלהם. אם תישארו עם מחשבות־המטמון שלכם, אם מוחכם ישלים את התשובה המובנת מאליה במהירות כזו שלא תאפשר לכם לבחון אותה במקוריות – לבטח לא תצליחו לחולל משבר אמונה.

אולי זה פשוט נובע מכך שמעטים מדי קראו את *גדל, אשר, באך* בגיל צעיר מספיק, אולם הבחנתי בכך שחלק ניכר מהאוכלוסייה – אפילו אנשים בעלי רקע מדעי – מתקשה לעקוב אחר טיעונים היורדים לעומק המטא הזה. בימיי הפסימיים יותר, אני תוהה אם לגמל אכן יש שתי דבשות.*

אפילו כשמצביעים על כך במפורש, ישנם אנשים שפשוט *אינם מצליחים לבצע את הקפיצה* מטיעון ברמת האובייקט של "השתמשו בתערו של אוקאם! אתם מוכרחים לראות שהאל שלכם הוא אמונה לא־נחוצה!" לטיעון ברמת המטא של "נסו למנוע מהמוח שלכם להשלים את התבנית בדרך הרגילה!" שכן, בדיוק כפי שכל חבריכם הרציונליסטים מדברים על תערו של אוקאם כדבר חיובי, ובדיוק כפי שתערו של אוקאם קופץ מיד לתודעתכם, כך התגובה הדתית המתבקשת והמקובלת־חברתית היא "נסתרות דרכי האל, וזוהי יוהרה להניח שאנו מסוגלים להבינן." לכן, אם תחשבו שהאסטרטגיה *הכללית* שעליכם לאמץ היא "השתמש בתערו של אוקאם," הדבר יהיה שקול לאדם דתי שיאמר שהאסטרטגיה הכללית היא להאמין.

"אבל – אבל תערו של אוקאם באמת טוב יותר מאמונה! זה לא כמו להעדיף טעם שונה של גלידה! כל אדם יכול לראות, מתוך התבוננות בהיסטוריה, שחשיבה אוקאמיאנית הייתה פרודוקטיבית הרבה יותר מאמונה –"

כל זה נכון. אך לא זה העניין. העניין הוא שכאשר אתם אומרים

---

* תכונות רבות (כמו IQ) מתפלגות באוכלוסיה *בהתפלגות נורמלית* – גרף בצורת פעמון. במאמר ידוע משנת 2006, *The Camel Has Two Humps* מאת דֶהנָאדי ובּוֹרנַט, נטען כי ההצלחה במבחן מסוים של חשיבה תכנותית מתפלגת בשני שיאים – "גמל עם שתי דבשות" – ולא בעקומת פעמון. כלומר: חלק מהאוכלוסיה יצליחו במבחן והחלק האחר ייכשלו בו כמעט לגמרי, וכמעט אין אנשים באמצע.

זאת, אתם מדקלמים הצדקה סטנדרטית שכבר קיימת במוחכם. האתגר במשבר אמונה הוא להתמודד עם מצב שבו, אולי, המסקנות וההצדקות הסטנדרטיות שלנו *שגויות*. לכן, אם ההצדקה הסטנדרטית שלכם עבור **א** היא "תערו של אוקאם!" וברצונכם לחולל משבר אמונה סביב **א**, עליכם לשאול את עצמכם האם תערו של אוקאם אכן תומך ב-**א**, האם אתם בכלל מבינים נכון את תערו של אוקאם, וכן – אם תרצו ספקות עמוקים באמת – האם פשטות היא בכלל הקריטריון שעבד היטב מבחינה היסטורית במקרה זה, או האם סביר לצפות שיעבוד, וכן הלאה. אם אתם מייעצים לאדם דתי להטיל ספק ב"אמונה" כהצדקה ל-**א**, עליכם להשקיע מאמץ כביר באותה המידה כדי לפקפק ב"תערו של אוקאם" כהצדקה ל-**א**.

(חשבו על כל אותם אנשים שאינם מבינים את ניסוחי תערו של אוקאם המבוססים על *אורך תיאור מינימלי* או על *אינדוקציית סולומונוף*,* וסבורים שתערו של אוקאם פוסל את *פרשנות העולמות המרובים* או את *השערת הסימולציה*.** הם יידרשו להטיל ספק בניסוחים שלהם לתערו של אוקאם, ובתפיסותיהם בנוגע לשאלה מדוע פשטות היא דבר טוב. יהא אשר יהא משפט **א** שהצדקתם זה עתה באמירה "תערו של אוקאם!", אני מוכן להתערב שאין זה ניצחון אוקאמיאני מוחץ וברור כמו כוח המשיכה.)

אם "תערו של אוקאם!" הוא התגובה הרגילה שלכם, התגובה הסטנדרטית שלכם, התגובה שכל חבריכם נותנים – הרי שמוטב לכם לעצור את מוחכם מהשלמה מיידית של התבנית הזו, אם אכן בכוונתכם לעורר משבר אמונה אמיתי.

מוטב לחשוב על כללים כגון: "דמיינו מה אדם ספקן היה אומר – ואז דמיינו מה הוא היה עונה לתגובה שלכם – ואז דמיינו מה עוד הוא

---

* ראו פרק 25.

** לפי *פרשנות העולמות המרובים* (many worlds theory) של מכניקת הקוונטים, פונקציית הגל אינה קורסת כאשר צופים בה, אלא כל הענפים של הסופרפוזיציה ממשיכים להתקיים כעולמות ממשיים במקביל. על פי *השערת הסימולציה* (simulation hypothesis), ייתכן שאנחנו חיים בסימולציה ממוחשבת שמריצה ציוויליזציה מתקדמת יותר.

עשוי לומר, שיהיה קשה יותר להשיב עליו."

או, "נסו לחשוב את המחשבה שכואבת יותר מכל."

ומעל הכל, הכלל: "השקיעו את אותה רמה של מאמץ נואש שהייתה נדרשת מאדם מאמין כדי לנטוש את דתו."

שכן אם *אינכם* מתאמצים עד כדי כך, הרי ש – *למיטב ידיעתכם* – הראש שלכם עלול להיות מלא בשטויות מגוחכות ממש כמו דת.

ללא מאמץ עוויתי ומטלטל להיות רציונליים – סוג המאמץ שדרוש כדי להשיל מעליכם דת – באיזו זכות אתם מעזים להאמין במשהו, בשעה שישראל אומן מאמין באלוהים?

מישהו (שכחתי מי זה היה) העיר פעם שלאנשים יש אפשרות לדחות מעליהם את אמונתם הדתית רק עד גיל מסוים. לאחר מכן, כבר יהיו בידיהם תשובות לכל ההתנגדויות, וזה יהיה מאוחר מדי. זהו סוג הקיום שעליכם להתעלות מעליו. זהו מבחן לכוחכם כרציונליסטים, והוא מחמיר מאוד. אך אם לא תוכלו לעמוד בו, תהיו חלשים יותר מילד בן עשר.

אך שוב, ברגע שבו אתם יודעים שאמונה היא טעות, היא כבר הובסה. איננו מדברים, אם כן, על מאמץ נואש ועוויתי לבטל את השפעותיו של חינוך דתי, *אחרי* שכבר הגעתם למסקנה שהדת שלכם שגויה. אנו מדברים על מאמץ נואש *להבין* האם עליכם להשליך מעליכם את הכבלים הללו או להותירם. הכנות העצמית הופכת לשברירית ביותר דווקא כשאיננו *יודעים* באיזה נתיב עלינו לבחור – זהו הרגע שבו רציונליזציות אינן נתפסות *מיד* כחטאים.

לא כל ספק דורש בימוי של משבר אמונה כולל. אך עליכם לשקול זאת כאשר:

- אמונה קיימת במוחכם זמן רב.
- היא מוקפת בענן של טיעונים והפרכות מוכרים.
- צברתם סביבה עלויות שקועות (זמן, כסף, הצהרות פומביות).
- לאמונה יש השלכות רגשיות (שימו לב: עובדה זו כשלעצמה אינה הופכת אותה לשגויה).
- היא נשזרה באישיותכם באופן כללי.

אף אחת מנורות האזהרה הללו אינה בבחינת הפרכה מיידית. תכונות אלו חושפות את האמונה למגוון רחב של סכנות, והופכות את דחייתה לקשה מאוד במקרה שהיא *אכן* שגויה. אולם, הן נכונות הן לגבי אמונתו של ריצ׳רד דוקינס בביולוגיה אבולוציונית, והן לגבי האמונה הקתולית של האפיפיור. אין משמעות הדבר שאנו עוסקים אך ורק בטעמים שונים של גלידה. רק בורים חושבים שכל האמונות האיתנות נמצאות באותו המעמד ללא קשר לראיות התומכות בהן, אך ורק משום שהן איתנות. המטרה איננה להחזיק באמונות שטחיות, אלא להחזיק במפה המתארת את השטח.

אני מדגיש זאת, כמובן, כדי שתוכלו להודות בפני עצמכם ״לאמונה שלי יש את נורות האזהרה האלה,״ מבלי שתיאלצו לומר לעצמכם בהכרח ״האמונה שלי שגויה.״

אך מה שנורות האזהרה הללו *כן* מסמנות, הוא אמונה שתדרוש *מאמץ מיוחד כדי להטיל בה ספק באופן אפקטיבי* – כך שאם היא אכן שגויה, תדחו אותה הלכה למעשה. ובמקום שבו אינכם מסוגלים להטיל ספק ביעילות, אתם עיוורים, משום שמוחכם יאחז באמונה ללא תנאי. כאשר רשתית העין משדרת את אותו האות בדיוק ללא קשר לפוטונים הנכנסים אליה, אנו מכנים זאת עיוורון.

מתי עליכם ליזום משבר אמונה?

שוב, חשבו על העצה שהייתם נותנים לאדם דתי: אם אתם מרגישים מעט מעורערים מבפנים, אך מנסים לעשות רציונליזציה ולמצוא סיבות לכך שהאמונה עדיין יציבה, ככל הנראה עליכם ליזום משבר אמונה. אם האמונה מבוססת בצורה מוצקה כמו כוח המשיכה, אינכם צריכים לטרוח – אך חישבו על כל אותם המאמינים שהיו מייחלים נואשות להסיק שאלוהים הוא עובדה מוצקה כמו כוח המשיכה. לכן, נסו לדמיין מה הספקנים שם בחוץ היו משיבים לטיעון ה״מוצק כמו כוח המשיכה״ שלכם. כמובן, אחת הסיבות לכך שאתם עשויים להיכשל ביצירת משבר אמונה היא שמעולם לא באמת עצרתם כדי להטיל ספק מלכתחילה – שמעולם לא אמרתם: ״הנה משהו שאני צריך להשקיע מאמץ כדי לפקפק בו כמו שצריך.״

אם מחשבותיכם הופכות מורכבות עד כדי כך, עליכם לגשת וליזום משבר אמונה. אל תנסו לעשות זאת באקראי, אל תנסו זאת ברגע פנוי מזדמן. אל תמהרו לסיים עם זה, רק כדי שתוכלו לומר "הטלתי ספק, כפי שהייתי מחויב לעשות." הרי גישה זו לא הייתה עובדת עבור מאמין דתי, והיא לא תעבוד גם עבורכם. נוחו היטב ביום שלפני, כדי שתהיו במיטבכם מבחינה מנטלית. פנו לעצמכם כמה שעות ללא הפרעות. מצאו מקום שקט לשבת בו. נקו את מוחכם מכל הטיעונים הסטנדרטיים, ונסו לראות הכל מהתחלה. השקיעו מאמץ נואש כדי להעלות ספק אמיתי – ספק שישמיד אמונה עמוקה שקרית, ורק *שקרית*.

אלמנטים מטכניקת "משבר האמונה" פוזרו לאורך פרקים רבים:

- **התחמקות מנקודות התורפה האמיתיות של אמונתכם:** אחד הפיתויים הראשונים במשבר אמונה הוא לתקוף את הנקודות החזקות ביותר של אמונתכם, כדי שתוכלו לדקלם את התשובות הטובות שבאמתחתכם. עליכם לחפש דווקא את הנקודות הכואבות ביותר, ולא את הטיעונים שהכי מרגיע להרהר בהם.
- **מדיטציית סקרנות:** רוג'ר זילאזני ערך פעם הבחנה בין "לרצות להיות סופר" לבין "לרצות לכתוב", ובאופן דומה קיימת הבחנה בין הרצון להיות מי שכבר חקר, לבין הרצון לחקור. לא די לומר "חובה עליי לבקר את האמונות שלי." עליכם להיות סקרנים, ורק אי־ודאות יכולה לייצר סקרנות. זיכרו את *חוק שימור הראיות הצפויות* – הוא יעזור לכם להתעדכן בהדרגה: עבור כל נקודה *בודדת* שאתם שוקלים, כל פיסת טיעון חדש וראיה חדשה, אינכם אמורים לצפות שאמונותיכם יזוזו (בממוצע) לכיוון אחד יותר מאשר לכיוון

השני – כך שתוכלו להיות סקרנים באמת בכל פעם לגבי הכיוון שאליו ייטו הדברים.

- **מבט מקורי:** השתמשו בטכניקה של פירסיג כדי למנוע ממחשבות סטנדרטיות ומוכנות־מראש לפלוש פנימה ולהשלים את התבנית.
- **המנטרה של גנדלין והמנטרה של טרסקי:** בני אדם יכולים לשאת את האמת, שהרי הם מתמודדים איתה ממילא. אם אמונה היא נכונה, אני רוצה להאמין בה, ואם היא שגויה, אני רוצה שלא להאמין בה. הייתם מייעצים לאדם דתי לנסות ולדמיין, במלוא העוצמה והעומק, עולם שבו אין אלוהים, ולהגיע – ללא תירוצים – להבנה המלאה *שאם* אין אלוהים, *אז* מוטב יהיה לו להאמין שאין אלוהים. אם אדם אינו מסוגל לקבל זאת ברמה רגשית עמוקה, הוא לא יהיה מסוגל לחוות משבר אמונה. לכן, עליכם להשקיע מאמץ כן לדמיין את *החלופה* לאמונתכם, בדיוק כפי שהספקן הטוב והדגול ביותר היה רוצה שתדמיינו אותה. חִשבו על המאמץ שיידרש ממאמין כדי לדמיין – מבלי לעוות לשם הנוחות – את השקפת עולמו של האתאיסט.
- **השקיעו מאמץ יוצא דופן:** ראו את המושג *אִישׁוּקֶנְמֶיי* – המאמץ הנואש והעוויתי להיות רציונליים.* המאמץ שדרוש כדי להתעלות מעל רמתם של ישראל אומן וכל המדענים הדגולים לאורך ההיסטוריה שמעולם לא הרפו מהדת שלהם.
- **ההיוריסטיקה הגנטית:** עליכם לחשוד מאוד אם יש לכם רעיונות רבים שהוצעו על ידי מקור שידוע לכם כעת שהוא אינו אמין, אבל – הפלא ופלא – נראה שכל הרעיונות הללו עדיין התבררו כנכונים. (למשל, אותו אדם שמודה שהתנ״ך נכתב בידי אדם,

* המושג נידון בפרק 309, בכרך השישי של *רציונליות: מ־א ועד ת*.

אך עדיין נאחז ברעיון שהוא מכיל חוכמה מוסרית שאין לה תחליף.)

- **חשיבותה של המילה "אופס":** באמת פחות כואב לבלוע את הגלולה המרה כולה בלגימה איומה אחת.
- **חַדְחוֹשֵׁב:** היפוכו של הדוּחוֹשֵׁב. הבחינו במחשבות שאתם נרתעים מפניהן, אלו שמופיעות בזווית התודעה שלכם לרגע קט בטרם תסרבו לחשוב עליהן. אם תהיו מודעים למה שאינכם חושבים עליו, תוכלו לחשוב עליו.
- **מעגלי מוות רגשיים (והתנגדות למעגל המוות המאושר):** מעגלי מוות רגשיים הם מחוללים ראשיים של אמונות שגויות, כאלה שיידרש משבר אמונה כדי לנער אותן. אך מכיוון שמעגלי מוות רגשיים יכולים להתחיל גם סביב דברים שהם אמיתיים וטובים באמת, אינכם חייבים להודות שאמונתכם שקרית כדי לנסות ולהתנגד לאפקט ההילה בכל שלב – סרבו לקבל שקרים שמציגים משהו באור חיובי, אפילו אם הוא באמת טוב. דיונים פוליטיים לא אמורים להרגיש חד־צדדיים.
- **המתינו לפני שתציעו פתרונות:** אל תציעו שום פתרון עד שהבעיה נדונה באופן יסודי ככל האפשר. הכריחו את מוחכם להמתין לפני שהוא קובע מה תהיה תשובתו. נסו זאת במשך חמש דקות לפני שאתם מרימים ידיים. זה נכון באופן כללי, ובמיוחד כאשר אתם מאמצים את נקודת המבט של פרקליט השטן.

והטכניקות הסטנדרטיות הללו רלוונטיות במיוחד:

- סדרת הפרקים על "השורה התחתונה" ועל רציונליזציה (חלק ז' בכרך זה), המסבירה מדוע זה אף פעם לא נכון לטעון באופן סלקטיבי לטובת צד אחד בלבד של הדיון.
- הטיה חיובית, ספקנות מגמתית ועצירה מוקדמת – פן תחפשו סלקטיבית אחר חיזוקים, תחפשו סלקטיבית אחר טיעוני־נגד לטיעוני־הנגד, ותעצרו את הדיון באופן סלקטיבי לפני שהוא הופך למסוכן. חלופות חסרות הן מקרה פרטי של עצירה

מוקדמת. מקרה פרטי של ספקנות מגמתית הוא ענווה מזויפת, שבה אתם מתוודים בביישנות ש"איש אינו יכול לדעת" דבר־מה שהייתם מעדיפים שלא לדעת. אל תדרשו סלקטיבית מטיעוני הנגד לעמוד ברף גבוה יותר מטיעוני הבעד.

- היזהרו מתמרורי עצור סמנטיים, נורות מחיאות כפיים, ומהבחירה שלכם האם להסביר, לסגוד או להתעלם.
- הרגישו את משקלם של פרטים מכבידים. כל פרט הוא משא בפני עצמו, קש שעשוי לשבור את גב הגמל.

אבל באמת שיש כמות גדולה למדי של חומר רלוונטי, כאן ובאתר *Overcoming Bias*. משבר אמונה הוא רק נקודת הקצה וההתנגשות הפתאומית של מאמץ *הָאִישׁוֹקֶנְמֶיי* הארוך יותר – המאמץ הבלתי מתפשר, הנמשך חיים שלמים, להיות כה רציונליים עד שתתעלו מעל טעויות ארורות ומטופשות. זהו הרגע שבו אתם זוכים בהזדמנות להשתמש במיומנויות שעליהן התאמנתם במשך זמן כה רב, במלוא הכוח, נגד עצמכם.

אני מאחל לכם את מירב ההצלחה מול יריבכם. שיהיה לכם משבר נפלא!

130

הטקס

החדר שבו קיבל ג׳פריסַאי את פני אורחיו שאינם בֵּייסוּטסוּקַאי היה רשמי ומאופק, מרוהט ללא דופי, ועל טהרת הטעם השמרני ביותר. אור שמש ואוויר צח זרמו מבעד לסבכת כסף מלוטשת. כמה קצוות חדים הבהירו שקיר זה לא נועד להיפתח לעולם. הרצפה והקירות היו עשויים זכוכית, עבה מספיק כדי לעוות את המראה שניבט מבעדה, כה עמוקה עד כי לא היה משנה כלל מה עשוי להימצא תחתיה. על פני משטחי הזכוכית היו שריטות עדינות, תבניות חסרות משמעות מיוחדת, חרוטות כאילו בידי ילד בעל נטייה אמנותית (וכך אכן היה).

במקומות אחרים בביתו של ג׳פריסאי היו חדרים בסגנונות שונים, אך חדר זה, כך גילה, תאם את ציפיותיהם של רוב הזרים ממאסטר בייסיאני – והוא בחר שלא להעמידם על טעותם. שעשוע שקט זה היה, אחרי הכל, אחת מהנאות החיים הקטנות.

האורחת ישבה מולו, ברכיה נחות על הכרית ועקביה משוכים לאחור. היא נכחה כאן אך ורק בענייני הקנוניה שלה, ולבושה העיד על כך: בגד גוף ורוד והדוק, עשוי עור, שכיסה אף את ידיה בכפפות – עוטף אותה לכל אורכה עד לברדס שכיסה את ראשה ואת שיערה, אם כי פניה נותרו גלויים לחלוטין מתחתיו.

זו הייתה הסיבה שבגינה בחר ג׳פריסאי לקבל את פניה בחדר זה.

ג׳פריסאי שאף נשימה ארוכה ונשף. ״את *בטוחה*?״

״הו,״ השיבה, ״האם אני חייבת להיות *בטוחה לחלוטין* לפני

שעצתי תוכל לשנות את דעותיך? האם לא די בכך שאני מומחית בנושא, ואתה אינך כזה?״

זווית פיו של ג׳פריסאי התעקלה כלפי מעלה בחצי חיוך. ״איך *את* בכלל יודעת כל כך הרבה על הכללים? מעולם לא קיבלת אפילו אורך־פְּלאנְק של הכשרה רשמית.״

״אתה בכלל צריך לשאול?״ אמרה ביובש. ״אם יש משהו שאתם *הבייסוטסוקאי* אוהבים לדבר עליו בלי סוף, הרי אלו הסיבות שבגללן אתם עושים דברים.״

ג׳פריסאי נרתע בתוכו למחשבה על הניסיון לספוג רציונליות מתוך צפייה באנשים אחרים מדברים עליה —

״ואל תירתע ממני בתוכך ככה,״ היא נזפה. ״אינני מנסה להיות רציונליסטית בעצמי, אני רק מנסה לנצח רציונליסט בוויכוח. יש הבדל, כפי שאני בטוחה שאתה מסביר לתלמידים שלך.״

האם היא באמת מסוגלת לקרוא אותי כל כך טוב? ג׳פריסאי הביט החוצה מבעד לסבכת הכסף, אל אור השמש שהשתקף מצלע ההר המגולפת. תמיד, תמיד שטף אור השמש הזהוב הזה מדי יום ביומו, במקום הזה, הרחק מעל העננים. דבר שאינו משתנה, האור הזה. השמש הרחוקה שאותו אור ייצג, עתידה לדעוך ולהישרף בעוד חמישה מיליארד שנים, אך כעת, ברגע *זה*, השמש עדיין זורחת. ודבר זה לעולם לא יוכל להשתנות. לשם מה לייחל שדברים יישארו כמות שהם לנצח, כאשר המשאלה הזו כבר התגשמה באורח מוחלט ככל שמשאלה כלשהי יכולה להתגשם? הפרדוקס שבין הנצחי לחולף: רק מבעד לנקודת המבט השנייה יכול היה להתקיים המושג של התקדמות, או של אובדן.

״תמיד השאת לי עצה טובה,״ אמר ג׳פריסאי. ״בלתי־משתנה, כך זה היה. לאורך כל הזמן בו אנו מכירים.״

היא הרכינה את ראשה לאות הסכמה. הדבר היה נכון, ולא היה כל צורך לומר במפורש את השלכותיו.

״אם כך,״ המשיך ג׳פריסאי. ״לא לשם הוויכוח, אלא רק מפני שאני רוצה לדעת את התשובה. את *בטוחה*?״ הוא אפילו לא הבין

כיצד יכלה *לנחש*.

"די בטוחה," אמרה. "אנחנו אוספים סטטיסטיקות כבר זמן רב, ובתשע מאות ושמונים וחמישה מתוך אלף מקרים כמו שלך —"

ואז היא פרצה בצחוק למראה הבעת פניו. "לא, אני צוחקת. ברור שאינני בטוחה. זהו דבר שרק אתה יכול להכריע לגביו. אולם אני *כן* בטוחה שעליך ללכת ולעשות את מה שזה לא יהיה שאתם עושים — אני די בטוחה שיש לכם טקס בשביל זה, גם אם אינכם דנים בו עם זרים — כשאתה *שוקל ברצינות תהומית* לנטוש הנחת יסוד קיומית שהחזקת בה זמן רב."

היה קשה להתווכח עם זה, הרהר ג'פריסאי, על אחת כמה וכמה כשמומחית לנושא אומרת לך שאתה ככל הנראה טועה.

"אני נכנע," אמר ג'פריסאי. כשיצא מפיו, המשפט נאמר בסופיות החלטית. *אין צורך להתווכח איתי יותר: ניצחת.*

"הו, תפסיק עם זה," פטרה אותו. היא התרוממה מן הכרית בתנועה חלקה אחת, ללא שמץ מאמץ מיותר. היא לא התהדרה בגילה, אך גם לא הסתירה אותו. היא לקחה את ידו המושטת והרימה אותה אל שפתיה לנשיקה רשמית. "חֲיֵה שלום, סנסיי."

"חֲיֵה שלום?" חזר ג'פריסאי. זו הייתה הצהרה שציינה פרידה מסדר גודל גבוה יותר מסתם 'להתראות'. "אני אכן מתכוון לבקר אותך שוב, גבירתי. ואת תמיד רצויה כאן."

היא צעדה לעבר הדלת מבלי להשיב. בפתח היא עצרה לרגע, מבלי להסתובב. "זה לא יהיה אותו הדבר," אמרה. ואז, מבלי שתנועותיה ייראו חפוזות ולו במעט, התרחקה בתנועה מהירה כל כך עד שדמתה כמעט להיעלמות.

ג'פריסאי נאנח. אבל לפחות מעתה ועד לאתגר עצמו, כל פעולותיו היו ידועות ומוכתבות מראש.

לאחר שעזב את אזור קבלת הפנים הרשמי, הוא הלך אל הזירה שלו והורה לשלוח שליחים לתלמידיו, כדי להודיע להם ששיעורי יום המחרת ייערכו במתכונת מאולתרת בהיעדרו, וכי בוחן ייערך במועד מאוחר יותר.

ואז הוא לא עשה שום דבר מיוחד. הוא קרא עוד מאה עמודים בספר הלימוד שהשאיל. הספר לא היה טוב במיוחד, אולם גם זה שמסר בתמורה לא היה כזה. הוא שוטט מחדר לחדר בביתו, בודק בבטלה מקומות אחסון שונים כדי לראות אם דבר־מה נגנב (חבילת קלפים הייתה חסרה, אבל זה היה הכל). מעת לעת נדדו מחשבותיו אל האתגר של יום המחרת, והוא הניח להן לשוטט. הוא נמנע לחלוטין מלכוון את מחשבותיו. הוא רק חסם כל מחשבה שכבר עלתה במוחו *בעבר*, ומנע כל סוג של מסקנה, או אפילו מחשבה על הכיוון שאליו מחשבותיו עשויות לנטות.

השמש שקעה, והוא צפה בה זמן מה כשמוחו מוחזק בקפידה בהילוך סרק. היה זה מפגן איזון פנטסטי – להעביר את המוח להילוך סרק כמעט מבלי לחשוב על כך, ומבלי להשקיע אנרגיה בשימור המצב הזה. לפני שנים הוא היה מזיע בשלב הזה, אך התרגול הוביל לשלמות כבר מזמן.

למחרת בבוקר הוא התעורר כשהכאוס של חלומות הלילה עדיין טרי במוחו. בעודו עושה את המיטב לשמר את תחושת הכאוס ואת זיכרונה, הוא ירד גרם מדרגות אחד, ואז גרם מדרגות נוסף, ואז עוד אחד, עד שלבסוף הגיע לחדר הכי פחות מעוצב בבית כולו.

הוא היה לבן. וזה פחות או יותר היה הכל, בכל הנוגע לסכֵמת הצבעים שלו.

לאורכו של קיר יחיד נתלו שלטים שאותם שרטט בעצמו ג׳פריסאי הצעיר יותר בקפידה רבה, בהתאם לשיטה הקלאסית והמומלצת, צורב את *העקרונות* אל תוך מוחו עם כל נגיעה של המכחול שכתב את המילים: *מה שהאמת יכולה להשמיד, ראוי שיושמד. בני אדם יכולים לשאת את האמת, שהרי הם מתמודדים איתה ממילא. סקרנות שואפת להשמיד את עצמה.* היה שם אפילו שלט קטן אחד שלא הציג דבר מלבד קו אופקי אדום. אפשר לגרום לסמלים לייצג *כל דבר*. זוהי דוגמה לגמישות הכוח החזותי שאפילו *הקנוניה הבָּאוּדית* הייתה נרתעת מלהודות בה בגלוי.

מתחת לשלטים נשרטו בקיר שתי עמודות של סימוני ספירה. תחת

עמודת הפלוס, שני סימנים. תחת עמודת המינוס, חמישה סימנים. שבע פעמים הוא נכנס לחדר הזה. חמש פעמים החליט שלא לשנות את דעתו. פעמיים הוא יצא ממנו כאדם מעט אחר. לא היה יחס שנקבע מראש, או טווח מוגדר – דבר כזה אכן היה הופך את העניין לנלעג. אולם, אם בחלוף הזמן לא הופיעו סימונים בעמודת הפלוס, יכולת באותה מידה להודות שאין טעם להחזיק בחדר כזה, משום שאין לך את היכולת שהוא מייצג. או זה, או שנולדת כשהאמת והצדק לגבי הכל כבר נתונים בידיך.

ג׳פריסאי התיישב, לא כשפניו אל השלטים, אלא כשגבו אליהם, פונה אל הקיר הלבן והריקני. עדיף היה להיות נטול הסחות דעת חזותיות. במוחו, הוא תרגל תחילה את המטא־מְנֶמוֹניקה,* ולאחריה את תת־המנמוניקות המקושרות אליה, עבור שבעת העקרונות המרכזיים ושישים ושתיים הטכניקות הספציפיות שהיו בעלות הסבירות הגבוהה ביותר להתגלות כנחוצות ב״טקס שינוי הדעה.״ על כך הוסיף ג׳פריסאי מנמוניקה נוספת, המזכירה לו את ארבע־עשרה הטעויות המביכות ביותר שלו עצמו.

הוא לא שאף נשימה עמוקה. נשימה רגילה עדיפה.

ואז הוא הציג לעצמו את השאלה.

* *מְנֶמוֹניקה* (mnemonic) היא טכניקה לשינון ושליפה מהירה של מידע מהזיכרון בעזרת תבניות (כגון ראשי־תיבות או אסוציאציות).