

# פילוסופיה של הנפש בתקופתנו

יחידות 6-7: פונקציונליזם ומודל  
המחשב של הנפש

אמיר הורוביץ

10679

מהדורה זמנית

פנימי • לא להפצה ולא למכירה

אוגוסט 2009

מק"ט 10679-1114

**כותב** פרופ' אמיר הורוביץ

**עורכת** רות רמות

#### **יועצים**

פרופ' עדי צמח – האוניברסיטה העברית  
פרופ' אורון שגריר – האוניברסיטה העברית  
ד"ר מאיר חמו – אוניברסיטת חיפה  
ד"ר מריס כהן – האוניברסיטה הפתוחה  
ד"ר יקיר לוי – אוניברסיטת בן גוריון

**עימוד** עינב צדוק

הדפסה דיגיטלית מתוקנת (II) – אוגוסט 2009

© תשס"ט – 2009. כל הזכויות שמורות לאוניברסיטה הפתוחה.  
בית ההוצאה לאור של האוניברסיטה הפתוחה, הקריה ע"ש דורותי דה רוטשילד, דרך האוניברסיטה 1, ת"ד 808, רעננה 43107.  
The Open University of Israel, The Dorothy de Rothschild Campus, 1 University Road, P.O.Box 808, Raanana 43107.  
Printed in Israel.  
אין לשכפל, להעתיק, לצלם, להקליט, לתרגם, לאחסן במאגר מידע, לשדר או לקלוט בכל דרך או בכל אמצעי אלקטרוני, אופטי, מכני או אחר כל חלק שהוא מהחומר שבספר זה. שימוש מסחרי בחומר הכלול בספר זה אסור בהחלט, אלא ברשות מפורשת ובכתב ממדור זכויות יוצרים של האוניברסיטה הפתוחה.

סדרת ספרים זו מוקדשת באהבה

לאבי ראובן הורוביץ

ולזכר אמי ורדה הורוביץ

אמיר הורוביץ, דצמבר 2008



## תוכן העניינים

7	1. הקדמה
8	1.1 פיזיקליזם התנהגותני
8	1.1.1 ריבוי מימושים וביהביריזם
9	1.1.2 מביהביריזם לפיזיקליזם התנהגותני
21	2. יסודות ההשקפה הפונקציונליסטית
29	2.1 פונקציונליזם ומעגליות
38	2.2 פונקציונליזם-פרט ופונקציונליזם-סוג
41	2.3 פונקציונליזם ופיזיקליזם, והבעיה הפסיכופיזית
50	3. פונקציונליזם ומודל המחשב של הנפש
56	3.1 מודל המחשב וקשרים סיבתיים – תמיהות אחדות
60	3.2 פונקציונליזם ומכונת טיורינג
60	3.2.1 מכונת טיורינג ופעולתה
70	3.2.2 מימושים פיזיקליים של מכונת טיורינג, ואפיון מצבי תוכנה
70	3.2.2.1 המימוש של תוכנה ומצביה
73	3.2.2.2 אפיונו של מצב תוכנה (מצב חישובי)
75	3.2.2.3 המימוש הפיזיקלי של מצבים חישוביים
79	3.2.3 זהות בקשרים סיבתיים, ומענה לתמיהות
81	3.2.4 הנפש כמכונת טיורינג
86	4. שתי גרסאות של פונקציונליזם
86	4.1 פונקציונליזם עממי (או אנליטי)
86	4.1.1 הפסיכולוגיה העממית
90	4.1.2 הפונקציונליזם העממי
95	4.1.3 פונקציונליזם עממי וידיעת טענות הפסיכולוגיה העממית
102	4.2 פונקציונליזם אמפירי
102	4.2.1 טיעונו של פטנאם בזכות ההשקפה הפונקציונליסטית
105	4.2.2 מהותו של הפונקציונליזם האמפירי
110	4.3 השוואות נוספות בין פונקציונליזם עממי לפונקציונליזם אמפירי
113	5. אפיון הקלט והפלט על פי הפונקציונליזם; ליברליזם ושובניזם
121	6. בעיית זהות התוכנה, ופונקציונליזם הומונקולארי
121	6.1 בעיית זהות התוכנה
124	6.2 פונקציונליזם הומונקולארי

126	7. מבחן טיורינג
134	8. טיעון החדר הסיני כנגד ההשקפה הפונקציונליסטית
135	8.1 החדר הסיני ומבחן טיורינג
136	8.2 התוכנה הנכונה, הבנה ומכוונות
140	8.3 ביקורת טיעון החדר הסיני – מכוונות, מחשב ורובוט
143	9. פונקציונליזם ו-qualia
144	9.1 טיעון הספקטרום המהופך
147	9.2 טיעון ה-qualia הנעדרות והמוח הגדול של סין
150	10. מילות סיכום
153	נספח: "ג'ון ר' סרל, "נפש, מוח ותוכנות מחשב"

## 1. הקדמה

התורה הפונקציונליסטית של הנפש היא, כפי הנראה, תורת הנפש המקובלת ביותר על ידי פילוסופים בני-זמננו. הפילוסוף האמריקני וויליאם לייקאן (William Lycan) ביטא את הערכתו לתורה זו באמרו

פונקציונליזם היא הדוקטרינה היחידה בכל הפילוסופיה שאני מוכן (אם לא מורשה) להרוג עבורה.<sup>1</sup>

פילוסוף אמריקני אחר, סטפן שיפר (Stephen Schiffer), תיאר את הרעיון הפונקציונליסטי כ"אחד הרעיונות רבי-ההשפעה ביותר שפותחו אי-פעם בפילוסופיה של הנפש".

חשיבותה של התורה הפונקציונליסטית – שראשית התפתחותה בשנות ה-60 של המאה העשרים – חורגת מתפקידה בהבהרת טבעם של המצבים הנפשיים: רבים רואים בה, או בפיתוח מסוים שלה, את התיאוריה העומדת בבסיסו של זרם מרכזי של המדעים הקוגניטיביים; לחלק מן הנושאים שתורה זו העלתה לדיון השלכות חשובות בתחומים פילוסופיים שונים, כמו הפילוסופיה של המדע והפילוסופיה של הלשון; והתפתחותה הייתה קשורה להתפתחויות אחדות בלוגיקה ובמתמטיקה. עיקר ענייננו בתורה זו ייסוב, כמובן, על אודות תרומתה לפילוסופיה של הנפש. במובן חשוב, תורה זו מהווה מהפכה מושגית – מהפכה בדרך החשיבה על טבעו של היסוד הנפשי ועל שאלת מיקומו בעולם. דרך אחת לעמוד בקצרה על המהפכה הכרוכה בפונקציונליזם היא להזכיר את הקביעה הקשורה במודל המחשב של הנפש, מודל המהווה היבט מהותי של תורה זו: היחס בין הנפש לגוף (או, באופן ספציפי, למוח) הוא כמו היחס בין התוכנה של מחשב לבין החומרה שלו; הנפש היא התוכנה של הגוף (או של המוח). כמובן, בבוא הזמן נסביר בהרחבה את פשרה של קביעה זו.

הפונקציונליזם, אפשר לומר, מנסה לקחת את הטוב מכל העולמות: לשמור על קשר הדוק בין היסוד הנפשי לבין ההתנהגות, אך עם זאת לא לוותר על הרעיון שיסוד זה כרוך במצבים פנימיים – שמצבים נפשיים הם מצבים פנימיים בעלי כוחות סיבתיים, ולעמוד על כך שהקשר בין היסוד הנפשי לבין ההתנהגות הוא סבוך ואינו ניתן לתיאור במונחים של מתאמי גירוי-תגובה. קסמה של תורה זו רב, ואמנם היא דחקה את רגלה של התורה של פיזיקליזם-סוג ממרכז הבמה הפילוסופית.

התורה הפונקציונליסטית שאבה ממקורות שונים, וניתן להציגה באופנים שונים. אנו נתחיל להציגה באמצעות הקשרים שבינה לבין עמדות פיזיקליסטיות, מחד גיסא, ובינה לבין העמדה הביהביוריסטית, מאידך גיסא.

W.G. Lycan, *Consciousness*, Cambridge (Mass.): Cambridge University Press, 1987, p. 37.

1

## 1.1 פיזיקליזם התנהגותני

### 1.1.1 ריבוי מימושים וביהיוריזם

ניזכר בטיעון של ריבוי המימושים אשר הועלה כנגד תורת-הזהות הסוגית. על פי טיעון זה, הימצאות במצבים נפשיים מאותו סוג – כאבי וכאבו של תמנון, או האמנתי שיורד גשם לעומת האמנתו של חזר שיורד גשם – אינה מחייבת הימצאות במצבים זהים מבחינה פיזיקלית, במצבים מאותו סוג פיזיקלי. אדרבא, הימצאות במצבים נפשיים זה-סוג מתיישבת עם שונות פיזיקלית גדולה. ניתן אם כך לומר, שבמידה רבה, לפי טיעון זה, אם עניינו בשאלה **באילו** מצבים נפשיים נמצא יצור נפשי כלשהו, אין חשיבות לשאלה ממה, מאיזה חומר, "עשוי" יצור נפשי זה; וניתן להוסיף ולומר, שגם אם עניינו בשאלה **האם** יצור מסוים הוא יצור נפשי, הרי, במידה רבה, אין חשיבות לשאלה ממה "עשוי" יצור זה (אנו מסייגים את הטענה באמצעות המילים "במידה רבה" משום שטענת ריבוי המימושים אין משמעותה שיצור נפשי יכול להיות עשוי מכל חומר שהוא **ללא כל מגבלה**. אנו נבהיר נקודה זו בהמשך).

#### שאלה

הרעיון בדבר אי-הרלוונטיות של החומר הספציפי ממנו עשויים יצורים נפשיים לחייהם הנפשיים עשוי להתקבל באהדה על ידי פילוסופים ביהיוריסטיים. מדוע?

ובכן, לפי הביהיוריסטים, כזכור, כאשר אנו מייחסים ליצור "מצב" נפשי כלשהו – ליתר דיוק, כאשר אנו מחילים על יצור **מושג** נפשי כלשהו – איננו מייחסים לו כל מאפיין פנימי. פנימיותו עשויה להיות עברו "קופסה אטומה", דהיינו, אנו עשויים שלא לדעת עליה דבר, ובכל-זאת אנו עשויים להחיל על היצור את המושגים הנפשיים שאמנם חלים עליו (הוא כועס, הוא משועשע, הוא מאמין שאלוהים קיים, וכדומה). מן הסתם, החלת מושגים נפשיים על יצורים היא ניטרלית באשר **לטבעו של החומר** שממנו הם עשויים. שאלת חלותו של מושג נפשי על יצור מועמדת במלואה על השאלה, אם היצור הנדון מאופיין על ידי הדגם ההתנהגותי ההולם, או הנטייה ההתנהגותית ההולמת. מה שחשוב הוא "מה שבחוץ", דהיינו הקלט (התפיסתי) והפלט (ההתנהגותי), והמתאמים בין קלטים ופלטים, ולא "מה שבפנים".

גם קביעה זו עלינו לסייג, משום שכפי שרמזנו ביחידה 3, יש הסוברים שלא ניתן לאפיין התנהגות מבלי לאפיין את חלקי הגוף הכרוכים בהתנהגות **באופן פיזיקלי**, כלומר לאפיינם במונחי החומר שממנו הם עשויים. נשוב ונדון בשאלת אפיונה של התנהגות בהמשך היחידה. אולם גם אם האפיון של התנהגות חייב להיכרך באפיון פיזיקלי, הרי שלפי הביהיוריסטים "מה שבפנים" אינו חשוב – הרכבו הפיזיקלי של המנגנון הפנימי שאחראי לקיומם של דגמי התנהגות ונטיות התנהגותיות (מנגנון שבמקרה האנושי ובמקרים רבים אחרים הוא כמובן



המוח) הוא שאינו חשוב לשאלת חלותם של המושגים הנפשיים לפי הביהביוריסטים – רק דגמי ההתנהגות חשובים.

העמדה הביהביוריסטית מתיישבת אפוא עם טענת ריבוי המימושים, ונראה שפילוסופים ביהביוריסטיים עשויים לדגול ברעיון הכרוך בטענת ריבוי המימושים, הרעיון בדבר אי-הרלוונטיות של החומר הספציפי ממנו עשויים יצורים נפשיים לחייהם הנפשיים.<sup>2</sup>

מדוע אנו מזכירים את הקשר בין טענת ריבוי המימושים לבין העמדה הביהביוריסטית? טענת ריבוי המימושים עניינה בשאלת הסיווג של המצבים הנפשיים, ובמילים אחרות, בשאלת התכונות הנפשיות. טענה זו מאיימת על קריטריון הסיווג הפיזיקליסטי של המצבים הנפשיים, הקריטריון של התורה של פיזיקליזם-סוג, על פיו מצבים נפשיים שייכים לאותו סוג (שניהם כאבים, שניהם ההאמנה שהשמש זורחת, וכדומה) אם הם חולקים באופן ייחודי תכונה פיזיקלית.

### 1.1.2 מביהביורזם לפיזיקליזם התנהגותני

גם העמדה הביהביוריסטית מספקת תשובה – אחרת, כמובן – לשאלת התכונות הנפשיות, שאלת הסיווג הנפשי. על פיה, נכון להחיל על שני יצורים אותו מושג נפשי – לייחס להם אותה תכונה נפשית – אם שניהם מתאפיינים על ידי אותה נטייה התנהגותית או אותו דגם התנהגות; דהיינו, אם הם מתנהגים באופנים דומים במצבים דומים – מגיבים באופנים דומים על גירויים דומים. ועתה, אם אמנם קריטריון הסיווג הנפשי של הביהביוריסטים מכבד, כפי שצינו, את טענת ריבוי המימושים, הרי פילוסופים הדוגלים בתורת הזהות ומתרשמים שאכן טענת ריבוי המימושים מהווה מגבלה מהותית על סיווגם של המצבים הנפשיים, מגבלה שכל תורה של הנפש חייבת לכבד, יכולים לשקול את האפשרות הבאה:

**מצבים נפשיים הם מצבים מוחיים, אולם קריטריון הסיווג של המצבים הנפשיים הוא התנהגותני. במילים אחרות, שני מצבים מוחיים שהם מצבים נפשיים מהווים מצבים נפשיים מאותו סוג, אם, ורק אם, הם כרוכים באותם דגמי התנהגות ונטיות התנהגותיות.**

השקפה זו מהווה שילוב של תורת הזהות, מחד גיסא, ושל יסוד מן התורה הביהביוריסטית, מאידך גיסא. נתייחס אליה, לעת עתה, כאל "ההשקפה המעורבת הפיזיקליסטית-התנהגותנית", או, בקיצור, "פיזיקליזם התנהגותני". חשוב להדגיש שהשקפה זו מקבלת רק

<sup>2</sup> כשאנו אומרים שהעמדה הביהביוריסטית מתיישבת עם טענת ריבוי המימושים, אנו מבינים את טענת ריבוי המימושים של המצבים הנפשיים כך שעל פיה הימצאות במצבים נפשיים מאותו הסוג מתיישבת עם הימצאות במצבים שונים מבחינה פיזיקלית **במובן החלש** של "מצבים", מובן שבו עצם חלותה של תכונה כלשהי על אובייקט כלשהו בזמן כלשהו משמעו שאובייקט זה נמצא במצב מסוים בזמן זה. הטענה שמצב נפשי, **במובן החזק** של "מצב", עשוי להתממש על-ידי מצבים פיזיקליים מסוגים שונים, אינה יכולה להתקבל על ידי הביהביוריסטים, משום שעל פיהם אין מצבים נפשיים במובן החזק. במובן זה, היות בעל נטייה מסוימת אינה מהווה הימצאות במצב מסוים, ועל פי הביהביורזם חלותו של מונח נפשי על יצור נפשי משמעה היותו של היצור בעל נטייה מסוימת ותו לא. האונטולוגיה הנפשית על פי הביהביורזם אינה אונטולוגיה של מצבים ואירועים במובן החזק (דנו בשני המובנים של "מצבים" בסעיף 3.5 ביחידה 3).

יסוד מסוים של התורה הביהיוריסטית, ואין היא זהה לתורה זו. השקפה זו דוחה את הטענה הביהיוריסטית כי האונטולוגיה הנפשית איננה אונטולוגיה של מצבים ואירועים, והיא מסכימה בנקודה זו עם תורת הזהות, לפיה מצבים ואירועים נפשיים הם מצבים ואירועים מוחיים. הבדל זה שבין השקפה מעורבת זו לבין התורה הביהיוריסטית קשור בהבדל נוסף ביניהן, הבדל שנוכל לעמוד עליו כשנבהיר את פשרה של הטענה כי מצבים נפשיים הם זהו סוג, אם הם **כרוכים** באותם דגמי התנהגות ונטיות התנהגותיות. מה פירוש "כרוכים" בהקשר זה?

#### שאלה

הייתכן, לפי ההשקפה המעורבת, שמונחים נפשיים כמו "כאב", "האמנה שיורד גשם", "ראיית אדום", וכדומה, מציינים דגמי התנהגות ונטיות התנהגותיות?

התשובה לשאלה זו שלילית, משום שלפי ההשקפה המעורבת מונחים כאלו מציינים **מצבים פנימיים**, מצבי-מוח.

#### שאלה

מה, אם כן, יכול להיות הקשר בין המצבים הנפשיים לבין דגמי ההתנהגות אשר קובעים את סיווגם?

ובכן, הקשר הוא **סיבתי**. המצבים הנפשיים אחראים מבחינה סיבתית לקיומם של דגמי ההתנהגות הנדונים. הטענה שמצב נפשי מסוים כרוך בדגם התנהגות מסוים – דגם המבטא מתאם בין גירויים תפיסתיים מסוג מסוים לבין תגובות התנהגותיות מסוג מסוים – משמעה הוא שמצב נפשי זה הוא **המתווך הסיבתי** שבין הגירויים והתגובות הללו. מצב זה הוא מצב הנוטה **להיגרם** על ידי הגירוי הנדון **ולגרם** את התגובה הנדונה.

גירוי X -----> תגובה Y

אפיון נפשי

#### תרשים 1. ביהיוריסם

גירוי X -----> מצב נפשי Z -----> תגובה Y

מצב נפשי כמתווך סיבתי

#### תרשים 2. ההשקפה המעורבת

כמובן, מה שמאפשר להשקפה המעורבת הפיזיקליסטית-התנהגותנית להתייחס למצבים הנפשיים כאל מתווכים סיבתיים זו בדיוק העובדה שהאונטולוגיה הנפשית על פיה היא אונטולוגיה של מצבים ואירועים (במובן החזק) – ספציפית, אלו הם מצבים פנימיים (ובאופן ספציפי יותר, מצבים מוחיים). כך, השקפה זו נמנעת מן הקושי שבו נתקלה, לפי מבקריה, העמדה הביהביוריסטית – אי-יכולתה לכבד את היות המונחים הנפשיים מונחים המציינים ישויות בעלות כוח סיבתי, ואת היותם של הסברים פסיכולוגיים של התנהגות הסברים סיבתיים.

אולם, כאמור, ההשקפה המעורבת הנדונה חולקת יסוד חשוב עם התורה הביהביוריסטית: קריטריון הסיווג של המצבים הנפשיים על פיה הוא קריטריון ביהביוריסטי: מה שמשיך, על פיה, מצב נפשי לסוג הנפשי שלו (כאב, האמנה שיורד גשם, וכדומה) זהו דגם ההתנהגות הספציפי שבו מצב נפשי זה כרוך, כלומר דגם ההתנהגות שמצב נפשי זה הוא המתווך הסיבתי שלו. ומה שנדרש, על פיה, משני מצבים נפשיים כדי שיהיו מצבים נפשיים השייכים **לאותו** סוג נפשי הוא ששניהם יהיו כרוכים – במובן סיבתי זה – באותו דגם התנהגותי.

קריטריון הסיווג של השקפה פיזיקליסטית זו, כפי שאנו רואים, איננו מתייחס לטבעו החומרי של המצב הנפשי – זהו איננו קריטריון פיזיקליסטי.

## שאלה

ביחידה 4-5 הוצגו עמדות פיזיקליסטיות אחדות. איזו מהן הולמת את ההשקפה המעורבת?

ההשקפה המעורבת, המזהה מצבים נפשיים עם מצבים פיזיקליים, היא עמדה של פיזיקליזם-פרט, ובאופן ספציפי יותר, כיוון שאין היא דוגלת בקריטריון סיווג פיזיקליסטי, היא עמדה של **פיזיקליזם-פרט בלעדי**. אנו רואים, אם כן, דוגמה מוחשית לעמדה של פיזיקליזם-פרט בלעדי, עמדה שהאונטולוגיה הנפשית על פיה היא אונטולוגיה פיזיקליסטית, ואילו קריטריון הסיווג הנפשי על פיה איננו פיזיקליסטי.

אפשר לראות בבירור כי ההשקפה המעורבת הפיזיקליסטית-התנהגותנית מתיישבת עם תופעת ריבוי-המימושים. כיוון שקריטריון הסיווג של השקפה זו איננו קריטריון פיזיקליסטי כיוון שאין הוא מתייחס לטבעו החומרי של המצב הנפשי – הוא מאפשר למצבים שוני מבנה חומרי להיות מצבים השייכים לאותו סוג נפשי. לפי השקפה זו, יהא טבעו החומרי של מצב נפשי אשר יהא, כל עוד הוא כרוך בדגם ההתנהגות של כאב, הרי הוא כאב, וכל עוד הוא כרוך בדגם ההתנהגות של ההאמנה שיורד גשם, הרי שהוא ההאמנה שיורד גשם. השקפה זו מתיישבת עם האפשרות שהטבע החומרי של האמנתי שיורד גשם והטבע החומרי של האמנתו של יצור נפשי אחר שיורד גשם שונים מאוד זה מזה, שאלה הם מצבים חומרים שונים לחלוטין. אין כל חשיבות לטבע החומרי. ובכלל, מה שעושה מצב פנימי מסוים של אורגניזם

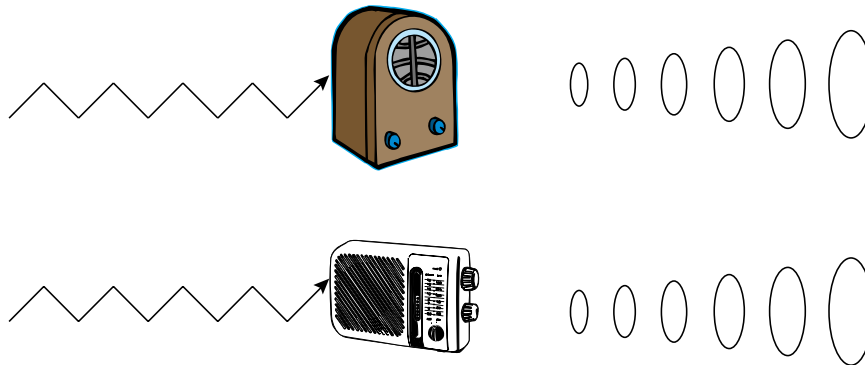
**למצב נפשי** אינו טבעו החומרי, אלא העובדה שהוא כרוך בדגם התנהגותי מסוים, ובמילים אחרות, שהוא אחראי מבחינה סיבתית לאותו מתאם בין גירויים לתגובות.

אפשר לתהות כיצד זה מאפשר קריטריון הסיווג הביהביוריסטי של ההשקפה המעורבת ריבוי מימושים – כיצד הוא מאפשר למצבים שונים מבחינה חומרית להשתייך לאותו סוג נפשי. הייתכן שמצבים בעלי מבנה חומרי שונה לחלוטין יהיו אחראים מבחינה סיבתית לאותו מתאם בין גירויים לתגובות? בעלי ההשקפה המעורבת הפיזיקליסטית-התנהגותנית יכולים להשיב לשאלה זו בשתי דרכים אפשריות (וקשורות זו בזו); נציג כעת דרך אחת ובשלב מאוחר יותר נציג את השנייה. התשובה הראשונה היא: מדוע עלינו להניח שהדבר לא ייתכן? ניתן לחשוב על דוגמאות שונות של עצמים העשויים מחומרים שונים לחלוטין ואשר מקיימים מתאמי גירוי-תגובה זהים.

### שאלה

התוכלו לחשוב על דוגמאות כאלו?

דוגמה אפשרית אחת היא של מכשירי רדיו שונים לחלוטין אשר בקולטם גלים אלקטרומגנטיים בתדירויות זהות (אלו הגירויים) מנפיקים צלילים זהים (אלו התגובות). בהמשך נדון בדוגמה חשובה אחרת של מכשירים שונים אשר מקיימים מתאמי גירוי-תגובה זהים.



### תרשים 3. מכשירי רדיו שונים בעלי מתאמי גירוי-תגובה זהים

עלינו להדגיש שעל פי ההשקפה המעורבת, הטענה האונטולוגית (הפיזיקליסטית) והטענה בדבר קריטריון הסיווג הנפשי (שהוא התנהגותני) הן טענות בעלות מעמד שונה. ההבדל במעמדן מתבטא בכך שהטענה בדבר קריטריון הסיווג הנפשי היא **טענה מושגית אנליטית**, ולפיכך הכרחית, ואילו הטענה האונטולוגית היא **טענה קונטינגנטית**, **טענה שבעובדה**. התייחסות נבדלת זו אל רכיביה השונים של ההשקפה המעורבת הפיזיקליסטית-התנהגותנית

משקפת את טבען של שתי העמדות שיסודותיהן מרכיבים השקפה זו: כזכור, העמדה הביהביוריסטית היא עמדה מושגית אנליטית, בעוד תורת הזהות הפיזיקליסטית היא עמדה קונטינגנטית. אולם לטבען השונה הזה של שתי העמדות הללו (הביהביורזם ותורת הזהות) יש חשיבות מיוחדת באשר להשקפה המעורבת. אפשר לעמוד על חשיבות זו לאור העובדה שטבע שונה זה עומד ביסודו של טיעון חשוב בזכות ההשקפה המעורבת. לפני שנציג טיעון זה, נקרא את הקטעים הללו מאת הפילוסוף דיוויד ארמסטרונג, שאותו פגשנו כבר ביחידה 4-5, ואשר אפשר לראותו כמייסדה של ההשקפה המעורבת הפיזיקליסטית-התנהגותית.<sup>3</sup>

הביהביורזם בוודאי שגוי, אבל אולי אין הוא שגוי לגמרי. אולי הביהביוריסטים שוגים בזהותם את הנפש ואת ההתרחשויות הנפשיות עם התנהגות, אבל אולי הם צודקים בחושבם שתפישתנו את הנפש ואת המצבים הנפשיים **קשורה קשר לוגי להתנהגות**. כי אולי מה שאנו מתכוונים לו ב"מצב נפשי" הוא איזשהו מצב של האדם אשר, בנסיבות מתאימות, **מביא** לטווח מסוים של התנהגויות. אולי ניתן להגדיר "נפש" לא כהתנהגות, אלא כגורם הפנימי של התנהגות מסוימת. מחשבה אינה דיבור בנסיבות מתאימות, אלא היא משהו באדם אשר, בנסיבות מתאימות, מביא לדיבור. ולמעשה, אני מאמין שזהו ההסבר האמתי, או בכל מקרה הסבר ראשון אמתי, של מה שאנו מתכוונים לו ב"מצב נפשי".<sup>4</sup>

אפשר לקרוא לניתוח המוצע "הניתוח הסיבתי של המושגים הנפשיים". לפי השקפה זו, מושג המצב הנפשי קשור באופן מהותי במושג של מצב אשר **נוטה להיות הסיבה לתוצאות מסוימות או אשר נוטה להיות התוצאה של סיבות מסוימות**, ומושג זה ממצה את מושג המצב הנפשי. דוגמה למושג סיבתי הוא מושג הרעל. מושג הרעל הוא המושג של משהו אשר כשהוא מוחדר לאורגניזם הוא גורם לאורגניזם לחלות ו/או למות...

אבל הנקודה המהותית לגבי מושג הרעל היא שזהו המושג של **דבר כלשהו אשר גורם תוצאות מסוימות**. כך נשארת בידנו האפשרות ל**זיהוי מדעי** של רעלים, לגילוי שסוג מסוים של חומר, למשל, ציאניד, הוא רעל, ובהמשך לגילוי של אותו דבר בחומר אשר עושה אותו לרעל. ועתה, אם המושגים של סוגי המצבים הנפשיים השונים הם מושגים של מה שהוא, באופנים שונים, בעל הנטייה להיות הסיבה של תוצאות מסוימות ובעל הנטייה להיות התוצאה של סיבות מסוימות, אז יהיה זה לגמרי לא מפתיע אם יתברר שמצבים נפשיים הם מצבים פיזיקליים של המוח.

מושג המצב הנפשי הוא המושג של משהו שהוא, באופן טיפוסי, הסיבה של תוצאות מסוימות והתוצאה של סיבות מסוימות. אילו סוגים של תוצאות ואילו סוגים של סיבות? התוצאות הנגרמות על ידי המצב הנפשי הן דפוסי התנהגות מסוימים של האדם הנמצא במצב זה. למשל, הרצון באוכל הוא מצב של אדם או חיה שבאופן טיפוסי מביא אדם זה או חיה זו להתנהגות

<sup>3</sup> למען הדיוק נציין כי לפי אופן הבנה רווח של עמדתו של ארמסטרונג, קיים הבדל מה בינה לבין ההשקפה המעורבת כפי שאנו הצגנו אותה. הבדל זה אינו מהותי לענייננו. כפי שנראה מייד, גישתו של ארמסטרונג שמה דגש חזק על ניתוח משמעויות המונחים הנפשיים.

<sup>4</sup> D.M., Armstrong. "The Nature of Mind", in C.V. Borst (ed.) *The Mind-Brain Identity Theory*, London: MacMillan, 1970, pp. 67-79 (pp. 72-73).  
 אדיון נרחב יותר של ארמסטרונג בנושא הנדון נמצא בספרו: *A Materialist Theory of Mind*, London: Routledge and Kegan Paul, 1968.

של חיפוש מזון ושל צריכת מזון. הסיבות של למצבים נפשיים יהיו אובייקטים ואירועים בסביבת האדם. למשל, תחושה של ירוק היא התוצאה האופיינית באדם של פעולתו של דבר מה ירוק על עיניו.

הדפוס הכללי של הניתוח ברור ביותר ומתקבל ביותר על הדעת במקרה של **כוונות**. אם כוונתו של אדם היא ללכת למטבח כדי לקחת משהו לאכול, טבעי לגמרי שנחשוב על כוונה זו כסיבה בתוכו אשר מביאה, או נוטה להביא, לסוג מסוים זה של התנהגות. יתר על כן, מן הידוע הוא שאיננו יכולים לאפיין כוונות **אלא** במונחים של מה שהן נוטות לגרום. כיצד נוכל להבדיל בין הכוונה ללכת למטבח כדי לקחת משהו לאכול לבין כוונה אחרת – ללכת לחדר השינה כדי להשתרע על המיטה? רק באמצעות התוצאות השונות ששתי הכוונות נוטות להביא להן. עובדה זו הייתה עידוד לביהיוריזם. עדיין סביר יותר להסביר אותה באמירה שמושג הכוונה הוא מושג סיבתי. ההשערה הנוספת, ששתי הכוונות הן מטבען דפוסים פיזיקליים שונים, או מצבים פיזיקליים שונים, של מערכת העצבים המרכזית, היא, אם כן, תוספת טבעית (אם כי, כמובן, לא בלתי נמנעת מבחינה לוגית) לניתוח הסיבתי.<sup>5</sup>

ובכן, התורה הביהיוריסטית שגויה, לפי ארמסטרונג, משום שלפיה מושגים נפשיים הם מושגים המציינים דגמי התנהגות. רעיון זה שגוי, לדעת ארמסטרונג, משום שמושגים נפשיים הם מושגים של **מצבים פנימיים** של יצורים נפשיים. אולם אף על פי שארמסטרונג שותף לביקורת זו על העמדה הביהיוריסטית, הוא סובר שהביהיוריסטים עמדו על אמתה חשובה. כמותם, הוא סובר שיש קשר הדוק בין היסוד הנפשי לבין התנהגות, ולמעלה מכך, הוא סובר שקשר זה הוא קשר לוגי, קשר מושגי. לתפיסתו, מתבטא קשר לוגי זה בכך שמושגים נפשיים הם מושגים של מצבים פנימיים של יצורים נפשיים, **כאשר המושגים הללו מציינים את המצבים הללו באמצעות דגמי ההתנהגות (או דגמי הגירוי החושי והתגובה ההתנהגותית) שבהם כרוכים המצבים הללו**. המושג הנפשי של "ההאמנה שקר בחוץ" הוא מושג של מצב פנימי הנוטה להוליד להתנהגויות מקבוצת התנהגויות מסוימת (למשל, ללבישת בגדים חמים), ובדומה, המושג הנפשי של "ההאמנה שחם בחוץ" הוא מושג של מצב פנימי הנוטה להוליד להתנהגויות מקבוצת התנהגויות אחרת (למשל, ללבישת בגדים קלים). דהיינו, על פי העמדה של ארמסטרונג (בניגוד לעמדה הביהיוריסטית), הטענה שמהם מאמין שקר בחוץ מייחסת למשהו מצב פנימי כלשהו, אך איננה מתייחסת לטבעו החומרי של מצב זה. משמעותה של טענה זו, סובר ארמסטרונג, היא שמהם נמצא במצב פנימי אשר כרוך (במובן האמור של "כרוך" – המובן של תיווך סיבתי) בדגם התנהגות ספציפי. המושג הנפשי כלל אינו מתייחס לטבעו של המצב הפנימי האחראי לדגם ההתנהגות הרלוונטי – הוא מתייחס אך ורק לדגם ההתנהגות שבו כרוך מצב זה; שני מצבים נפשיים שייכים לאותו סוג נפשי, על פי קריטריון זה (למשל, שניהם כאבים), אם, ורק אם, שניהם כרוכים באותו דגם התנהגותי,

<sup>5</sup> M.D. Armstrong, "The Causal Theory of the Mind", in *Mind and Cognition*, G.W. Lycan (ed.), Oxford: Blackwell, pp. 20-27 (pp. 22-23).

דיון נרחב יותר של ארמסטרונג בנושא הנדון ניתן למצוא בספרו *A Materialist Theory of Mind*, London: Routledge and Kegan Paul, 1968.

באותן נטיות התנהגותיות, וכדומה.<sup>6</sup> קריטריון הסיווג של היסוד הנפשי, על פי "ההשקפה המעורבת" שמציג ארמסטרונג, מקביל לחלוטין לקריטריון הביהביוריסטי, אולם האונטולוגיה שונה. האונטולוגיה הנפשית, לפי השקפה זו, היא אונטולוגיה של מצבים ואירועים פנימיים. אונטולוגיה זו מתחייבת מעצם טיבם של המושגים הנפשיים – מעצם משמעותיותם של המונחים הנפשיים – **בהיותם מושגים של סיבות**, של מתווכים סיבתיים.

נוכל להבין טוב יותר את עמדתו של ארמסטרונג אם נזכר באחד מהיבטיה שאותו הצגנו ביחידה 4-5 בהקשר שונה. כזכור, ארמסטרונג התווה אנלוגיה בין מושגים נפשיים לבין מושגים מדעיים כדוגמת המושג הביולוגי של הגן. מושג זה היה ידוע במדע כמושג המציין רכיב גופני הממלא תפקיד סיבתי מסוים – רכיב האחראי להעברתם של מאפיינים תורשתיים (לכך שלצאצאי אדם יהיו תכונות מסוימות שיש לאדם עצמו). מהן תכונותיו הכימיות-פיזיקליות של הגן – זאת לא ידעו עד אמצע המאה ה-20, ותכונות אלו אינן חלק מתוכנו של מושג הגן (שהוא, כאמור, "ניטרלי מבחינה נושאית" – הוא מציין את מה שממלא בפועל תפקיד סיבתי זה). המושגים הנפשיים דומים מבחינה זאת למושג של גן. כאשר אנו מתארים מצבים נפשיים כמו כאבים במונחים נפשיים כדוגמת "כאב", "ההאמנה שירדת גשם", "פחד", וכדומה, אנו מתארים אותם באמצעות תפקידים הסיבתיים. המונח "כאב" הוא מונח שמשמעותו מצב פנימי אשר נגרם, או נוטה להיגרם, על ידי נזקים גופניים מסוימים, וגורם, או יש לו נטייה לגרום, התנהגויות מסוגים מסוימים (גניחה, אחיזה בחלק הגוף הפגוע, וכדומה).

### שאלה

האם, לפי השקפה זו, כפי שהיא מתבטאת בקטעים מדברי ארמסטרונג שהבאנו לעיל, עצם טיבם זה של המושגים הנפשיים מחייב אונטולוגיה **פיזיקליסטית** של היסוד הנפשי? האם, במילים אחרות, מתחייב מעצם טיבם של המושגים הנפשיים כי המצבים הנפשיים – הסיבות של התנהגויותינו – הם מצבים פיזיקליים?

ובכן, טיבם של המושגים הנפשיים בוודאי אינו מחייב מחויבות לוגית, או מושגית, שהמצבים עליהם הם חלים יהיו מצבים פיזיקליים (למעשה, אין כל השקפה מטריאליסטית הטוענת למחויבות **לוגית** כזו). אולם, כפי שאומר ארמסטרונג, ההשערה שמצבים אלו הם אכן מצבים פיזיקליים היא השלמה טבעית של הניתוח של מושגים נפשיים כמושגים המציינים סיבות של התנהגויות.

<sup>6</sup> כפי שראינו ביחידה 3, שיתוף בדגם התנהגותי אין משמעו אותה התנהגות בפועל – מדובר באותה **נטייה התנהגותית**.

חשוב לשים לב לכך שלהשקפות בדבר טבעו של היסוד הנפשי ניתן לייחס **אונטולוגיה נפשית** בשני מובנים שונים. כפי שאנו רואים, השקפות כאלו עשויות לדגול באונטולוגיה נפשית של **מצבים ואירועים**. אולם ביחסנו להשקפה בדבר טבעו של היסוד הנפשי אונטולוגיה של מצבים ואירועים עדיין לא מיצינו את עמדתה ביחס לאונטולוגיה הנפשית, שכן נותרת השאלה, אם המצבים והאירועים הנפשיים הם פיזיקליים או לא. המחלוקת האונטולוגית שבין הדואליזם הקרטזיאני לבין ההשקפה הפיזיקליסטית עניינה, כמובן, בשאלה אחרונה זו. אין מחלוקת בין שתי ההשקפות הללו באשר לשאלה האונטולוגית הראשונה: שתיהן מסכימות כי האונטולוגיה הנפשית היא אונטולוגיה של מצבים ואירועים. **ההשקפה הביהביוריסטית** חולקת על-כך.

### שאלה

התוכלו לשער בעטיו של איזה טעם סובר ארמסטרונג שההשערה הפיזיקליסטית מהווה השלמה טבעית של הניתוח הסיבתי של מצבים נפשיים?

הטעם פשוט: המדע מספק בסיס מוצק להנחה שמצבים פיזיקליים – באופן ספציפי, מצבי מוח – מהווים סיבות להתנהגויותינו, ומהווים מתווכים סיבתיים בין הגירויים התפיסתיים שלנו לבין תגובותינו ההתנהגותיות. למעשה, ארמסטרונג ופילוסופים נוספים (הבולט שבהם הוא האמריקאי דיוויד לואיס)<sup>7</sup> מציעים טיעון הדומה במבנהו לטיעון מן הסיבתיות הנפשית, שבו דנו בהרחבה ביחידה 4-5, בזכות הטענה שמצבים נפשיים הם מצבים פיזיקליים:

(1) המושגים הנפשיים הם מושגים של מצבים הממלאים תפקידים סיבתיים מסוימים.

(2) כפי שמגלה המדע, תפקידים אלו ממלאים על ידי מצבי מוח מסוימים. לפיכך,

המצבים הנפשיים (המצבים שמצוינים על ידי המושגים הנפשיים) הם מצבי מוח.

### שאלה

למעשה, בטיעון זה מובלעת הנחה נוספת (הנחה, שאם לא נפרש את הטיעון ככולל אותה, הוא לא יהיה תקף). מהי הנחה זו? (יקל עליכם להשיב לשאלה זו אם תעינו בטיעון מן הסיבתיות הנפשית ביחידה 4-5).

<sup>7</sup> ראו, בין היתר, את מאמרו של לואיס

"An Argument for the Identity Theory", *Journal of Philosophy* 63 (1966), pp. 17-25.



ההנחה המובלעת דומה להנחה השלישית של הטיעון מן הסיבתיות הנפשית, שעל פיה:

(3) כל האפשרויות שעל פיהן התפקידים הסיבתיים האמורים ממולאים הן על ידי מצבי מוח והן על ידי מצבים אחרים (למשל, מצבים "רוחניים" קרטזיאניים) אינן מתקבלות על-הדעת.

הנחה זו באה לשלול אפשרויות דוגמת אלו של היקבעות-יתר סיבתית, או של שרשרות פסיכופיזיות מעורבות. כמובן, מי שאינם משוכנעים כי אמנם יש הצדקה לפסילתה של אפשרות אחת לפחות מן האפשרויות הללו, לא ישוכנעו על ידי הטיעון הנדון, בדיוק כשם שהם לא ישוכנעו על ידי הטיעון מן הסיבתיות הנפשית.

כיוון שדנו בהרחבה בטיעון מן הסיבתיות הנפשית, לא נעריך כאן את הטיעון הנוכחי. נסתפק בהצבעה על העובדה שהנחתו הראשונה של טיעון זה היא במובן חשוב חזקה יותר מהנחתו הראשונה של הטיעון מן הסיבתיות הנפשית, ולפיכך היא חשופה יותר לביקורת.<sup>8</sup> הנחה זו אינה מתחייבת רק לכך שהמצבים שעליהם חלים המושגים הנפשיים הם סיבות של התנהגויותינו, היא כרוכה גם במחויבות באשר לאופיים של המושגים הללו עצמם: על פיה, מושגים אלו הם, אפשר לומר, מושגים סיבתיים – הם מציינים את המצבים הנפשיים כמצבים הממלאים תפקידים סיבתיים אלו ואלו. מצב נפשי כדוגמת כאב – המצב המצוין על ידי המונח "כאב" – מוגדר כמצב הנפשי שהינו, בהתאם להנחה זו, על ידי התפקיד הסיבתי הספציפי שהוא ממלא. כדי לבסס את המסקנה הפיזיקליסטית אין הכרח בהנחה כה חזקה, וניתן להסתפק בהנחה שעל פיה המצבים הנפשיים הם בפועל סיבות של התנהגויותינו (כפי שאמנם מניח הטיעון מן הסיבתיות הנפשית). אולם ארמסטרונג ופילוסופים אחרים השותפים לדעתו סוברים שהביהיוריסטים צדקו בכך שטענו לקיומו של קשר מושגי-לוגי בין מושגים נפשיים לבין התנהגויות או דגמי התנהגויות, וטענתם הייתה נעוצה רק בטענה כי המושגים הנפשיים מציינים את דגמי ההתנהגויות עצמם, ואילו לאמתו של דבר (לפי ארמסטרונג ושותפיו לדעה) הם מציינים את המצבים הפנימיים האחראיים סיבתית לדגמי ההתנהגויות. סברה זו שלהם מתבטאת בהנחתו הראשונה של הטיעון הנוכחי.

כאשר קוראים את מסקנת הטיעון יחד עם הנחה זו, רואים שהטיעון אינו סתם טיעון בזכות זיהויים של מצבים נפשיים עם מצבי מוח, אלא שהוא טיעון בזכות פיזיקליזם **התנהגותני**: מצבים נפשיים הם מצבי מוח, והם מאופיינים כמצבים הנפשיים שהינם (האמנה שיורד גשם, כאב, פחד, וכדומה) באמצעות תפקידיהם הסיבתיים; תפקידים סיבתיים אלו של המצבים הנפשיים הם תפקידיהם "ההתנהגותניים", דהיינו תפקידיהם ביצירת התנהגות או בתיווך סיבתי בין גירויים חושיים לתגובות התנהגותיות.

<sup>8</sup> נחזור ונבהיר כי באמרנו שטענה מסוימת היא **חזקה** יותר מטענה אחרת איננו מתכוונים לכך שהיא סבירה יותר. אנו מתכוונים לכך שתוכנה של הטענה הראשונה רב מתוכנה של הטענה השנייה, ובמילים אחרות, שהראשונה מתחייבת לעובדות נוספות, מעבר למחויבותה של הראשונה. טענה המתחייבת ליותר מרעותה מן הסתם חשופה לביקורת במידה רבה יותר מרעותה.

ההנחה הראשונה של הטיעון מהווה, כפי שראינו, את אחד משני הרכיבים של ההשקפה המעורבת הפיזיקליסטית-התנהגותנית, "הרכיב המושגי". וראינו כי רכיב זה של ההשקפה המעורבת אינו מחייב את הרכיב השני שלה – אין הוא מחייב אונטולוגיה פיזיקליסטית; הטיעון הנזכר בזכות הטענה שמצבים נפשיים הם פיזיקליים נזקק להנחות נוספות: להנחה המדעית-אמפירית שעל פיה מצבים פיזיקליים מסוימים הם אלו שממלאים את התפקידים הסיבתיים אשר מאפיינים את המצבים הנפשיים, ולהנחה שעל פיה כל האפשרויות, שעל פיהן התפקידים הסיבתיים האמורים ממולאים הן על ידי מצבי מוח והן על ידי מצבים אחרים, אינן מתקבלות על-הדעת. אם הרכיב המושגי מותיר את השאלה האונטולוגית כשאלה פתוחה (שאלה שההכרעה בה תלויה בין היתר בהכרעה אמפירית), ואין הוא מחייב אונטולוגיה פיזיקליסטית, מן הסתם אין הוא מחייב אונטולוגיה פיזיקליסטית **מסוימת**. כלומר, אם אין הכרח שמצבים נפשיים יהיו פיזיקליים, ודאי שאין הכרח שהם יהיו בעלי מבנה פיזיקלי X או Y; למשל, אין הכרח שהחומר שהם מצביו יהיה חומר פחמני. הרכיב המושגי של ההשקפה המעורבת מהווה את קריטריון הסיווג או האינדיווידואציה של היסוד הנפשי, וקריטריון סיווג זה הינו קריטריון המאפשר **ריבוי מימושים**. הוא איננו קריטריון סיווג פיזיקליסטי. לו סברו הדוגלים בהשקפה המעורבת שאמנם בפועל ממולאים התפקידים הסיבתיים המאפיינים מצב נפשי זה או אחר אצל יצורים שונים על ידי מצבים בעלי מבנים פיזיקליים שונים, היה עליהם לקבל את **מציאותה של תופעת ריבוי המימושים** של המצבים הנפשיים (ולא רק את אפשרותה), שכן המצבים הנפשיים הם על פיהם המצבים הממלאים תפקידים סיבתיים.

#### הבהרה

**תופעת ריבוי המימושים** של המצבים הנפשיים היא העובדה (הנטענת על ידי פילוסופים שונים) כי מצבים נפשיים מאותו סוג ממומשים **בפועל** על ידי מצבים חומריים מסוגים שונים. קריטריון סיווג נפשי **המאפשר** ריבוי מימושים אינו מחויב לכך שאמנם בפועל ממומשים מצבים נפשיים מאותו סוג על ידי מצבים חומריים מסוגים שונים (כלומר, אין הוא מחויב לתופעת ריבוי המימושים).

לא כל הפילוסופים המקבלים את הניתוח הנדון של מושגים נפשיים דנים באפשרות של ריבוי המימושים (עבודותיהם הרלוונטיות של רובם קדמו למאמרו של פטנאם אשר הצביע על תופעת ריבוי המימושים). חלקם אף סוברים שאף על פי שהמושגים הנפשיים בפני עצמם הם מושגים המאפשרים ריבוי מימושים, את התכונות הנפשיות יש לזהות לא עם תכונותיהם הסיבתייות-התנהגותיות של המצבים הנפשיים, דהיינו עם היותם כרוכים בדגמי גירוי-תגובה מסוימים (תכונות הניתנות לריבוי מימושים פיזיקלי כך שאינן תכונות פיזיקליות), אלא עם **התכונות הפיזיקליות** של המצבים הללו אשר אחראיות לכך שהמצבים כרוכים בדגמים אלו. לפי גישה זו, התכונות הנפשיות הינן תכונות פיזיקליות. אם אמנם בפועל ממולא התפקיד

הסיבתי המאפיין כאבים, למשל, על-ידי מצבים פיזיקליים החולקים באופן ייחודי תכונה פיזיקלית מסוימת, הרי שתכונת הכאב היא תכונה פיזיקלית זו, ואין היא התכונה ההתנהגותית המאפיינת כאבים (היותם כרוכים בדגם מסוים של גירוי ותגובה) – תכונה הניתנת לריבוי מימושים.

עמדה זו מהווה, למעשה, מעין פיזיקליזם-סוג. לא נוכל לדון כאן בשאלת המניע שביסוד עמדה זו.<sup>9</sup> לענייננו חשוב להדגיש, כי פילוסופים המקבלים את הניתוח הסיבתי של המושגים הנפשיים שאותו הצגנו, וסוברים כי אמנם התפקידים הסיבתיים המאפיינים מצבים נפשיים מאותו סוג ממולאים על ידי מצבים בעלי מבנים פיזיקליים שונים – ובמילים אחרות – שאחראיות להן תכונות פיזיקליות **שונות**, בוודאי יקבלו שקריטריון הסיווג של המצבים הנפשיים והתכונות הנפשיות איננו קריטריון פיזיקליסטי (דהיינו, הם ידחו פיזיקליזם-סוג) אלא קריטריון התנהגותי. עמדה כזאת משלבת קריטריון סיווג לא פיזיקליסטי של היסוד הנפשי – קריטריון סיווג המאפשר ריבוי מימושים – עם אונטולוגיה פיזיקליסטית. בין שנכון ליחס עמדה זו לפילוסופים אלו או אחרים, ובין שזוהי עמדה אפשרית בלבד, עמדה זו חשובה לנו בדרכנו להסביר את העמדה הפונקציונליסטית. למעשה, כפי שנבין בהמשך, עמדה זו מהווה מעין גרסה פשוטה של פונקציונליזם.

הזכרנו כי הרכיב המושגי של "העמדה המעורבת" – הטענה כי המושגים הנפשיים הם מושגים של מצבים הממלאים תפקידים סיבתיים מסוימים – היא טענה חזקה, ולפיכך היא עשויה להיתקל בביקורת. ואמנם, אחד מסוגי הביקורות שכוונו כנגד העמדה הביהביוריסטית (ביקורות שאותן הצגנו ביחידה 3) הוא סוג של ביקורות שאם הן תקפות הן חותרות גם תחת "ההשקפה המעורבת". באיזה סוג של ביקורות מדובר? ברור שאין מדובר בביקורת שעל פיה העמדה הביהביוריסטית אינה יכולה לכבד את היות המונחים הנפשיים מונחים המציינים ישויות בעלות כוח סיבתי ואת היותם של הסברים פסיכולוגיים של התנהגות הסברים סיבתיים. ביקורת זו אינה ישימה כנגד ההשקפה המעורבת, שהרי השקפה זו למדה, ניתן לומר, את לקחה של העמדה הביהביוריסטית – את הקושי הכרוך בה שעליו מתיימרת ביקורת זו להצביע: העמדה המעורבת מחויבת להיות האונטולוגיה הנפשית אונטולוגיה של מצבים ואירועים, שאין בעיה ליחס להם כוחות סיבתיים; עמדה זו אינה מזהה את היסוד הנפשי עם התנהגויות או דגמי התנהגויות, אלא עם המצבים האחראיים להתנהגויות.

## שאלה

איזו ביקורת כנגד העמדה הביהביוריסטית תחתור, אם היא תקפה, גם תחת ההשקפה המעורבת הפיזיקליסטית-התנהגותית?

<sup>9</sup> נקודה זו תובן טוב יותר בהמשך, כשנחدد את ההבחנה בין פונקציונליזם עממי לבין פונקציונליזם אמפירי. למעשה, מקובל לפרש את עמדת ארמסטרונג עצמו ככזו הדוגלת בפיזיקליזם-סוג מן הדגם המוזכר כאן, דהיינו עמדה שעל פיה למרות שהמושגים הנפשיים הם מושגים סיבתיים-התנהגותיים, התכונות הנפשיות זהות לתכונות הפיזיקליות שאחראיות בפועל למילוי התפקידים הסיבתיים הרלוונטיים על ידי המצבים הנפשיים. וכך הצגנו גם אנחנו את עמדתו של ארמסטרונג ביחידה 4-5.

מדובר בביקורת שעל פיה אי אפשר להעמיד ייחוסים נפשיים על ייחוסים של התנהגויות או דגמי התנהגויות מבלי לכלול גם התייחסות למאפיינים נפשיים אחרים. אם ביקורת זו תקפה, היא בהחלט חותרת גם תחת ההשקפה המעורבת. הרעיון שביסוד ביקורת זו הוא שאיננו יכולים לומר כי מצב נפשי, כמו רצייה לבקר בטיבט, הוא המצב הפנימי אשר משמש כמתווך סיבתי בין, למשל, קבלת כרטיס טיסה לטיבט לבין טיסה לטיבט. כפי שראינו, לפי ביקורת זו אין מתאם כזה בין רציות לבין התנהגויות: מי שרוצה לבקר בטיבט יטוס לטיבט אם יינתן לו כרטיס רק אם יתקיימו תנאים נפשיים נוספים לגביו (למשל, אם הוא **מאמין** שהטיסה בטוחה, ועוד). אין מתאם בין רציית אדם לטוס לטיבט לבין טיסתו לטיבט כאשר ניתן לו כרטיס. ואם כך, אם אין מתאם כזה, הרי ברור שלא קיים כל מצב פנימי האחראי לקיומו של מתאם כזה: אצל מי שרוצה לבקר בטיבט לא קיים כל מצב פנימי הממלא תפקיד של תיווך סיבתי בין קבלת הכרטיס לבין הטיסה, ולא ניתן לאפיין רצייה זו בתור המצב הפנימי הממלא תפקיד סיבתי כזה. דהיינו, בניגוד להשקפה המעורבת, לא ניתן לאפיין מצבים נפשיים כמצבים הממלאים תפקיד של תיווך סיבתי בין גירויים תפיסתיים לבין תגובות התנהגויות.

### שאלה

האם מתחייב מביקורת זו כי מצבים נפשיים כלל אינם ממלאים תפקידים סיבתיים באשר להתנהגות?

חשוב לשים לב לכך שמביקורת זו לא מתחייבת מסקנה כה חזקה. ביקורת זו מתיישבת עם האפשרות שמצבים נפשיים ממלאים תפקיד סיבתי **סבך יותר** באשר להתנהגות, תפקיד שאינו מתמצה בתיווך בין גירויים תפיסתיים, מצד אחד, לבין תגובות התנהגויות, מצד אחר. ראינו כי ביקורת זו מבוססת על הטענה שאפיון של מושג נפשי במונחי דגמים התנהגוטיים הוא לקוי בכך שהוא מתעלם ממעורבותם של תנאים נפשיים נוספים בדגמים ההתנהגוטיים (בדוגמה שלעיל, למשל, מההאמנה שהטיסה בטוחה). ובכן, ביקורת זו אינה שוללת את האפשרות שמצבים נפשיים משמשים כמתווכים סיבתיים בין גירויים תפיסתיים **ומצבים נפשיים אחרים** לבין התנהגויות. על פיה, התפקיד הסיבתי שממלא מצב נפשי אינו פשוט – אין הוא מתמצה ביחסי גירויים ותגובות: אין זה נכון שיצור נפשי הנמצא במצב נפשי מסוים (למשל, יצור נפשי הרוצה לבקר בטיבט) יגיב תמיד באותם אופנים על אותם גירויים. תגובתו תלויה גם במצביו הנפשיים האחרים. בבואנו לאפיין את התפקיד הסיבתי של מצב נפשי מסוג מסוים, אנחנו חייבים אפוא להתייחס גם לקשריו עם מצבים נפשיים אחרים. אך ביקורת זו אינה מחייבת שלמצבים נפשיים אין תפקידים סיבתיים באשר להתנהגות, ולכן לא נובע ממנה שאי אפשר לאפיין מצבים נפשיים באמצעות תפקידם הסיבתי. **נותנת האפשרות שאפשר לאפיין מצבים נפשיים כמצבים פנימיים הממלאים תפקידים סיבתיים סבוכים**, סבוכים יותר מאלו שמתארת ההשקפה של פיזיקליזם-התנהגותני. רעיון זה בדבר אפיון סיבתי מורכב של המצבים הנפשיים עומד בבסיסה של ההשקפה הפונקציונליסטית.

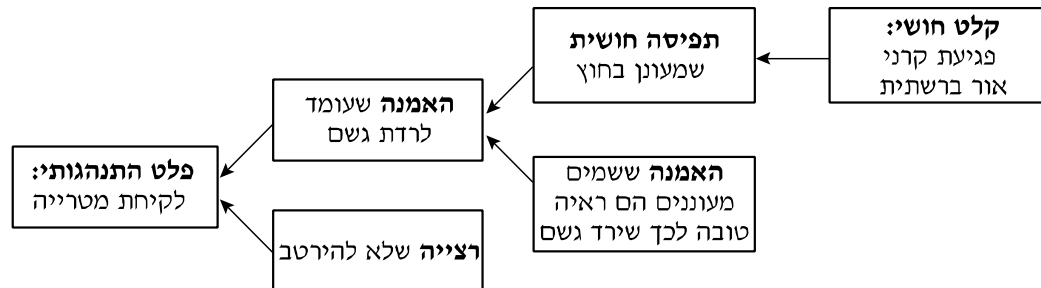
## 2. יסודות ההשקפה הפונקציונליסטית

כמו ההשקפה המעורבת הפיזיקליסטית-התנהגותנית, מקבלת ההשקפה הפונקציונליסטית את הטענה שהאונטולוגיה הנפשית היא אונטולוגיה של מצבים ואירועים. מרבית הפילוסופים הפונקציונליסטים מקבלים גם את הטענה שמצבים נפשיים הם מצבים פיזיקליים, מצבי מוח, אולם טענה זו אינה נחשבת כחלק מן ההשקפה הפונקציונליסטית עצמה, שכפי שנראה בהמשך, במובן חשוב היא **ניטרלית** באשר לטיבם האונטולוגי של המצבים והאירועים הנפשיים. מאחר שהאונטולוגיה הנפשית על פי הפונקציונליזם היא אונטולוגיה של מצבים ואירועים, השקפה זו יכולה לכבד – ואכן מכבדת – את הרעיון שמצבים ואירועים נפשיים הם מתווכים סיבתיים. ההשקפה הפונקציונליסטית מאפיינת מצבים נפשיים באמצעות "התפקיד" (או "הפונקציה") שהם ממלאים בחיים הפסיכולוגיים של הפרט. בכך דומה ההשקפה הפונקציונליסטית להשקפה הפיזיקליסטית-התנהגותנית, אולם היא מתקנת את ההשקפה הפיזיקליסטית-התנהגותנית בכך שהיא מאפיינת באופן שונה – באופן שנראה הולם יותר את מציאות חיינו הנפשיים – את תפקידם הסיבתי של המצבים הנפשיים. מיד נרחיב בעניינו של התפקיד הסיבתי הזה, ונסה להבהיר באופן מפורט את האפיון של המצבים הנפשיים על פי הפונקציונליזם. אך לפני-כן נדגיש שיחד עם חשיבותה של הצגת ההשקפה הפונקציונליסטית כ"גרסה משופרת" של ההשקפה הפיזיקליסטית-התנהגותנית, ההשקפה הפונקציונליסטית הכניסה גורמים חשובים נוספים לתמונה בדבר טבעו של היסוד הנפשי ובדבר יחסי הגוף והנפש – גורמים שעליהם נעמוד, כמובן, בהמשך.

ביסוד ההשקפה הפונקציונליסטית עומדת הטענה שמה שמאפיין יצורים נפשיים – מה שמקנה להם נפשיות – אינו החומר שממנו הם עשויים, אלא **הארגון הפונקציונלי** שלהם – הקשרים בין מצביהם השונים, מאלו מצבים עובר היצור לאלו מצבים. להיות יצור נפשי משמעו להיות בעל ארגון פונקציונלי מסוים; אין משמעו להיות בעל הרכב פיזיקלי מסוים. התרמוסטט השולט על פעולת מזגן האוויר אינו יצור נפשי לא משום שהרכבו הפיזיקלי אינו ההרכב הנדרש, אלא משום שארגונו הפונקציונלי (הפשוט באופן יחסי) – הקשרים הסיבתיים בין מצביו השונים – אינו הארגון הפונקציונלי הנדרש. ואנו הננו יצורים נפשיים לא בשל החומר שממנו עשוי מוחנו, אלא בשל הארגון הפונקציונלי שמוחנו מאפשר לנו. כך מהווה הפונקציונליזם תורה בדבר מהותו של היסוד הנפשי, מהותה של נפשיות.

אולם הפונקציונליזם מהווה גם תורה בדבר סוגים נפשיים, בדבר אינדיווידואציה נפשית: מה שמאפיין מצבים נפשיים ומשייכם לסוגיהם אינו החומר שממנו הם עשויים, אלא **קשריהם למצבים אחרים מסוימים**: קשריהם לגירויים חושיים, לתגובות התנהגותיות, ולמצבים נפשיים אחרים. מצב נפשי כמו האמנה שיורד גשם עשוי להיות מאופיין כמצב הנגרם, למשל, על ידי תפיסה חושית מסוימת (ראיית טיפות המים היורדות מן השמים), ואשר כשהוא חובר לרצייה לצאת לרחוב, להאמנה שהציאה לרחוב כשיורד גשם תיכרך בהירטבות אלא אם כן

ינקט אמצעי כלשהו, לרצייה שלא להירטב, ולהאמנה ששימוש במטרייה הוא אמצעי יעיל למניעת ההירטבות, מוביל על פי-רוב (נוטה להוביל) לשימוש במטרייה. התרשים הזה מדגים לנו רשת קשרים של מצבים נפשיים, גירויים חיצוניים ותגובות התנהגותיות:



#### תרשים 4. רשת קשרים של מצבים נפשיים, גירויים חיצוניים ותגובות התנהגותיות

הפונקציונליסטים מאפיינים מצבים נפשיים כמצבים הממלאים תפקידים סיבתיים סבוכים כאלו. מה שמשייך מצב נפשי לסוגו הוא התפקיד הסיבתי שהוא ממלא, כאשר תפקיד זה מאופיין לא רק על ידי מתאם פשוט של גירוי-תגובה, אלא על ידי מערכת מורכבת של קשרים. מצבים נפשיים נוטים, לפחות על פי רוב, לגרום התנהגויות למיניהן רק בתיווכם של מצבים נפשיים נוספים; והם נוטים להיגרם על ידי גירויים חיצוניים, על פי רוב, רק בתיווכם של מצבים נפשיים אחרים. הפונקציונליזם מביא עובדות אלו בחשבון בקבעו מה מאפיין סוגים של מצבים נפשיים. ה"תפקיד" (או הפונקציה) המאפיין מצבים כמו רעב, פחד, האמנה שפורץ מסתתר מאחורי הדלת, וכדומה, הוא תפקיד תיווכי מורכב – ניתן לאפיין סוגי מצבים כאלו רק על ידי קשריהם עם מכלול רחב של מצבים, ביניהם מצבים נפשיים אחרים.

מאחר שהאפיון הפונקציונליסטי של מצבים נפשיים הוא אפיון של מצבים במונחי קשריהם למצבים אחרים, נכון לתארו **כאפיון יחסותי** (או "חיצוני"). הוא מאפיין מצבים נפשיים באמצעות תכונות יחסותיות שלהם, ולא באמצעות תכונות "פנימיות" (או "אינטרינסיות") שלהם (דהינו תכונות שיש להם בפני עצמם, בלי תלות בקיומם של מצבים או דברים אחרים כלשהם).

אפיונים פונקציונליים – אפיונים של אובייקטים במונחי התפקיד שלהם – רווחים בהקשרים שונים. כשאנו מאפיינים אובייקט **ככיסא**, אנו מאפיינים אותו באמצעות הפונקציה שלו – לשמש לישיבה – ולא באמצעות החומר שממנו הוא עשוי. וכך גם כשאנו מאפיינים אובייקטים **כמכונית, מנוע, לב מלאכותי**, ועוד ועוד.

שמה של ההשקפה – "פונקציונליזם" – מבטא את העובדה שלפי השקפה זו מצבים נפשיים מאופיינים על פי הפונקציה, דהיינו התפקיד, שהם מאפיינים בחייו הנפשיים של הפרט (פונקציה הבאה לידי ביטוי, כאמור, בקשרים שהמצב מקיים עם גירויים חושיים, עם תגובות התנהגותיות ועם מצבים נפשיים אחרים). למעשה, גם ההשקפה המעורבת הפיזיקליסטית-התנהגותית מאפיינת מצבים נפשיים על פי הפונקציה שהם ממלאים – ההבדל טמון בכך שהפונקציה הרלוונטית, על פי הפיזיקליזם ההתנהגותני, היא פשוטה יותר, היא פונקצית גירוי-תגובה, או, במילים אחרות, פונקציה של פלט-קלט. במובן זה, ניתן לומר שהפיזיקליזם ההתנהגותני הוא סוג של פונקציונליזם. אולם השם "פונקציונליזם" בפילוסופיה של הנפש יוחד כשמה של ההשקפה המאפיינת מצבים נפשיים במונחי הפונקציה הסבוכה יותר, הכוללת גם את קשריהם של מצבים נפשיים מסוימים למצבים נפשיים אחרים.

על פי ההשקפה הפונקציונליסטית, להיות יצור נפשי משמעו להיות בעל ארגון פונקציונלי מסוים; ומה שמשייך מצב נפשי לסוגו אלו הם היחסים הפונקציונליים המתקיימים בינו לבין גירויים תפיסתיים, תגובות התנהגותיות ומצבים אחרים של היצור הנפשי.

תיארנו לעיל רשת של קשרים שבה עשויה להיכרך ההאמנה שיורד גשם. אין הכוונה שרשת קשרים זו מאפיינת בפועל כל פרט ספציפי של האמנה זו (כל מקרה ומקרה שבו מישחו מאמין שיורד גשם). ודאי שאין זה נכון כי בכל מקרה ומקרה שבו אני מאמין שיורד גשם מוליכה אותי האמנתי זו (בצירוף למצבים נפשיים אחרים) להשתמש במטרייה; שהרי אני עשוי להאמין שיורד גשם מבלי לרצות שלא להירטב, למשל. מצבים נפשיים רבים, כדוגמת ההאמנה שיורד גשם, אינם מאופיינים על פי הפונקציונליזם רק על ידי הקשרים שהם מקיימים בפועל עם מצבים נפשיים אחרים, עם תפיסות חושיות ועם תגובות התנהגותיות. אלא, רבים מהקשרים המאפיינים אותם (אם לא כולם) הם **קשרים מותניים**: האמנתו של אדם שיורד גשם היא, לפי הפונקציונליזם, המצב אשר יובילו על פי רוב לשימוש במטרייה, **אם**, למשל, אדם זה מאמין גם שהציאה לרחוב כשיורד גשם תיכרך בהירטבות אלא אם כן יינקט אמצעי כלשהו, אם הוא רוצה שלא להירטב, ואם הוא מאמין ששימוש במטרייה הוא אמצעי יעיל למניעת הירטבות.

וכשם שרשת קשרים זו איננה מאפיינת כל מקרה ומקרה של ההאמנה שיורד גשם, כך **היא גם אינה ממצה** את אפיונה הפונקציונלי של האמנה זו: במצבים אחרים – כאשר האמנה זו חוברת למצבים נפשיים אחרים – היא תוביל את האדם להתנהגויות אחרות (או למצבים נפשיים אחרים). במילים אחרות, האמנה זו, כמו המצבים הנפשיים האחרים, מאופיינת על פי הפונקציונליזם על ידי **רשתות רבות של קשרים**. חלק נוסף מאפיונה הפונקציונלי של ההאמנה שיורד גשם הוא שהיא תוליך את האדם לשמות, אם הוא חושב שגשם יעלה את מפלס הכנרת ואם הוא רוצה מאוד שמפלס הכנרת יעלה; חלק נוסף מאפיונה הוא שהיא

תוליך את האדם הזה להצטער, אם הוא מאמין שירידת גשם תגרום את ביטול של משחק הכדורגל, שלו הוא ציפה בכליון עיניים, וכדומה. אפיונה המלא של האמנה זו, על פי הפונקציונליזם, ניתן על ידי **כל** רשתות הקשרים מן הסוגים הללו, כל רשתות הקשרים המתקיימים בינה לבין מצבים נפשיים אחרים, תפיסות חושיות ותגובות התנהגותיות. מובן שהדברים נכונים לגבי כל המצבים הנפשיים: כל מצב נפשי מאופיין, על פי הפונקציונליזם, על ידי **כל** רשתות הקשרים המתקיימות בינו לבין מצבים נפשיים אחרים, תפיסות חושיות ותגובות התנהגותיות (מובן מאליו, שאיננו יכולים להתוות רשימה ממצה של כל רשתות הקשרים המאפיינות מצבים נפשיים).

למעשה, הפונקציה המאפיינת את המצבים הנפשיים היא פונקציה סבוכה עד מאוד. אפיון פונקציונליסטי של מצב נפשי אחד כרוך בהתייחסות **לכל** מצביו הנפשיים האחרים של היצור הנפשי. זאת משום שקיימים קשרים אפשריים (דהיינו, בנסיבות מסוימות **יתקיימו** קשרים) בין כל מצביו הנפשיים לבין עצמם. ההאמנה שיוורד גשם קשורה, כפי שראינו, ברצייה שלא להירטב, אולם בתנאים מסוימים היא תהיה קשורה בהאמנות על כדאיות החקלאות או על גובה מפלס המים בכנרת, או בתקווה שמשחק הכדורגל לא יבוטל, וכדומה; וכל אחד מן המצבים הנפשיים האחרים הללו, בתורו, עשוי להיות קשור עם מצבים נפשיים אחרים רבים (וזהותו, לפי הפונקציונליזם, נקבעת על ידי קשריו עמם), וכך הלאה וכך הלאה. אפיון פונקציונליסטי מלא של מצב נפשי כלשהו כרוך אפוא בתיאור רשת הקשרים המסועפת שבה כרוכים כל מצביו הנפשיים. לפי ההשקפה הפונקציונליסטית, אי אפשר לאפיין מצב נפשי אחד מבלי להתייחס **לכל** המצבים הנפשיים. בשל מחויבותה של ההשקפה הפונקציונליסטית לרעיון זה, **השקפה זו היא השקפה "הוליסטית"** (holism – בעברית: "כוליות", מלשון "כול"). בהמשך נראה כי לעובדה שההשקפה הפונקציונליסטית מחויבת להוליסם יש השלכות חשובות, ולאחת מהן לפחות יש בעיני רבים השלכה בעייתית במיוחד.

חשוב להדגיש שעל פי הפונקציונליזם אין זה נכון לומר שמצבים נפשיים הם **פונקציות**. על פי השקפה זו מצבים נפשיים הם **מצבים פנימיים** של היצור הנפשי. אולם מה שעושה את המצבים הללו למצבים נפשיים, ומה שעושה אותם למצבים הנפשיים שהינם (דהיינו, לכאב חד, לרצייה בעוגת תפוחים חמה, להאמנה שהמשבר הכלכלי קרב לסופו, וכדומה) זוהי הפונקציה שלהם. מצבים נפשיים הם אותם מצבים פנימיים של היצור הנפשי הממלאים – או, במינוח המקובל בקרב פילוסופי הנפש, **המממשים** – פונקציות מסוימות.

אפשר לומר שדווקא לפי **הביהיוריזם** אפיונים נפשיים הם פונקציות (הם פונקציות גירוי-תגובה פשוטות). המושגים הנפשיים אינם, לפי ההשקפה הביהיוריסטית, מושגים של מצבים פנימיים **המממשים** פונקציות משום שאינם כלל מושגים של מצבים פנימיים. מטעם זה נתקלה השקפה זו בקושי בדבר התפקיד הסיבתי של היסוד הנפשי, שעליו דברנו ביחידה 3.



הזכרנו מספר השקפות אשר דוגלות בקריטריון אינדיווידואציה של מצבים נפשיים שאינו קריטריון אינדיווידואליסטי – השקפות המכבדות את האפשרות של ריבוי מימושים של המצבים הנפשיים: ביהיוריזם, פיזיקליזם התנהגותני, ופונקציונליזם. למעשה, ראינו שהשקפת הפיזיקליזם ההתנהגותני היא השקפת ביניים בין ההשקפה הביהביוריסטית לבין ההשקפה הפונקציונליסטית: היא "מתקנת" פן אחד של ההשקפה הביהביוריסטית – את הפן האונטולוגי, ואילו ההשקפה הפונקציונליסטית "מתקנת" הן פן זה והן פן נוסף של ההשקפה הביהביוריסטית – את קריטריון האינדיווידואציה שלה.

---

♦ אפשר לחשוב על השקפת ביניים אפשרית נוספת בין הביהיוריזם לבין הפונקציונליזם. התוכלו לתאר בקצרה את הרעיון המרכזי של השקפה אפשרית כזו?

---

אף על פי שההשקפה הפונקציונליסטית "מתקנת" את ההשקפה הביהביוריסטית בשני היבטים חשובים, רבים מאפיינים אותה כיורשת של ההשקפה הביהביוריסטית, כפי הנראה משום שרעיון הפונקציה כעומדת ביסוד הסיווג הנפשי מהותי לשתי ההשקפות הללו. אולם אל לנו לשכוח שההשקפה הפונקציונליסטית ניזונה גם מן ההשקפה הפיזיקליסטית: לפונקציונליזם, בדיוק כמו לפיזיקליזם, מהותי גם הרעיון שמצבים נפשיים הם מצבים פנימיים של היצור הנפשי, מצבים שיש להם כוחות סיבתיים; וכפי שרמזנו כבר, ונראה בהמשך, מרבית הפונקציונליסטים סוברים שמצבים פנימיים אלו הם מצבים פיזיקליים.

אמרנו כי על פי ההשקפה הפונקציונליסטית, מה שמשייך מצב נפשי לסוגו (דהיינו, מה שעושה אותו לכאב, לרעב, להאמנה שיורד גשם, וכדומה) אלו הם היחסים הפונקציונליים המתקיימים בינו לבין גירויים תפיסתיים, תגובות התנהגותיות ומצבים אחרים של היצור הנפשי. במובן זה, ההשקפה הפונקציונליסטית מציעה **קריטריון סיווג** (קריטריון אינדיווידואציה) למצבים נפשיים. כפי שכבר ראינו בדיוננו על פיזיקליזם, קריטריון סיווג של מצבים נפשיים הוא למעשה קריטריון לזהותן של **תכונות נפשיות**, שהרי מצבים נפשיים זהו סוג הם מצבים נפשיים בעלי אותן תכונות נפשיות. אם כן, אנו יכולים לומר כי לפי הפונקציונליזם **תכונות נפשיות הן תכונות פונקציונליות**; ובדברנו על תכונות **פונקציונליות** בהקשר זה כוונתנו להיותם של מצבים נפשיים קשורים בקשרים סיבתיים אלו עם אלו, ועם תפיסות חושיות ותגובות התנהגותיות.

הרעיון שעל פיו מצבים נפשיים הם מצבים פנימיים המממשים פונקציות מסוימות הוא רעיון מרכזי מאוד בחשיבה הפונקציונליסטית. אחד מהיבטיו של רעיון זה הוא הרעיון בדבר **אי-המהותיות של החומר הספציפי של המצבים הנפשיים להיותם נפשיים ולסוג הנפשי שלהם**. אם מה שקובע את היותו של מצב פנימי מסוים מצב נפשי ואת סוג המצב הנפשי שהינו, זו הפונקציה שהוא מממש, הרי, במובן חשוב, החומר של המצב אינו מהותי לנפשיותו ולסוגו הנפשי. היותו של המצב מצב חומרי מסוים – מצב בעל מאפיינים פיזיקליים אלו ולא אחרים – זו עובדה שאינה מהותית להיותו של המצב מצב נפשי, ולזהותו הנפשית הספציפית.

לפי ההשקפה הפונקציונליסטית, שני מצבים שייכים לאותו סוג נפשי, אם, ורק אם, הם מממשים אותה פונקציה בחייו הנפשיים של הפרט (כלומר, אם הם קשורים באותם קשרים מותנים עם גירויים תפיסתיים, תגובות התנהגותיות, ומצבים נפשיים אחרים). מצבים השייכים לאותו סוג נפשי אינם חייבים להשתייך לאותו סוג פיזיקלי, ובכלל, ההשקפה הפונקציונליסטית איננה מייחסת למצבים נפשיים כל אפיון פיזיקלי. הפיזיקה אינה מהותית – הפונקציה היא המהותית. וכך, ההשקפה הפונקציונליסטית מאפשרת ריבוי מימושים של מצבים נפשיים.

האם פירוש הדבר שעל פי ההשקפה הפונקציונליסטית מצב נפשי עשוי להיות בעל כל אפיונים פיזיקליים שהם, יהיו אשר יהיו? האם לפיזיקה אין כל חשיבות באשר למהות המצבים הנפשיים, לפי השקפה זו? כזכור, העלינו את השאלה כיצד ההשקפה המעורבת הפיזיקליסטית-התנהגותית מאפשרת ריבוי מימושים: כיצד ייתכן שמצבים בעלי מבנה חומרי שונה לחלוטין יהיו אחראים מבחינה סיבתית לאותו מתאם בין גירויים לתגובות? התשובה שנתנו לנו הייתה שיש דוגמאות שונות של אובייקטים העשויים מחומרים שונים לחלוטין ואשר מקיימים מתאמי גירוי-תגובה זהים, למשל, מכשירי רדיו שונים לחלוטין אשר בקולטם גלים אלקטרומגנטיים בתדירויות זהות (הגירויים) מנפיקים צלילים זהים (אלו התגובות). ניתן לשאול שאלה מקבילה באשר להשקפה הפונקציונליסטית: כיצד ייתכן שמצבים בעלי מבנה חומרי שונה לחלוטין יממשו את אותה פונקציה ויקיימו את אותם קשרים סיבתיים (עם גירויים חושיים, תגובות התנהגותיות, ומצבים נפשיים אחרים)?

לשאלה מקבילה זו ניתן להשיב תשובה מקבילה: יש דוגמאות שונות של מערכות העשויות מחומרים שונים לחלוטין, אשר מממשות את אותה פונקציה – מקיימות רשתות מקבילות לחלוטין של קשרים בין "גירויים" חיצוניים, תגובות "התנהגותיות" ומצבים פנימיים. למעשה, מכשיר הרדיו מהווה דוגמה גם למערכת כזו. אולם דוגמה מובהקת יותר היא דוגמת המחשב: מחשבים העשויים מחומרים שונים מאוד זה מזה עשויים לממש את אותן פונקציות – אותן רשתות של קשרים בין מצביהם הפנימיים, "גירויים" חיצוניים ו"תגובות התנהגותיות". בהמשך נעסוק בהרחבה בדוגמת המחשב, המשחקת תפקיד חשוב מאוד בהשקפה הפונקציונליסטית. עתה אנו מזכירים אותה רק כדי להבהיר באמצעותה כי אין קושי עקרוני ברעיון שמצבים בעלי מבנה חומרי שונה מאוד זה מזה עשויים לממש את אותה פונקציה (אותה רשת קשרים במובן האמור); וכך, אם אמנם, כפי שגורסת ההשקפה הפונקציונליסטית, מצבים נפשיים מאופיינים על ידי הפונקציות שהם מממשים, הרי אין קושי עקרוני ברעיון שמצבים נפשיים עשויים להיות ממומשים על ידי מצבים השונים זה מזה במידה ניכרת מבחינה חומרית. **התורה הפונקציונליסטית מאפשרת אפוא ריבוי מימושים של מצבים נפשיים.**

אולם בכך עדיין לא השבנו לשאלה שהוצגה לעיל: האם פירוש הדבר שעל פי ההשקפה הפונקציונליסטית מצב נפשי עשוי להיות בעל כל אפיונים פיזיקליים שהם? האם אין כל

מגבלה על הטבע הפיזיקלי של המצבים המממשים תכונות נפשיות? התשובה לשאלה זו היא במובן אחד חיובית, ובמובן אחר שלילית. מבחינה תיאורטית, על פי הפונקציונליזם מצב נפשי אמנם עשוי להיות בעל כל אפיונים פיזיקליים שהם. הוא עשוי להיות בעל כל אפיונים פיזיקליים שהם, כל עוד יממש את הפונקציה הדרושה. במובן זה התשובה חיובית. אולם, אף על פי שלפי הפונקציונליזם לטבעו החומרי של מצב אין תפקיד מהותי באפיונו כמצב נפשי (וכמצב נפשי מסוים), בכל-זאת יש לו תפקיד. זאת משום שלא כל גוף חומרי יכול לממש כל פונקציה. הפונקציות הנפשיות – הפונקציות אשר בהן כרוכים המצבים הנפשיים – הן פונקציות מורכבות למדי (ולו במובן זה שגורמים רבים רלוונטיים להן, שינוי במצב נפשי מסוים עשוי לגרור שינויים ניכרים במצבים נפשיים אחרים, או בתגובות התנהגותיות לגירויים חושיים מסוימים, וכדומה). כדי שיוכל לממש פונקציה מורכבת, על גוף חומרי להיות בעצמו מורכב, ועליו להיות בעל מורכבות מסוימת דווקא ולא אחרת. מדחום המגיב לשינויי הטמפרטורה בסביבה באמצעות שינויים באורך עמוד הכספית שלו מממש פונקציה מסוימת (פונקציה הקושרת בין הטמפרטורה לבין אורך עמוד הכספית), אבל אין הוא יכול לממש פונקציות מסובכות ומורכבות. כדי לממש את הפונקציות הנפשיות שלנו, נדרש גוף חומרי בעל דרגת מורכבות גבוהה מאוד, ואומנם המוח האנושי הוא גוף חומרי בעל דרגת מורכבות גבוהה מאוד.

דומה שכדי להיות בעל מצבים נפשיים כשלנו – כדי לממש את הפונקציות הכרוכות במצבינו הנפשיים – נדרש גוף חומרי בעל דרגת מורכבות דומה לזו של המוח האנושי. אין פירוש הדבר שרק המוח האנושי יכול לממש פונקציות אלו – בהחלט תיתכנה מערכות שונות ממנו מבחינה חומרית המסוגלות לממש את אותן פונקציות. אולם לא כל מערכת מתאימה לכך. מערכות בעלות מבנים חומריים מסוימים מסוגלות לממש את הפונקציות הנפשיות, ואילו מערכות בעלות מבנים חומריים אחרים אינן מסוגלות לממש את הפונקציות הנפשיות. אם כן, המבנה הפיזיקלי של היצורים הנפשיים ומצביהם אמנם אינו מהותי לחיי הנפש של היצורים הללו, אך עם זאת הוא בהחלט רלוונטי: לא כל מבנה פיזיקלי יאפשר חיים נפשיים; ובפרט, לא כל מבנה פיזיקלי יאפשר חיים נפשיים כשלנו. המבנה הפיזיקלי רלוונטי לחיים הנפשיים במידה שהוא רלוונטי למימושן של הפונקציות הנפשיות.

התיישבותה של ההשקפה הפונקציונליסטית עם אפשרות ריבוי המימושים אינה מאפיין שולי של השקפה זו. למעשה, אחת המוטיבציות החשובות שעמדו ביסוד הפונקציונליזם הייתה הרצון להיענות לאתגר שהציבה חשיפתה של תופעת ריבוי המימושים: לאתר קריטריון לזהותם של מצבים נפשיים שמתיישב עם תופעת ריבוי-המימושים. במילים אחרות, אחת המוטיבציות שעמדו ביסוד הפונקציונליזם הייתה הרצון להציע קריטריון "ליברלי" לזהותם של מצבים נפשיים – קריטריון שלא ילקה ב"שוביניזם" של פיזיקליזם-סוג, שוביניזם אשר התבטא בכך, שלפי השקפה זו, רק יצורים אשר הפיזיקה שלהם דומה לשלנו עשויים להיות בעלי מצבים נפשיים מן הסוגים שיש לנו. ראינו ביחידה 4-5 כי העמדה של פיזיקליזם-סוג הואשמה בכך, שעל פיה תמנונים, למשל, אינם יכולים להימצא במצב של כאב, חייזרים שמוחם עשוי מסיליקון לא יוכלו להאמין האמנות שונות על אודות סביבתם, וכדומה; בעיני

המבקרים, זו תוצאה שאין לקבלה. דומה כי הפונקציונליזם אינו נתקל בבעיה זו: העובדה שהפיזיקה של היצורים הללו שונה מאוד מן הפיזיקה שלנו אינה מהווה מכשול בפני האפשרות שיהיו להם מצבים נפשיים כשלנו – כאבים, האמנות, וכדומה.

בשלב מאוחר יותר נראה כי יש כאלו המבקרים את הפונקציונליזם בטענה שהוא **ליברלי מדי** – דהיינו, שנובע ממנו כי מערכות שונות שדומה כי אין להן כלל מצבים נפשיים, הן בעלות מצבים נפשיים (והדוגמה המובהקת של מערכות כאלו היא דוגמת המחשב, שנדון בה בהרחבה בהמשך היחידה). לעומת זאת, יהיו גם מי שמנסים להראות כי הפונקציונליזם, או גרסה מסוימת שלו, גם הם שוביניסטים.

נחזור (במילים אחרות) על רעיון שהצגנו ביחידה 1: אחד המבחנים להלימותה של השקפה בדבר טבעם של המצבים הנפשיים הוא שאין היא שוביניסטית ואין היא ליברלית מידי; דהיינו, מחד גיסא, שלא מתחייב ממנה שיצורים מסוימים שאנו יודעים שהם בעלי מצבים נפשיים הם חסרי מצבים נפשיים, ומאידך גיסא, שלא מתחייב ממנה שיצורים או מערכות שאנו יודעים שאין להם מצבים נפשיים הם בעלי מצבים נפשיים. כך, מרבית הפילוסופים ידחו השקפה שעל פיה לכלבים או לאוטיסטים, למשל, אין מצבים נפשיים,<sup>10</sup> ומרבית הפילוסופים ידחו השקפה שעל פיה למדי חום או לווסתי חום, למשל, יש מצבים נפשיים. ההשקפה הראשונה תיחשב שוביניסטית, וההשקפה השנייה תחשב ליברלית מידי.

בהקשר של פונקציונליזם וריבוי המימושים עשויה להתעורר השאלה, האם בכלל מחויבת ההשקפה הפונקציונליסטית – ההשקפה המאפיינת מצבים נפשיים באמצעות הפונקציות שהם מממשים – לכך שמצבים נפשיים הם מצבים **פיזיקליים**? האם מחויבותה לאפשרות של ריבוי מימושים אינה מגעת עד כדי כך, שהמצבים הנפשיים – דהיינו המצבים המממשים את הפונקציות הנפשיות – אינם חייבים כלל להיות מצבים פיזיקליים, והם עשויים גם להיות מצבים רוחניים? נדון בשאלה חשובה זו בהמשך היחידה.

<sup>10</sup> כן, יש פילוסופים שזו השקפתם. אנו נזכיר השקפה כזו ביחידה 9.

## 2.1 פונקציונליזם ומעגליות

העובדה שההשקפה הפונקציונליסטית מאפיינת מצבים נפשיים, בין היתר, על פי יחסיהם למצבים נפשיים אחרים, מעלה את התמיהה הזו, שדיון בה עשוי להבהיר יסודות נוספים של השקפה זו: האם אין האפיון הפונקציונליסטי של מצבים נפשיים **אפיון מעגלי**?

### שאלה

התוכלו להעלות בדעתכם מדוע ניתן לטעון כנגד האפיון הפונקציונליסטי של מצבים נפשיים שהוא אפיון מעגלי?

כאשר ניתן אפיון פונקציונליסטי של מצב נפשי מסוג מסוים במונחים של קשריו, בין היתר, עם מצבים נפשיים אחרים, הרי, כדי שבאמת יובהר באופן זה טבעו של המצב הנפשי הנדון, עלינו לאפיין גם את המצבים הנפשיים האחרים: אם אנו מסבירים את  $X$  באמצעות (בין היתר) קשריו עם  $Y$ , הסבר זה אינו מועיל, אם איננו יכולים להסביר את טבעו של  $Y$ . ולפי הגישה הפונקציונליסטית, הדרך לאפיין את המצבים הנפשיים האחרים היא, כמובן, אותה דרך: ציון הקשרים שלהם עם (בין היתר) מצבים נפשיים אחרים... למשל, על פי הדוגמה דלעיל של ההאמנה שיורד גשם, אפיונה הפונקציונליסטי כרוך, בין היתר, בהתייחסות לרצייה להימנע מלהירטב. וכדי שאפיון ההאמנה הנדונה על ידי התייחסות לרצייה זו יהיה מועיל ואינפורמטיבי, עלינו לאפיין רצייה זו, והדרך לאפיין, על פי הפונקציונליזם, היא באמצעות קשריה עם (בין היתר) מצבים נפשיים אחרים, שגם אותם עלינו לאפיין, וגם זאת באמצעות (בין היתר), קשריהם למצבים אחרים, וכך הלאה וכך הלאה... הבעיה חמורה במיוחד, מכיוון שבשלב כלשהו יהא עלינו להתייחס גם למצב הנפשי המקורי – זה שאפיונו היווה את ראשית התהליך (זכרו את האופי ההוליסטי של רשת הקשרים בה כרוכים מצבינו הנפשיים); וכך נמצא את עצמנו מאפיינים מצב נפשי מסוים באמצעות (בין היתר) אפיון מצב נפשי זה עצמו. במילים אחרות, **האפיון הפונקציונליסטי של מצבים נפשיים כרוך במעגליות**; ודומה שאין הוא מסוגל לספק לנו הבהרה אמיתית של טבעם של מצבים נפשיים.

מצב ביש זה מיוחס לניסיון הפונקציונליסטי לאפיין **סוגים ספציפיים** של מצבים נפשיים, כדוגמת ההאמנה שכדאי להצטייד במים לעתות חרום, או הרצייה לשמור על הבריאות. אולם נראה כי גם התפיסה הפונקציונליסטית של **מהות היסוד הנפשי עצמו** – מהו מצב נפשי לפי תפיסה זו (ולא רק מהו, לפיה, מצב נפשי מסוג ספציפי כלשהו) – נקלעת למצב ביש דומה, גם היא נקלעת לבעיה של מעגליות: שהרי הניתוח הפונקציונליסטי של השאלה "מהו מצב נפשי" – הניתוח על פיו מצבים נפשיים מאופיינים כמצבים המקימים קשרים סיבתיים מסוימים עם גירויים חושיים, תגובות התנהגותיות ומצבים נפשיים אחרים – הוא **ניתוח של מהות המצבים הנפשיים במונחי מצבים נפשיים**. על פי ניתוח זה, **מצב נפשי** הוא מצב שקשור בקשרים מסוימים עם... ועם... **ועם מצבים נפשיים**. כיצד יכול אפוא ניתוח כזה להבהיר מהו מצב נפשי? דומה שהוא מבהיר מהו מצב נפשי רק בעבור מי שכבר יודע מהו מצב נפשי...

וטענה זו משמעותה, כמובן, שניתוח כזה אינו מבהיר כלל מהו מצב נפשי. האם, לפיכך, נאלצים הפונקציונליסטים לוותר על היומרה (יומרה שאפיינה את הביהביוריסטים, למשל) להבהיר את טבעו של היסוד הנפשי **מבלי להשתמש במונחים נפשיים**?

פילוסופים פונקציונליסטים מאמינים כי ביכולתם להתמודד עם קושיות אלו. את אחת מן הדרכים להתמודדות כזו פיתח הפילוסוף דייוויד לואיס (David Lewis), מחשובי הפונקציונליסטים. נקרא את הקטע הבא שלו, קטע שלכאורה אינו קשור כלל לסוגיית המצבים הנפשיים:

אנו מקובצים בחדר האורחים של בית הכפר. הבלש משחזר את הפשע. כלומר, הוא מציע **תיאוריה** המיועדת להיות ההסבר הטוב ביותר של תופעות שהבחנו בהן: מותו של מר בודי, הדם על הטפט, היותו של הכלב שקט בלילה, העובדה ששעון הקיר ממחר בשבע עשרה דקות, וכך הלאה. הוא פותח בסיפורו:

X, Y ו-Z קשרו קשר לרצוח את מר בודי. שבע עשרה שנה קודם לכן, בשדה הזהב של אוגנדה, X היה שותפו של בודי... בשבוע שעבר Y ו-Z נועצו ביניהם בבר ברדינג... ביום שלישי בלילה ב- 11:17 עלה Y לעליית הגג והתקין פצצת זמן... שבע עשרה דקות מאחר יותר פגש X את Z בחדר הביליארד ונתן לו את צינור העופרת... בדיוק כשהפצצה התפוצצה בעליית הגג, ירה X שלוש יריות לתוך חדר העבודה דרך החלונות הצרפתיים...

וכך זה ממשיך – סיפור ארוך. הבה נניח שזהו משפט קוניונקטיבי אחד ארוך.<sup>11</sup>

הסיפור מכיל את שלושת השמות 'X', 'Y' ו-'Z'. הבלש משתמש בשמות חדשים אלו ללא הסבר, כאילו שידענו את משמעותם. אך אנחנו איננו יודעים אותה. מעולם לא השתמשנו בהם קודם לכן, לפחות לא במובן שיש להם בהקשר הנוכחי. כל מה שאנו יודעים על אודות המשמעות שלהם זה מה שאנחנו מבינים באופן הדרגתי מהסיפור עצמו. הבה נקרא להם **מונחים תיאורטיים (מונחי T, בקיצור)**, משום שהם מוצגים על-ידי תיאוריה.<sup>12</sup> נקרא לשאר המונחים בסיפור **מונחי O**. אלו הם כל שאר המונחים למעט **מונחי T**; אלו הם כל המונחים **הישנים, המקוריים**, שהבנו עוד לפני שהתיאוריה הועלתה. נוכל לקרוא להם המונחים 'הקדם-תיאורטיים' שלנו...

נניח, שלאחר ששמענו את סיפורו של הבלש, נודע לנו שהוא אמתי לגבי שלושה אנשים מסוימים: פלאם, פיקוק ומסטרב. אם נציב את השם 'פלאם' במקום 'X', את השם 'פיקוק' במקום 'Y' ואת השם 'מסטרב' במקום 'Z' לאורך כל הסיפור, נקבל סיפור אמיתי על מעשיהם של שלושה אנשים אלו. אנו נאמר שפלאם, פיקוק ומסטרב **מממשים** יחדיו את (או מהווים **מימוש** של) סיפורו של הבלש.

<sup>11</sup> "משפט קוניונקטיבי" משמע משפט המורכב ממשפטים שונים המקושרים ביניהם על-ידי וו החיבור או על-ידי מילת חיבור אחרת. "השמש זורחת והפרחים פורחים" הוא משפט קוניונקטיבי.

<sup>12</sup> התיאוריה שבה מדובר היא התיאוריה שמעלה הבלש על אודות הרצח.

אנו אף עשויים לגלות שהסיפור אינו אמתי לגבי אף שלשה אחרת של אנשים.<sup>13</sup> הציבו בו שלושה שמות כלשהם שאינם שמות של פלאם, פיקוק ומסטרב (בסדר זה), והסיפור שנקבל יהיה שקרי. אנו נאמר שפלאם, פיקוק ומסטרב **מממשים באופן בלעדי** את (מהווים את **המימוש הבלעדי** של) התיאוריה.

אנו עשויים לגלות שתי עובדות אלו (ייתכן שהבלש ידע אותן כל הזמן, אך נמנע מלגלותן כדי להציב את מלכודתו; או שהוא, כמונו, עשוי לגלות אותן רק לאחר שסיפר את סיפורו). ואם נגלה אותן, נסיק בוודאי ש-Y, X ו-Z שבסיפור היו פלאם פיקוק, ומסטרב. אני טוען שנהיה חייבים להסיק זאת, בהינתן מובני המונחים Y, X ו-Z, בשל האופן שבו הבלש מציג אותם בהציגו את התיאוריה שלו, ובהינתן המידע שלנו על פלאם, פיקוק ומסטרב.

בספרו את סיפורו, הציג הבלש שלושה תפקידים, ואמר שמילאו אותם Y, X ו-Z. בעשותו כך הוא ציין בוודאי את המשמעויות של שלושת מונחי ה-T, 'X', 'Y' ו-'Z', שכן הייתה להם משמעות לאחר מכן, לא הייתה להם משמעות לפני כן, ולא נעשה כל דבר אחר, כדי לתת להם משמעות. הם הוצגו באמצעות הגדרה פונקציונלית מובלעת, שיועדה לכנות בשם את ממלאי שלושת התפקידים. כשאנו מגלים מי הם ממלאי שלושת התפקידים, אנו מגלים מי הם Y, X ו-Z. זהו הזיהוי התיאורטי שלנו...<sup>14</sup>

נסה לפצח את מטרת הסיפור שמספר לנו לואיס ולהבין את הרלוונטיות שלו לענייננו.

## שאלה

התוכלו להעלות בדעתכם מה עשוי להיות הקשר בין הסיפור שמספר לואיס לבין **אפיון פונקציונלי** של מצבים נפשיים?

ובכן, סיפורו של לואיס אמור להדגים אפיון פונקציונלי. שלוש הדמויות הנזכרות – פלאם, פיקוק ומסטרב – מאופיינות באמצעות אפיונים פונקציונליים. המשפטים המופיעים בסיפור הנסבים על אודות Y, X ו-Z, אמורים לספק אפיונים פונקציונליים של כל אחת מן הדמויות הללו. באיזה מובן אמורים משפטים אלו לספק אפיונים פונקציונליים של הדמויות? הם מאפיינים את הדמויות באמצעות **התפקידים שיש להן בעלילה** (על פי התיאוריה שמעלה הבלש). תפקידיהן בעלילה, בהקשר זה, הם היחסים שמתקיימים ביניהן לבין גורמים נוספים בעלילה וביניהן לבין עצמן. אוסף המשפטים הנאמרים על X, למשל – העובדה שהוא היה שותפו של מר בודי, העובדה שהוא פגש את Y בבר, וכולי – מתארים את התפקיד של X בעלילה. אפשר לומר שהמשפטים הללו (או, בלשונו של לואיס, המשפט הקוניונקטיבי המורכב מן המשפטים הללו) קובעים את זהותו של X: X הוא אותו אדם אשר המשפטים הללו אמיתיים באשר אליו; X הוא אותו אדם הממלא את התפקיד המתואר על ידי

<sup>13</sup> [הערה של דייוויד לואיס עצמו] הסיפור עצמו עשוי לרמוז זאת. אם, למשל, נאמר בסיפור: 'X' ראה את Y נותן ל-Z את הפמוט בזמן ששלושתם היו לבד בחדר הביליארד ב-9:17, אז הסיפור אינו יכול להיות אמתי עבור יותר משלושה אחת של אנשים.

<sup>14</sup> D. Lewis, "Psychophysical and Theoretical Identifications", *Australasian Journal of Philosophy* 50 (1972), pp. 249-258.

המשפטים הללו. **פלאם** הוא  $X$  משום שפלאם הוא האדם היחיד הממלא את התפקיד המיוחס ל- $X$  בסיפור. אפשר להסיק ש- $X$  זהה לפלאם באופן הבא:  $X$  הוא אדם המאופיין כמי שממלא תפקיד כזה וכזה בעלילה; פלאם הוא היחיד הממלא תפקיד זה בעלילה; לפיכך  $X$  הוא פלאם. הסיפור מתאר את פלאם מבלי לנקוב בשמו, על ידי תיאור אוסף עובדות הנכונות באשר אליו ורק באשר אליו.

שימו לב שהיסק זה דומה במבנהו (ולא במקרה, כמובן) להיסק הנזכר של ארמסטרונג ולואיס על פיו מצבים נפשיים הם מצבים פיזיקליים מאחר שמצבים פיזיקליים מסוימים הם אלו הממלאים את התפקידים הסיבתיים המאפיינים מצבים נפשיים. היסק כזה ניתן להחלה גם על מצבים נפשיים ספציפיים: למשל, כאב הוא גירוי סיבי  $C$ , משום שהמצב של גירוי סיבי  $C$  ממלא את התפקיד הסיבתי המאפיין כאבים.

הסיפור מסופר במונחים של  $X$ ,  $Y$  ו- $Z$  וזהות האנשים הספציפיים איננה נתונה באופן מפורש. הסיפור אינו מספק לנו ישירות את המידע ש**פלאם** עשה כך וכך אלא ש- $X$  עשה כך וכך, ואינו אומר לנו ש- $X$  הוא פלאם. מה, אם כך, פשר הטענה שבסיפור ש- $X$  עשה כך וכך? פשרה הוא, **שקיים אדם** (אחד) אשר עשה כך וכך. את הסיפור כולו, כאשר הוא מסופר במונחים של  $X$ ,  $Y$  ו- $Z$ , יש להבין כך: קיים אדם  $X$  שעשה כך וכך, וקיים אדם  $Y$  שעשה כך וכך, וקיים אדם  $Z$  שעשה כך וכך.

הסיפור מראה שניתן לאפיין דברים למיניהם (במקרה זה אלו הם בני-אדם) באופן הקובע את זהותם – מבלי לנקוב בשמותיהם או לאזכר תכונה "פנימית" כלשהי שלהם, תכונות שאינן קושרות אותם לדברים אחרים. ניתן לאפיין דברים על ידי אזכור של תכונות יחסיות שלהם, כאשר אנו מבינים אזכור כזה כאומר: קיים משהו (או מישהו) אשר קשור ביחס זה עם חפץ, ביחס אחר עם פלוני, ביחס אחר עם אלמוני, וכדומה; והאפיון חל על מי שבפועל מקיים יחסים אלו עם הפריטים (או האנשים) הנזכרים. הרעיון הפונקציונליסטי הוא שאפשר לאפיין מצבים נפשיים – למשל כאבים, האמנות, רציות, פחדים, וכדומה – על ידי אזכור תכונות יחסיות שלהם, כלומר קשריהם עם מצבים אחרים. זהו אפיון פונקציונלי.

אולם מהם המצבים האחרים הללו על פי הפונקציונליסטים? המצבים האחרים הללו, על פי הפונקציונליסטים, הם גירויים חושיים, תגובות התנהגותיות, ו**מצבים נפשיים אחרים**. וכאן בדיוק מתעוררת בעיית המעגליות שתיארנו לעיל! כדי ש"סיפור" על כאב – סיפור המתאר את קשריו למצבים אחרים – יהיה אינפורמטיבי, כדי שיהיה בעל משמעות, עלינו לדעת מה הם המצבים האחרים הללו, אולם כפי שראינו, אפיון של  $X$  במונחים של קשריו עם  $Y$  ו- $Z$  אינו אפיון יעיל – למעשה, אין הוא באמת אפיון היכול לזהות את  $X$  – אם אין בידינו אפיונים של  $Y$  ו- $Z$  עצמם. אבל האפיונים של  $Y$  ו- $Z$  עצמם ניתנים בסיפור, בין היתר, באמצעות קשריהם של  $Y$  ו- $Z$  בינם לבין עצמם ובינם לבין  $X$ . כך, הזיהוי של  $X$  תלוי בזיהוי של  $Y$  ו- $Z$ , שתלוי בזיהוי של  $X$ ... זוהי המעגליות!



אבל שימו לב שבסיפור של לואיס המעגליות כלל איננה מהווה בעיה; היא אינה מפריעה לזיהוי של  $Y, X$  ו- $Z$  עם הדמויות הספציפיות (פלאם, פיקוק ומסטרב). המידע שמספקת העלילה מספיק כדי שנוכל לזהותן. מדוע? משום שהקשרים שמתקיימים על פי העלילה בין שלוש הדמויות הללו, והמאפיינים הנוספים שמאפיינים אותם (קשריהם לפריטים אחרים), דיים לייחד את השלישייה הזו. ההנחה של הסיפור היא שיש רק שלישייה אחת של אנשים אשר ביניהם מתקיימים קשרים כאלו (והמקיימים קשרים כאלו עם פריטים אחרים מסוימים). העלילה נכונה באשר לשלישייה זו של אנשים – היא נכונה באשר לפלאם, לפיקוק ולמסטרב – ורק באשר לשלישייה זו. ולכן ניתן לזהות שהדמויות הללו הן פלאם, פיקוק ומסטרב על-סמך העלילה, אף על פי שהעלילה אינה מזכירה אף אחת מהן בשמה, ואף על פי שהאפיון של כל דמות מן השלישייה מסתמך על אפיון של הדמויות האחרות. **שלוש הדמויות מאופיינות יחדיו, בו זמנית!**

דוגמה נוספת תסייע להבהרת העניין: נניח שכתבה מסוימת מתארת את  $X$  כאחיו של  $Y$ , המבוגר ממנו ב-30 שנה, ו- $Y$  מתואר כאחיו של  $Z$ , המבוגר ממנו ב-25 שנה. ל- $X$  דירה ברחוב הרצל בבאר-שבע, ול- $Z$  קיוסק ברחוב מצדה בבאר-שבע. האם נוכל לזהות מיהם  $Y, X$  ו- $Z$  על-סמך מידע זה? נניח שיש עוד שלישיות של אחים שביניהם פערי גילים כאלו, ונניח שיש עוד שלישיות אחים שלאחד מהם דירה ברחוב הרצל בבאר-שבע, ולאח נוסף קיוסק ברחוב מצדה בבאר-שבע. אולם, אם יש רק שלישייה אחת של אחים שביניהם מתקיימים פערי גילים כאלו ואשר לאחד דירה ברחוב הרצל בבאר-שבע ולאחר קיוסק ברחוב מצדה בבאר-שבע, הרי מידע זה מספיק כדי לזהות את השלישייה. הוא מספיק לשם זיהוי השלישייה, אף על פי שחלק מן המידע החיוני הוא מידע בדבר קשריהם של האחים בינם לבין עצמם. האפיונים כולם מזהים את השלישייה כולה – הם מזהים את שלושת האחים יחדיו, בו-זמנית.

על פי הפונקציונליסטים, מצבים נפשיים ניתנים לזיהוי בדיוק באופן זה. אם אנו מתארים את רשת הקשרים של מצב נפשי מסוים עם גירויים חושיים, תגובות התנהגותיות, ומצבים נפשיים אחרים, תיאור זה יזהה את המצב הנפשי, אף על פי שהוא כולל התייחסות למצבים נפשיים אחרים שמאופיינים גם הם על ידי קשריהם עם (בין היתר) מצבים נפשיים אחרים, משום שתיאור מלא כזה מייחד את כל רשת המצבים הנפשיים. כל המצבים הנפשיים מאופיינים יחדיו, בו-זמנית, באמצעות תיאור מלא של רשת הקשרים בינם לבין עצמם ובינם לבין גירויים חושיים ותגובות התנהגותיות. אפיון של מצב נפשי – היותו כאב, האמנה שירד גשם, רציה בגלדת תות, וכדומה – הוא, לפי ההשקפה הנדונה, אפיון פונקציונליסטי: מה שמשייך מצב נפשי לסוגו אלו הם קשריו עם המצבים האחרים הללו. אולם למרות שאפיון של מצב נפשי תלוי גם באפיוניהם של מצבים נפשיים אחרים, ואפיוניהם תלויים גם באפיון שלו, אין בכך מכשול לזיהוי המצב, כי תיאור רשת הקשרים כולה מאפיינת את כל המצבים הנפשיים הקשורים ביניהם יחדיו, בו-זמנית. מידע זה מייחד את קבוצת המצבים הנפשיים כולה ומזהה אותה.

על פי הפונקציונליזם, כל המצבים הנפשיים מאופיינים יחדיו על ידי תיאור מלא של רשת הקשרים בינם לבין עצמם ובינם לבין גירויים חושיים ותגובות התנהגותיות.

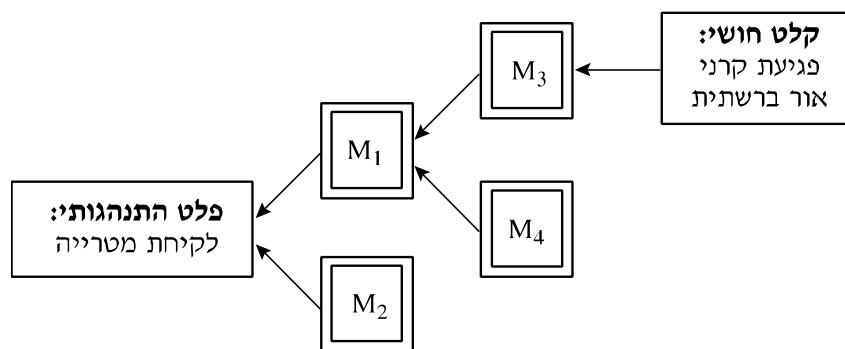
אם אכן המידע בדבר רשת הקשרים של מצב נפשי מזהה את קבוצת המצבים הנפשיים כולה, הרי שכך מזהה גם כל מצב נפשי ספציפי. אם זיהינו את שלוש הדמויות בעלילה של לואיס, הרי זיהינו גם את  $X$ , למשל, משום של- $X$  תפקיד מסוים בסיפור – תפקיד השונה מזה של שתי הדמויות האחרות. כך, תיאור רשת הקשרים שבה כרוך מצב נפשי מסוים מזהה את המצב הנפשי הזה: **הוא מזהה על ידי תפקידו הספציפי ברשת הקשרים הזו, או, כפי שנהוג לומר, על ידי "מיקומו ברשת"**. מצב הוא כאב, אם הוא משחק תפקיד מסוים, או ממוקם במקום מסוים, ברשת קשרים כזו; מצב הוא רצייה בשוקולד, אם הוא ממוקם במקום אחר ברשת קשרים כזו, וכדומה. למצב נפשי מכל סוג שהוא מקום ברשת הקשרים תפקיד אופייני, תפקיד המקנה לו את זהותו ככאב, כרצייה בשוקולד, וכדומה.

על פי הפונקציונליזם, מצבים נפשיים מאופיינים על ידי תפקידם או "מיקומם" ברשת הקשרים בינם לבין עצמם ובינם לבין גירויים חושיים ותגובות התנהגותיות.

שימו לב, שהאנשים בעלילה של לואיס, או שלושת האחים שהזכרנו, מזהים על-סמך המידע מבלי שאף אחד מהם מוזכר בשמו – הם מצוינים, כזכור, כ- $X$ ,  $Y$ , ו- $Z$ . מה חשיבותה של עובדה זו באשר לפונקציונליזם? ובכן, באותו אופן ניתן לתאר את רשת הקשרים של המצבים הנפשיים מבלי להזכיר בשמו אף סוג של מצב נפשי – מבלי להשתמש באף מונח נפשי (מונחים כמו "כאב", "האמנה", וכדומה): אנו יכולים לומר פשוט שיש קבוצה של מצבים נפשיים –  $X$ ,  $Y$ ,  $Z$ , וכדומה – כאשר  $X$  נגרם על ידי גירוי חושי מסוים, ויחד עם  $Y$  הוא גורם ל- $Z$ , וכך הלאה. ואם כך, אם תיאור רשת הקשרים אינו חייב לכלול כל אפיון נפשי, הרי שהאפיון הפונקציונלי של המצבים הנפשיים אינו נתקל בבעיה של מעגליות – בעיית מעגליות קיימת אם בבואנו לאפיין מצב נפשי, אנו **חייבים** להשתמש באפיון נפשי. אולם אם האפיון הפונקציונליסטי של מצבים נפשיים יכול להיפטר מאפיונים נפשיים – אם אפשרי אפיון פונקציונליסטי של מצבים נפשיים אשר מזהה אותם רק באמצעות קשריהם למצבים אחרים, ולא בתור מצבים נפשיים, הרי שאפיון כזה אינו מאוים על ידי בעיית מעגליות; אפיון כזה מסביר את היסוד הנפשי, בסופו של דבר, מבלי להסתמך על אפיונים נפשיים. ניתן לומר שהוא מעמיד את היסוד הנפשי על יסודות שאינם נפשיים – על קשריהם ההדדיים של המצבים הנפשיים ועל קשריהם עם גירויים חושיים ותגובות התנהגותיות.

כדי להמחיש רעיונות אלו, נפנה לדוגמה קונקרטית. נחזור לתרשים שהצגנו קודם, המתאר מערכת קשרים ספציפית בין מצבים נפשיים לבין עצמם ובינם לבין גירויים חושיים ותגובות התנהגותיות. בתרשים זה צוינו המצבים הנפשיים על ידי מונחים נפשיים: "האמנה שיורד

גשם", רצייה שלא להירטב".... עתה נציין את המצבים הללו באמצעות אותיות בלבד: את ההאמנה שירד גשם נציין כ- $M_1$ , את הרצייה שלא להירטב נציין כ- $M_2$ , וכו'. מאחר ותיאורי הגירויים החושיים והתגובות ההתנהגותיות אינם כוללים מונחים נפשיים, נוכל להשאירם כפי שהם. באופן זה אנו מציגים את רשת הקשרים שבה כרוכים המצבים הנפשיים מבלי להשתמש בכל מונח נפשי שהוא.<sup>15</sup>



**תרשים 5.** רשת קשרים של מצבים נפשיים, גירויים חיצוניים ותגובות התנהגותיות

כמובן, רשת הקשרים שבה כרוכים מצבינו הנפשיים סבוכה הרבה יותר – התרשים מציג תמונה חלקית בלבד. לפי ההשקפה הפונקציונליסטית, ההאמנה שירד גשם היא המצב שתופס את המקום של  $M_1$  בכל מערכת נפשית שמאופיינת על ידי רשת של קשרים מעין זו (אולם, כאמור, סבוכה יותר). כלומר, הטענה שיצור נפשי מסוים מאמין ברגע מסוים שירד גשם שקולה לטענה שיצור נפשי זה מאופיין על ידי רשת קשרים מעין זו (אולם סבוכה יותר), ושהוא נמצא כרגע במצב  $M_1$ . דהיינו הוא נמצא במצב "שמיקומו ברשת" הוא מיקומו של  $M_1$  – במצב שקשור עם מצבים אחרים באופן שבו  $M_1$  קשור עם מצבים אחרים, כפי שמתאר התרשים. בדומה, ההאמנה שירד גשם מתוארת כמצב שעומד ביחסים כאלו וכאלו עם מצבים אחרים – אף אחד מהמצבים הללו אינו מתואר במונחים נפשיים, נאמר עליו רק שהוא מצב המקיים קשרים סיבתיים כאלו וכאלו; הוא מאופיין אפוא באמצעות התפקיד הסיבתי שהוא ממלא.

כך, ההשקפה הפונקציונליסטית – ההשקפה על פיה מצבים נפשיים מאופיינים על ידי הפונקציות שלהם בחייו הנפשיים של הפרט – אינה **מניחה** אפיונים נפשיים כלשהם, והאפיון של כל מצב נפשי ספציפי איננו כרוך באזכור מצבים נפשיים אחרים **בתור** מצבים נפשיים; הוא אינו מזכיר שום מונחים נפשיים שהם. הרעיון הכללי הוא שאומנם איננו יכולים לאפיין מצב נפשי באופן שאיננו תלוי במצבים נפשיים אחרים ואינו מתייחס אליהם; אבל המצבים הנפשיים הללו אינם מתוארים או מאופיינים כמצבים נפשיים. המצבים מאופיינים רק

<sup>15</sup> הרעיון שביסוד תרשים זה לקוח מתוך:

J. Heil, *Philosophy of Mind: A Contemporary Introduction*, London: Routledge, pp. 100-101.

באמצעות יחסיהם זה לזה, ולגירויים חושיים ולתגובות התנהגותיות, ובאופן זה כולם מאופיינים בו-זמנית באמצעות רשת הקשרים הזו ומיקומם או תפקידם בה.

בלשון הנהוגה בלוגיקה, כשאנו מציינים את המצבים הנפשיים באמצעות  $M_1, M_2$  וכדומה, אנו מציינים אותם באמצעות **משתנים** (דהיינו,  $M_1, M_2$  כדומה הם משתנים). הטכניקה של ההיפטרות מן המונחים הנפשיים והחלפתם במשתנים כרוכה בכך שבמקום לומר, למשל, "ההאמנה שיוֹרד גשם גורמת [בתנאים מסוימים שהתיאור מפרט] רצייה לקחת מטרייה... והיא נגרמת על ידי...", אנו אומרים: "ישנו מצב,  $M_1$ , וישנו מצב,  $M_2$ , וישנו... ו- $M_1$  גורם (בתנאים מסוימים שהתיאור מפרט) ל- $M_2$ , ונגרם על ידי..." טכניקה כזו, שניתן להחילה על תחומים שונים (לאו דווקא על התחום הנפשי) מכונה **"רמזיפיקציה"** (Ramsification), או "תרגום למשפטי רמזי", על-שם הפילוסוף והלוגיקן פרנק רמזי (Frank Ramsey) שפיתח אותה (לואיס החיל טכניקה זו לעניין האפיונים הפונקציונליסטיים של מונחים נפשיים).

אם כן, כך מוסבר היסוד הנפשי על ידי ההשקפה הפונקציונליסטית מבלי להסתמך על כל אפיון נפשי שהוא. הן היותו של מצב פנימי מסוים מצב נפשי, והן השתייכותו לסוג נפשי מסוים – כאב, רצייה בעוגת גבינה, וכדומה – מוסברים באופן זה, ללא מעגליות: מצב "פנימי" (דהיינו, מצב שאינו מצב של גירוי המערכת מבחוץ, או תגובה של המערכת כלפי חוץ) הממוקם ברשת הקשרים המתאימה הוא מצב נפשי; מיקומו הספציפי של מצב פנימי ברשת הקשרים מקנה לו את אופיו הנפשי הספציפי ומשייכו לסוגו הנפשי.

**על פי הפונקציונליזם, האפיון של כל מצב נפשי איננו כרוך באזכור מצבים נפשיים אחרים כמצבים נפשיים, ואינו משתמש בשום מונחים נפשיים.**

כיוון שהיסוד הנפשי מוסבר על ידי ההשקפה הפונקציונליסטית מבלי להסתמך על כל אפיון נפשי שהוא, ניתן לומר שההשקפה הפונקציונליסטית היא **השקפה רדוקציוניסטית**. עם זאת, היומרה הרדוקציוניסטית של הפונקציונליזם חלשה מן היומרה הרדוקציוניסטית של הביהביוריזם, משום שהאפיון הפונקציונליסטי של מצב נפשי כרוך באזכור מצבים נפשיים אחרים (אם כי, כפי שאך ראינו, אין הוא כרוך באזכורם כמצבים נפשיים). האפיון הביהביוריסטי של מצב נפשי, כזכור, לא היה כרוך כלל בהתייחסות למצבים נפשיים אחרים (אם כי בעניין זה כרוך קושי ניכר).

במובן חשוב, הרדוקציה הפונקציונליסטית של היסוד הנפשי ליסודות שאינם נפשיים **אינה רדוקציה פיזיקליסטית**, שהרי על פי הפונקציונליסטים מצב נפשי מסוג מסוים אינו ניתן לזיהוי עם מצב מסוג פיזיקלי מסוים.<sup>16</sup>

נקודה נוספת בעניין טיבה של הרדוקציה הפונקציונליסטית קשורה בשאלה האם היא רדוקציה **לוגית** או רדוקציה **אונטולוגית** בלבד. כשנדון, בהמשך, בגרסאות שונות של פונקציונליזם, נראה כי התשובה לשאלה זו משתנה בהתאם לגרסה הפונקציונליסטית שבה מדובר.

#### תזכורת:

רדוקציה לוגית היא יחס בין מילים או משפטים: מילים או משפטים המציינים ישויות מסוג אחד הם רדיוסיביליים ל- (במילים אחרות, ניתנים להעמדה על) מילים או משפטים המציינים ישויות מסוג אחר, אם אפשר לתרגם את המילים והמשפטים מן הקבוצה הראשונה (מבלי לאבד דבר ממשמעותם) לאלו מן הקבוצה השנייה.

רדוקציה אונטולוגית היא יחס בין אובייקטים או מצבים, או תכונות, או מאפיינים כלשהם של ישויות: אובייקטים, או מצבים, או תכונות, או מאפיינים כלשהם מסוג אחד רדיוסיביליים ל- (או ניתנים להעמדה על) אובייקטים, או מצבים, או תכונות, או מאפיינים כלשהם מסוג אחר, אם המאפיינים מן הסוג הראשון אינם אלא מאפיינים מן הסוג השני. לדוגמה, שולחנות ניתנים להעמדה על אוספי מולקולות, אם שולחנות אינם אלא אוספי מולקולות.

אנו רואים, שהפונקציונליסטים אינם מתיימרים לספק אפיונים מבודדים של מצבים נפשיים, כלומר, אפיונים שאינם מזכירים מצבים נפשיים אחרים (זאת, כזכור, הייתה היומרה הביהביוריסטית). הפונקציונליסטים סבורים שמתן אפיונים מבודדים כאלו אינו אפשרי, בדיוק משום שמהותו של מצב נפשי – היותו מצב נפשי והיותו המצב הנפשי מסוגו הספציפי – מכוננת על ידי קשריו עם מצבים אחרים – על ידי מיקומו הספציפי ברשת מסוימת של קשרים בין מצבים שונים – שחלקם מצבים נפשיים אחרים (אולם, כאמור, אפיון הולם של מצב נפשי באמצעות מיקומו ברשת הקשרים אינו חייב להזכיר את המצבים הנפשיים **כמצבים נפשיים**). מה שעושה מצב נפשי למצב נפשי ולמצב נפשי מסוגו הספציפי אלו הן העובדות שהוא "חבר" ברשת קשרים מסוימת ושהוא ממלא ברשת קשרים זו תפקיד מסוים (כפי שראינו, לא רק

<sup>16</sup> אנו אומרים "במובן חשוב", משום שיש פילוסופים הטוענים שבמובן אחר הרדוקציה הפונקציונליסטית הינה פיזיקליסטית. לא נוכל לדון בטענות אלו במסגרת זו.

שאפיונים פונקציונליסטיים של המצבים הנפשיים אינם יכולים להיות מבודדים, אלא שהם הוליסטיים – אפיון פונקציונליסטי מלא של מצב אחד כרוך בהתייחסות **לכל** מצביו הנפשיים האחרים של היצור הנפשי).

אפיונו של מצב נפשי עולה, לפי הפונקציונליזם, מתיאור רשת הקשרים שבין מצב זה לבין גירויים חושיים, תגובות התנהגותיות, ומצבים נפשיים אחרים. תיאורים כאלו מספקת **הפסיכולוגיה**. היא היא שמתארת את רשתות הקשרים שבהן כרוכים מצבינו הנפשיים – את כללי המעבר ממצב אחד למשנהו. "העלילה" של מצבינו הנפשיים, "העלילה" שממנה עולים אפיוניהם, היא התיאוריה הפסיכולוגית החלה עלינו. כפי שנראה, קיימות אפשרויות שונות באשר לזהותה של התיאוריה הפסיכולוגית שממנה עולים אפיוני מצבינו הנפשיים – גישות המהוות גרסות שונות של פונקציונליזם.

### שאלה

בקטע מדברי דייויד לואיס שציטטנו לעיל מוזכרים "**מונחים תיאורטיים**", אלו שאת משמעויותיהם אנו למדים רק מן העלילה, ומונחים אחרים, אלו שאת משמעויותיהם היכרנו אף לפני סיפור העלילה. בעלילה הפסיכולוגית, מהם המונחים התיאורטיים ומהם המונחים האחרים?

המונחים התיאורטיים של העלילה הפסיכולוגית הם המונחים הנפשיים – המונחים המציינים את המצבים הנפשיים שבהם עוסקת התיאוריה. הרעיון הפונקציונליסטי הוא שמהותם של המצבים הנפשיים נתונה לנו באמצעות התיאוריה הפסיכולוגית. המהות של ההאמנה שיוורד גשם, למשל, נתונה לנו על ידי רשת הקשרים אשר, לפי התיאוריה הפסיכולוגית, האמנה זו קשורה בהם (קשרים בינה לבין גירויים חושיים, תשובות התנהגותיות ומצבים נפשיים אחרים). המונחים האחרים הם המונחים המציינים את הגירויים החושיים והתגובות ההתנהגותיות. אולם כפי שנראה בהמשך היחידה, שאלת המונחים האחרים היא נושא סבוך.

## 2.2 פונקציונליזם-פרט ופונקציונליזם-סוג

הטענה שמצבים נפשיים הם מצבים פונקציונליים היא טענה המקובלת על רבים – למעשה כמעט כל מי שחושבים על נושא זה יקבלוה, ובתוכם גם כאלו הדוחים את ההשקפה הפונקציונליסטית. שהרי קשה לחלוק על הטענה שמצבים נפשיים משחקים תפקידים סיבתיים, ובאופן ספציפי יותר, שהם קשורים ברשתות סיבתיות המורכבות מתפיסות חושיות, תגובות התנהגותיות, ומצבים נפשיים אחרים (דומה שרק אפינומנליסטים יחלקו על טענה זו). אף לפילוסופים הדוגלים בפיזיקליזם-סוג אין בעיה להכיר בכך שמצבים נפשיים הם מצבים פונקציונליים: אין קושי בצירוף שתי הטענות הבאות: מחד גיסא, המצבים

הנפשיים ממלאים פונקציה כלשהי בחיי הסובייקט, הם משחקים תפקיד סיבתי – תפקיד של תיווך סיבתי בין תפיסות חושיות, תגובות התנהגותיות, ומצבים נפשיים אחרים; מאידך גיסא, קריטריון הסיווג הנפשי הוא פיזיקלי, כלומר מה שמשייך (פרטים של) מצבים נפשיים לאותו סוג נפשי (למשל, מה שמשותף לכל הכאבים ורק להם) הוא אותה תכונה פיזיקלית. לומר שמצבים נפשיים הם מצבים פונקציונליים משמע לשייך להם **תכונות מסוג אחד** (תכונות פונקציונליות), אולם בכך אין כדי לפסול את האפשרות שתכונות מסוג אחר הן אלו המשייכות אותם לסוגים נפשיים (שתכונות מסוג אחר הן התכונות הנפשיות). שימו לב, שההשקפה הפונקציונליסטית איננה מסתפקת בטענה שמצבים נפשיים הם **מצבים פונקציונליים**. ואומנם, הפילוסופים הפונקציונליסטים טוענים טענה חזקה יותר מן הטענה שמצבים נפשיים הינם מצבים פונקציונליים: הם טוענים **שקריטריון הסיווג הנפשי הוא פונקציונלי**. הם טוענים שמה שמשייך מצב נפשי לסוגו הנפשי (להיותו כאב, להיותו האמנה שהמשיח מבושש להגיע, וכדומה) זו העובדה שהוא משחק תפקיד סיבתי מסוים (תפקיד מסוים של תיווך בין קלטים חושיים, פלטים התנהגותיים, ומצבים נפשיים אחרים); זו העובדה שהוא ממלא פונקציה מסוימת.

דרך פשוטה לבטא את הטענה הפונקציונליסטית החזקה הזו היא לומר שתכונות נפשיות הן תכונות פונקציונליות. אנו מבחינים אם כן בבירור בין הטענה של **מצבים נפשיים יש תכונות פונקציונליות** (טענה שמקבלים, כאמור, גם רבים שאינם פונקציונליסטים), לבין הטענה **שתכונות נפשיות** – התכונות המקנות למצבים נפשיים את אופייהם הנפשיים הספציפיים ומשייכות אותם לסוגיהם הנפשיים – **הן תכונות פונקציונליות**.<sup>17</sup> למעשה, השקפה פונקציונליסטית מלאה מחויבת לכך ש**כל** המאפיינים המהותיים של מצבים נפשיים (כל המאפיינים העושים אותם למצבים הנפשיים שהינם), הם מאפיינים פונקציונליים. לפי השקפה פונקציונליסטית מלאה, היסוד הנפשי על כל היבטיו ניתן להעמדה על מאפיינים פונקציונליים.

ניתן לבטא הבחנה זו גם באופן אחר.

## שאלה

ביחידה 4-5 דנו בהבחנה בין פרט (token) לבין סוג (type) באשר לפיזיקליזם – הבחנו בין פיזיקליזם-פרט לבין פיזיקליזם-סוג. התוכלו לבטא את ההבחנה בין הטענה שלמצבים נפשיים יש תכונות פונקציונליות (במילים אחרות, שהם מצבים פונקציונליים) לבין הטענה שתכונות נפשיות הן תכונות פונקציונליות **במונחים של פרט וסוג**?

<sup>17</sup> יש המשתמשים בביטויים הנזכרים כאן באופנים שונים מכפי שאנו מציגים אותם. לדוגמה, הילארי פטנאם, במאמרו "The Nature of Mental States" – מאמר הנחשב על-ידי רבים כמבטא החשוב ביותר של ההשקפה הפונקציונליסטית – הביע השקפה זו באמצעות הטענה שמצבים נפשיים הם **מצבים פונקציונליים** (ואינם מצבים פיזיקליים, מצבי-מוח). אולם נראה שפטנאם לא התכוון לטענה החלשה יותר, על פיה למצבים נפשיים יש תכונות פונקציונליות, אלא לטענה החזקה יותר על פיה תכונות נפשיות הן תכונות פונקציונליות; הוא התכוון להביע את הרעיון שמה שעושה מצבים נפשיים למצבים נפשיים, ולמצבים הנפשיים שהינם (חרדה, האמנה ששוויון כלכלי הוא ערך חשוב, וכדומה), אלו הם מאפייניהם הפונקציונליים (ולא מאפייניהם הפיזיקליים).

ובכן, אפשר לומר שהטענה שמצבים נפשיים הם מצבים פונקציונליים היא **פונקציונליזם-פרט** (token functionalism), ואילו הטענה שתכונות נפשיות הן תכונות פונקציונליות היא **פונקציונליזם-סוג** (type functionalism). הטענה הראשונה היא פונקציונליזם-פרט משום שעל פיה כל פרט ופרט של מצב נפשי הוא מצב פונקציונלי – היא מתייחסת רק למצבים נפשיים ספציפיים ואינה אומרת דבר על סוגיהם; הטענה השנייה היא פונקציונליזם-סוג משום שעל פיה סוגיהם של מצבים נפשיים (היותם כאבים, פחדים, רציות בשוקולד, וכדומה) נקבעים על ידי תכונותיהם הפונקציונליות של המצבים הללו. ובכן, ההשקפה הפונקציונליסטית היא למעשה **פונקציונליזם-סוג** – היא השקפה בדבר התכונות הנפשיות ובדבר קריטריון הסיווג של מצבים נפשיים. השקפה זו דוחה את קריטריון הסיווג הפיזיקליסטי של המצבים הנפשיים, דהיינו את העמדה של פיזיקליזם-סוג, ומציעה, במקומו, קריטריון סיווג פונקציונליסטי.<sup>18</sup>

**ההשקפה הפונקציונליסטית היא פונקציונליזם-סוג – היא השקפה בדבר מהותן של תכונות נפשיות ובדבר מהותם של סוגים נפשיים. לפי השקפה זו תכונות נפשיות הן תכונות פונקציונליות.**

#### שאלה

הדוגלים בהשקפה הפונקציונליסטית – דהיינו בפונקציונליזם-סוג – מחויבים גם לעמדה של פונקציונליזם-פרט. התוכלו להראות כיצד פונקציונליזם-פרט נובע מפונקציונליזם-סוג?

לפי פונקציונליזם, סוג תכונותיהם הנפשיות של מצבים נפשיים הוא תכונות פונקציונליות. מן הסתם, השקפה זו מחויבת לכך שלמצבים נפשיים יש תכונות פונקציונליות. זוהי בדיוק הטענה של פונקציונליזם-פרט: הטענה שמצבים נפשיים הם מצבים פונקציונליים מפורשת כטענה שיש להם תכונות פונקציונליות, כלומר שהם ממלאים חסרים של תיווך סיבתי בין גירויים חושיים, תגובות התנהגותיות ומצבים נפשיים אחרים.

הטענה כי תכונות נפשיות הן תכונות פונקציונליות וסוגים נפשיים הם סוגים פונקציונליים אינה כה מובנת מאליה כמו הטענה שמצבים נפשיים הם מצבים פונקציונליים. נראה בהמשך האם ניתן להגן על טענה זו.

<sup>18</sup> מי שהבחין בין פונקציונליזם-פרט לבין פונקציונליזם-סוג והציג את ההשקפה הפונקציונליסטית כפונקציונליזם-סוג הוא הפילוסוף דניאל דנט (Daniel C. Dennett), בהקדמה של ספרו *(MIT) Brainstorms* (Cambridge, Mass, 1981, Press).



## 2.3 פונקציונליזם ופיזיקליזם, והבעיה הפסיכופיזית

כפי שצינו זה עתה, ההשקפה הפונקציונליסטית מתנגשת בפיזיקליזם-סוג: היא מציעה קריטריון סיווג חלופי, קריטריון סיווג שאינו פיזיקליסטי. אולם מה באשר לשאלה האונטולוגית? האם ההשקפה הפונקציונליסטית היא השקפה פיזיקליסטית או לא? מה היחס בין השקפה זו לבין ההשקפה של פיזיקליזם-פרט?

ראשית, עלינו לשים לב לכך שההשקפה הפונקציונליסטית מתיישבת עם פיזיקליזם-פרט. הטענה שתכונות נפשיות הן תכונות פונקציונליות מתיישבת עם הטענה שמצבים נפשיים הם מצבים פיזיקליים. אמנם תכונות פונקציונליות אינן תכונות פיזיקליות במובן שבו אנו משתמשים בביטוי "תכונות פיזיקליות" (ודרך אחת להיווכח בכך היא לשים לב שתכונות פונקציונליות ניתנות לריבוי מימושים)<sup>19</sup>; אולם, כפי שראינו ביחידה 4-5, לאובייקטים פיזיקליים ולמצבים פיזיקליים יש תכונות שאינן תכונות פיזיקליות במובן זה – תכונות הניתנות לריבוי מימושים (היזכרו בדוגמת העוגיות, בדוגמת משאבת הדם, ובדוגמת המדחום), אך שהתממשויותיהן נקבעות על ידי התממשויותיהן של תכונות פיזיקליות. אין כל קושי ברעיון שבין התכונות הללו תהייה תכונות פונקציונליות. אין כל בעיה באפשרות שמצבים פיזיקליים ימלאו פונקציות שונות, ואנו אף מכירים תכונות פונקציונליות רבות של מצבים פיזיקליים. למעשה, כמעט לכל מכשיר יש מצבים פיזיקליים הממלאים פונקציה סיבתית זו או אחרת, הנקבעת על ידי התכונות הפיזיקליות שלו ומצביו. חשבו על מכונה למכירת משקאות, מכונת כביסה, מנוע של מכונית, וכדומה. למכשירים אלו, שהם לחלוטין חלק מן העולם הפיזיקלי, יש מצבים פנימיים בעלי פונקציות או תפקידים סיבתיים שונים. כך, אין קושי ברעיון שלמצבים פיזיקליים טהורים יש תכונות פונקציונליות. במילים אחרות, אין קושי ברעיון שהתממשויותיהן של תכונות פונקציונליות נקבעות על ידי התממשויותיהן של תכונות פיזיקליות. הפונקציונליזם מתיישב אפוא עם אונטולוגיה פיזיקליסטית: אונטולוגיה פיזיקליסטית מכירה בקיומן של תכונות שאינן פיזיקליות במובן הנדון (תכונות הניתנות לריבוי מימושים פיזיקלי), ובכללן תכונות פונקציונליות. אם כך, הרי מי שמקבלים את ההשקפה הפונקציונליסטית באשר ליסוד הנפשי – מי שמקבלים שתכונות נפשיות הן תכונות פונקציונליות – יכולים לקבל אונטולוגיה פיזיקליסטית באשר ליסוד הנפשי, דהיינו פיזיקליזם-פרט.

למעשה, כדי שניתן יהיה לעבור באופן לגיטימי מן הטענה שאין קושי ברעיון על פיו למצבים פיזיקליים יש תכונות פונקציונליות, לטענה שההשקפה הפונקציונליסטית – **כהשקפה בדבר המצבים הנפשיים והתכונות הנפשיות** – מתיישבת עם אונטולוגיה פיזיקליסטית, עלינו להסיר עוד מכשול. ניתן לטעון שאף אם תכונות פונקציונליות עשויות להיות תכונותיהם של

<sup>19</sup> תזכורת: ביחידה 4-5 אמרנו כי הטענה שתכונות נפשיות הן תכונות הניתנות לריבוי מימושים משמעה שתכונות נפשיות אינן תכונות פיזיקליות, אך הוספנו כי יש המבקרים לומר שמשמעה הוא כי תכונות נפשיות אינן תכונות פיזיקליות מסדר ראשון – הן **תכונות פיזיקליות מסדר גבוה**. הבחירה אלו מבין שתי אמירות אלו לאמצץ היא בחירה מינוחית בלבד – אין כאן מחלוקת פילוסופית.

מצבים חומריים, הרי שהתכונות הפונקציונליות אשר מאפיינות את מצבינו הנפשיים אינן כאלו; הן אינן יכולות להיות תכונותיהם של מצבים חומריים, משום שהן פונקציות מורכבות וסבוכות מאוד.

ביחידה 2 הוצג טיעון אשר ביסודו עמד רעיון דומה, אף על פי שטיעון זה לא השתמש במונחים של פונקציות. הכוונה לטיעון של דקארט, שעל פיו לא ניתן להסביר את ההתנהגויות הלשוניות שלנו ואת הביצועים המתמטיים שלנו על-סמך ההנחה שסיבותיהם הן סיבות פיזיקליות. נוכחנו באופנים אחדים כי טיעון זה אינו עומד בפני הביקורת. נזכיר את אחד האופנים הללו: ראינו כי הדרך האפשרית היחידה לבסס את הטענה שלא ניתן להסביר ביצועים כלשהם על-סמך ההנחה שסיבותיהם הן סיבות פיזיקליות היא להראות שמערכת פיזיקלית גרידא אינה יכולה להיות בעלת רמת מורכבות מספקת כדי להוליך לביצועים כאלו; או לפחות, שהמערכת הפיזיקלית הרלוונטית – במקרה של ביצועים אנושיים המדובר במוח – אינה בעלת רמת מורכבות מספקת. אולם אין כל טעם סביר להחזיק בהנחה שמערכת פיזיקלית אינה יכולה להגיע לרמת המורכבות הדרושה כדי שתוכל להיות את המנגנון הסיבתי העומד ביסוד פעילות אנושית כלשהי, ובכלל זה פעילויות אינטלקטואליות מסובכות כמו פעילות מתמטית או פעילות לשונית; ובפרט, אנו יודעים שהמוח האנושי הוא בעל רמת מורכבות עצומה, כדי כך שאין טעם סביר להניח שמורכבות זו אינה מספקת כדי לאפשר לו להיות מנגנון סיבתי כזה. פירוש הדבר, שאין טעם סביר להתנגד לטענה שמצבים פיזיקליים, כמו מצבי המוח שלנו, עשויים לממש פונקציות מורכבות מאוד. במילים אחרות, אין טעם סביר להתנגד לטענה שתכונות פונקציונליות מורכבות מאוד – ובפרט, התכונות הפונקציונליות המאפיינות את מצבינו הנפשיים – עשויות להיות תכונותיהן של מצבים פיזיקליים דוגמת מצבי המוח. ומשמעות האמירה שהתכונות הפונקציונליות המאפיינות את מצבינו הנפשיים עשויות להיות תכונותיהן של מצבי המוח שלנו היא שאין בעיה להניח שהפונקציונליזם מתיישב עם אונטולוגיה פיזיקליסטית לא רק באופן תיאורטי אלא באופן ממשי.

**ההשקפה הפונקציונליסטית, שעל פיה תכונות נפשיות הן תכונות פונקציונליות ואינן תכונות פיזיקליות, מתיישבת עם הטענה הפיזיקליסטית (הטענה של פיזיקליזם-פרט) שעל פיה מצבים נפשיים הם מצבים פיזיקליים (ובפרט, מצבים מוחיים).**

פונקציונליזם, אם כן, שקודם ראינו כי הוא מתנגש בפיזיקליזם-סוג, מתיישב עם פיזיקליזם-פרט. אולם מובן כי אין בכך כדי לומר שפונקציונליזם מחייב פיזיקליזם-פרט. העובדה שהתכונות הפונקציונליות המאפיינות את מצבינו הנפשיים עשויות להיות תכונות של מצבים פיזיקליים כמצבי-מוח, אין משמעה שהתכונות הללו הן אמנם תכונות של מצבים פיזיקליים כמצבי-מוח. כפי שציין הילארי פטנאם, שרבים רואים בו את אבי התורה הפונקציונליסטית, **הרעיון הפונקציונליסטי מתיישב אף עם דואליזם!**<sup>20</sup>

<sup>20</sup> H. Putnam, "The Nature of Mental States", in his *Mind, Language and Reality: Philosophical Papers*, Volume 2, Cambridge: Cambridge University Press, 1975, pp. 429-440.

דרך אחת להיווכח בכך שהרעיון הפונקציונליסטי מתיישב עם הדואליזם היא זו: בסיסי להשקפה הפונקציונליסטית הוא רעיון ריבוי המימושים של מצבים נפשיים ותכונות נפשיות, על פיו מצבים נפשיים זה-סוג עשויים להיות מצבים בעלי מבנים פיזיקליים שונים מאוד אלו מאלו, ובאופן אחר, תכונות נפשיות עשויות להיות ממומשות על ידי מצבים שונים מאוד אלו מאלו מבחינה פיזיקלית. אמנם האמנתי שיורד גשם היא מצב הממומש במוחי, שהוא אובייקט פיזיקלי הבנוי על-יסוד פחמן, אבל האמנה כזו עשויה להיות ממומשת גם במוח עשוי סיליקון, במוח עשוי משבבים אלקטרוניים, ואולי אף (כפי שפונקציונליסטים אחדים אוהבים לומר) במוח העשוי מגבינה צהובה (אם רק מסוגל "מוח" כזה להיות בעל דרגת המורכבות הדרושה).

רעיון זה משתלב היטב עם ההשקפה הפונקציונליסטית, משום ש**פונקציות** ניתנות לריבוי מימושים פיזיקלי – אותה פונקציה עשויה להיות ממומשת על ידי מצבים פיזיקליים השונים מאוד זה מזה. עתה יכולים הפונקציונליסטים לצעוד צעד נוסף: אם תכונות נפשיות עשויות, באופן עקרוני, להיות ממומשות על ידי מגוון רחב של מצבים פיזיקליים, כשהסייג היחיד לכך שמצבים פיזיקליים מסוימים יממשו תכונות מסוימות הוא היותם בעלי רמת מורכבות מספקת – אם, במילים אחרות, הטבע החומרי איננו חשוב למימוש התכונות הנפשיות – הרי **ייתכן, באופן עקרוני, שגם מצבים שאינם חומריים יממשו את התכונות הנפשיות**. באופן ספציפי, ייתכן שגם מצבים **רוחניים** יממשו תכונות אלו. בהתאם לרעיון הפונקציונליסטי, מערכות שיכולות לממש את הפונקציות הנפשיות הן מערכות היכולות לממש את התכונות הנפשיות (שהרי, לפי הפונקציונליסטים, התכונות הנפשיות **זהות** לפונקציות הללו); ולכן, כיוון שהפונקציות הנפשיות ניתנות למימוש על ידי מגוון רחב של מערכות, כך גם התכונות הנפשיות: מערכת נפשית, מערכת בעלת תכונות נפשיות, עשויה (באופן עקרוני, בכפוף לסייג האמור) להיות ממומשת במערכת הבנויה על בסיס פחמן, על בסיס סיליקון, במערכת העשויה מגבינה צהובה, ובמערכת שאינה חומרית כלל. אנו יכולים אפוא לומר, שמבחינה מסוימת (ובהמשך נבהיר את הגיונו של הסיוג "מבחינה מסוימת"), **הרעיון הפונקציונליסטי אדיש לשאלה האונטולוגית – הוא ניטרלי באשר לשאלה זו**. מהותם של מצבים נפשיים אינה תלויה בשאלת מימושם – אין היא תלויה בטבע האונטולוגי של המערכת המממשת אותם.

ובכן,

**הרעיון הפונקציונליסטי מתיישב הן עם מטריאליזם נפשי (ובאופן ספציפי, עם פיזיקליזם-פרט) והן עם דואליזם.**

אולם שימו לב שאנו כותבים כי **הרעיון הפונקציונליסטי** מתיישב עם שתי ההשקפות הללו, ואדיש לשאלה האונטולוגית. מרבית **הפילוסופים** הדוגלים בפונקציונליזם אינם אדישים

ההתייחסות להתיישובותם של פונקציונליזם ודואליזם נמצאת בעמוד 436.  
פילוסוף המאמץ גרסה מסוימת של פונקציונליזם דואליסטי הוא דייוויד צ'למרס. ראו D. Chalmers, *The Conscious Mind: A Search of Fundamental Theory*, Oxford: Oxford University Press, 1996.

לשאלה זו – הם דוגלים באונטולוגיה פיזיקליסטית; על פי רוב הם מקבלים את העמדה של **פיזיקליזם-פרט בלעדי**. דהיינו, הם סוברים שמצבים נפשיים הם מצבים פיזיקליים (ובאופן ספציפי, באשר לבני אדם ויצורים רבים אחרים, הם מצבי-מוח), ותכונותיהם הנפשיות של המצבים הללו הן תכונות פונקציונליות שלהם. כלומר, תכונות נפשיות, על פי השקפה זו, הן תכונות פונקציונליות של מצבים פיזיקליים, ולגבי בני-אדם ויצורים רבים אחרים, תכונות נפשיות הן תכונות פונקציונליות של **מצבי-מוח**.

ניתן לראות את הקרבה בין השקפתם של הפונקציונליסטים הללו – שאפשר לכוונתם "פונקציונליסטים פיזיקליסטיים" – לבין ההשקפה של פיזיקליזם התנהגותני; שתיהן השקפות פיזיקליסטיות: על פיהן מצבים נפשיים הם מצבים פיזיקליים (וספציפית, מצבי-מוח).

### שאלה

מה ההבדל בין שתי ההשקפות הללו?

ההבדל בין ההשקפות הללו הוא הבדל בין טיבן של הפונקציות הנפשיות על פיהן: לפי פיזיקליזם התנהגותי אלו הן פונקציות גירוי-תגובה פשוטות; לפי "הפונקציונליזם הפיזיקליסטי" אלו הן פונקציות מורכבות יותר, הכרוכות לא רק בקשרים בין המצב הנפשי לבין גירויים תפיסתיים ותגובות התנהגותיות, אלא גם בינו לבין מצבים נפשיים אחרים. כאמור, ההשקפה הפונקציונליסטית ניטרלית מבחינה אונטולוגית. עם זאת, אם אמנם הרעיון הפונקציונליסטי נכון, אם התכונות הנפשיות הן תכונות פונקציונליות, אם ניתן להסביר את טיבם של המצבים הנפשיים באמצעות הסבר פונקציונליסטי, הריי **הגרעין הקשה של הבעיה הפסיכופיזית מתפרק מנשקו**. זאת משום שמשמעות הטענה שאפשר להסביר את טיבם של המצבים באמצעות הסבר פונקציונליסטי, היא שאין צורך להתייחס לקטגוריות אונטולוגיות **בהסבר היסוד הנפשי**. די להתייחס לקטגוריות פונקציונליות, שהן, כפי שראינו, ניטרליות מבחינה אונטולוגית (דהיינו, מבחינה עקרונית הן ניתנות הן למימוש חומרי והן למימוש רוחני). לרעיון הפונקציונליסטי יש, לפיכך, השלכה חשובה באשר לשאלת האונטולוגיה של המצבים הנפשיים: אם אמנם אין צורך להתייחס לקטגוריות אונטולוגיות בהסבר היסוד הנפשי, הרי שהיסוד הנפשי איננו מהווה סיבה לסטות מן האונטולוגיה המטריאליסטית. הרעיון הפונקציונליסטי מאפשר להישאר בגבולותיה של אונטולוגיה מטריאליסטית, ובקרב פילוסופים רווחת העמדה שאם **ניתן** להחזיק באונטולוגיה מטריאליסטית, הרי שיש להחזיק בה.

אמנם הרעיון הפונקציונליסטי מתיישב, כאמור, גם עם דואליזם – קבלתו אינה מחייבת קבלתה של אונטולוגיה מטריאליסטית (ובאופן יותר ספציפי, קבלתה של אונטולוגיה פיזיקליסטית של היסוד הנפשי). אולם פילוסופים רבים סוברים שאין סימטריה בין הדואליזם לבין המטריאליזם בהקשר הנדון כאן. לדעתם, אם רק אפשר, עדיף להימנע

מדואליזם, ובהיעדר טיעונים מוחצים בזכות ההשקפות השונות בסוגיה הפסיכופיזית, ההשקפה המטריאליסטית היא ברירת המחדל ויש לקבלה. יש המנמקים העדפה זו של המטריאליזם בטענה שההשקפה הדואליסטית היא ביטוי של מצוקה אינטלקטואלית יותר מאשר פתרון של הבעיה הפסיכופיזית. יש המקבלים אותה בשל טעמי פשטות, ויש הסוברים (בדומה) שנדרשים טעמים חזקים כדי לקבל את הטענה שהיסוד הנפשי מהווה חריג ביקום. קבלת הרעיון הפונקציונליסטי אמנם מאפשרת לשמור על אונטולוגיה מטריאליסטית, גם למי שמשתכנעים מטעם זה או אחר שתכונות נפשיות אינן תכונות פיזיקליות במובן האמור. למעשה, אחד המניעים החשובים ביסוד הצעתו של הקריטריון הפונקציונליסטי כקריטריון לזהותם של מצבים נפשיים ותכונות נפשיות היה בדיוק זה: להציע קריטריון אשר יאפשר למי שהשתכנעו מטעם זה או אחר שתכונות נפשיות אינן תכונות פיזיקליות לשמור על אונטולוגיה מטריאליסטית (ובאופן יותר ספציפי, על אונטולוגיה פיזיקליסטית של היסוד הנפשי). ראינו כבר ביחידה הקודמת כי למצבים פיזיקליים יש תכונות שאינן פיזיקליות (אף על פי שהן מתיישבות עם אונטולוגיה מטריאליסטית); הפונקציונליסטים טוענים כי **תכונות נפשיות** הן כאלו – הן תכונות פונקציונליות.

נסכם אפוא בקצרה:

הרעיון הפונקציונליסטי מתיישב הן עם פיזיקליזם והן עם דואליזם, אולם מרבית הפונקציונליסטים מאמצים גרסה של פיזיקליזם – פיזיקליזם-פרט בלעדי: על פיהם מצבים נפשיים הם מצבים פיזיקליים, ובאשר לבני-אדם ויצורים אחרים, מצבים נפשיים הם מצבי-מוח, ותכונות נפשיות הן תכונות פונקציונליות של מצבי-מוח.

מעתה, בדברנו על ההשקפה הפונקציונליסטית, נתייחס להשקפה שעל פיה (במקרה של בני-אדם ויצורים אחרים) מצבים נפשיים הם מצבי-מוח, ותכונות נפשיות הן תכונות פונקציונליות של **מצבי-מוח**. כלומר, את המונחים "פונקציונליזם" ו"ההשקפה הפונקציונליסטית" נבין כמתייחסים להשקפה שכינינו "פונקציונליזם פיזיקליסטי" (אלא אם כן יצוין ששימוש ספציפי של מונח מן המונחים הללו סוטה משימוש זה).

ראוי להעיר כי יש המציגים את ההשקפה הפונקציונליסטית כהשקפה אנטי-פיזיקליסטית **השוללת** את הרעיון שמצבים נפשיים הם מצבים פיזיקליים. לפי גישה זו, השקפה המקבלת ריבוי מימושים פיזיקלי של מצבים נפשיים, ויותר מכך, המתירה באופן עקרוני אף את האפשרות של מימוש בלתי-פיזיקלי של מצבים נפשיים, אינה מאפשרת לומר שמצבים נפשיים הם מצבים פיזיקליים. כאבים, לדוגמה, אינם מצבים פיזיקליים לפי הפונקציונליזם, משום שאין כל מבנה פיזיקלי המהותי להיותם כאבים (או להיותם מצבים נפשיים) – כאבים אינם ניתנים לזיהוי עם מצבים פיזיקליים כלשהם, אלא רק עם מצבים פונקציונליים. ומצבים פונקציונליים **אינם זהים** למצבים פיזיקליים: מצבים פונקציונליים **ממומשים על ידי** מצבים

פיזיקליים (הם ממומשים על ידי אותם מצבים פיזיקליים שממלאים את הפונקציות הרלוונטיות).<sup>21</sup>

בהציגה את ההשקפה הפונקציונליסטית כהשקפה אנטי-פיזיקליסטית, אין גישה זו מתכוונת להציגה כהשקפה המחויבת לדואליזם. כלל וכלל לא. על פיה, ההשקפה הפונקציונליסטית מתנגשת גם עם הרעיון שמצבים נפשיים זהים למצבים רוחניים. הרעיון שביסוד גישה זו הוא שלפי ההשקפה הפונקציונליסטית מצבים נפשיים אינם יכולים להיות **זהים למצבים ממשיים כלשהם**; מצבים ממשיים כלשהם – בין אם פיזיקליים ובין אם רוחניים – אינם יכולים אלא **לממש** מצבים נפשיים (אם הם יכולים למלא את הפונקציות הרלוונטיות). השקפה זו מציגה את הפונקציונליזם כחלופה הן לפיזיקליזם והן לדואליזם – היא מציגה את הפונקציונליזם כהשקפה מסוג אחר לחלוטין.

כמובן, הצגה זו של ההשקפה הפונקציונליסטית מנוגדת לאופן שבו הצגנו אותה אנו – כהשקפה המתיישבת עם פיזיקליזם (כמו גם עם דואליזם), ואשר רוב מצדדיה סוברים שהגרסה הפיזיקליסטית שלה (הגרסה שכינינו "פונקציונליזם פיזיקליסטי") היא ההשקפה האמיתית בדבר טבעו של היסוד הנפשי. אנו איננו סוברים שהעובדה שאין כל מבנה פיזיקלי המהותי להיותו של מצב נפשי המצב הנפשי שהינו (רעב, מחשבה על שכונת מונטיפיורי, וכדומה), או לעצם היותו מצב נפשי, מחייבת שאין לזהות מצבים נפשיים עם מצבים פיזיקליים. הטענה שמה שעושה מצב נפשי למצב הנפשי שהינו, ומה שעושה אותו למצב נפשי בכלל, אינו מאפיין פיזיקלי, מתיישבת עם הטענה שבפועל מצבים נפשיים הם מצבים פיזיקליים. אי-מהותיותו של המבנה הפיזיקלי לנפשיות ולזהות נפשית ספציפית מתיישב עם הטענה שהמצבים המממשים בפועל את הפונקציות הנפשיות הם מצבים פיזיקליים. פונקציונליסטים בהחלט יכולים לומר שהמצבים הללו הם המצבים הנפשיים. במלים אחרות, הטענה של דוגלי הגישה האמורה לפונקציונליזם, שלפיה מצבים פיזיקליים עשויים לממש מצבים נפשיים אך לא להיות זהים עמם, אינה נראית מוצדקת: **התכונות** הנפשיות אינן זהות לפי הפונקציונליזם לתכונות פיזיקליות (הן תכונות פונקציונליות), אולם המצבים המממשים תכונות אלו (מצבים פיזיקליים, על פי מרבית הפונקציונליסטים) הם המצבים הנפשיים.<sup>22</sup>

גישתנו כלפי ההשקפה הפונקציונליסטית מדגישה את היותה השקפה השונה באורח ניכר הן מההשקפה הדואליסטית והן מגרסותיה של ההשקפה הפיזיקליסטית שרווחו בתקופה שלפני פיתוחה של ההשקפה הפונקציונליסטית; והיבט חשוב של השוני בין ההשקפה הפונקציונליסטית לבין ההשקפות הללו הוא שעל פי הפונקציונליזם טיבו האונטולוגי של

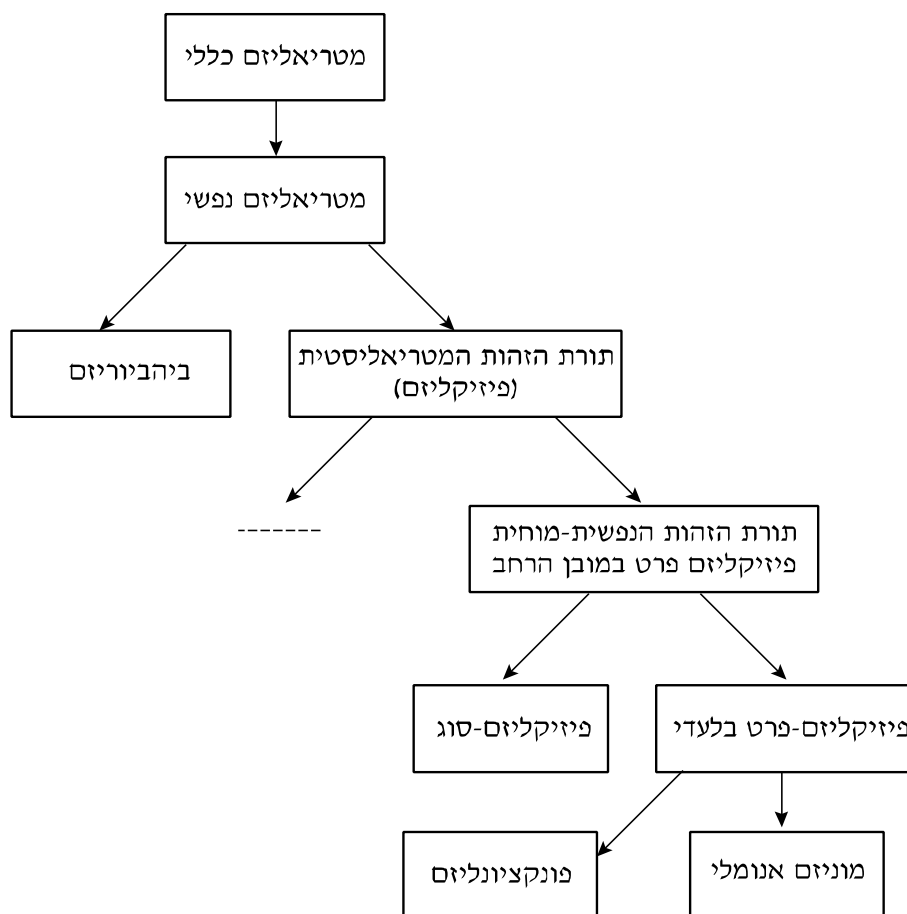
<sup>21</sup> לעניין זה ראו, למשל:

J. Heil, *Philosophy of Mind: A Contemporary Introduction*, London: Routledge, 2004, pp. 90-91.

<sup>22</sup> הזכרנו בהערה 17 לעיל את טענתו של פטנאם כי מצבים נפשיים אינם מצב-מוח אלא מצבים פונקציונליים, אולם, כפי שאמרנו, נראה שפטנאם התכוון להביע את הטענה שעל פיה התכונות המקנות למצבים נפשיים את אופייהם הנפשיים – תכונות נפשיות – הן תכונות פונקציונליות ואינן תכונות פיזיקליות.

היסוד הנפשי – ממה הוא "עשוי" – אינו מהותי (אין הוא מהותי לעצם הנפשיות ואין הוא מהותי לזהות נפשית ספציפית). כך עומדים אפוא על הבט חשוב של הפונקציונליזם. אולם כאמור, עמידה על הבט זה אינה מחייבת את המסקנה כי הפונקציונליזם משמעו אנטי-פיזיקליזם, במובן זה שמצבים נפשיים אינם יכולים להיות זהים עם מצבים פיזיקליים. הפונקציונליזם מתיישב עם פיזיקליזם, וכאמור מרבית הפונקציונליסטים דוגלים בכך שבפועל מצבים נפשיים הם מצבים פיזיקליים. אף ראינו כי יש פילוסופים הסוברים שהפונקציונליזם עשוי דווקא לבסס את הפיזיקליזם.

עתה נוכל לעדכן את תרשימים העמדות המטריאליסטיות: נוכל לציין בו את ההשקפה הפונקציונליסטית כגרסה של פיזיקליזם-פרט בלעדי.



**תרשים 6.** העמדות המטריאליסטיות (תרשים מעודכן)

חשוב להבהיר: הפונקציונליסטים המקבלים כי מצבים נפשיים הם מצבי-מוח, יטענו כי **המוח אינו מהותי ליסוד הנפשי**. הוא אמנם מממש את התכונות הנפשיות, אולם גם מערכות פיזיקליות אחרות עשויות, באופן עקרוני, לממש תכונות אלו: כל שנדרש הוא שהמערכת

תממש את הפונקציות הנפשיות (שהרי תכונות נפשיות הן, לפי הפונקציונליסטים, תכונות פונקציונליות), ומערכות פיזיקליות אחרות (וכאמור, אף מערכות שאינן פיזיקליות, אם יש כאלו) עשויות לממש פונקציות אלו. המוח אינו מהותי לזהות הנפשית של יצור נפשי. המוח חשוב לזהותו הנפשית של יצור נפשי רק משום שהוא **מממש את הפונקציות** המכוננות את חייו הנפשיים של היצור. כך, **כפיל נפשי** שלי – מי שמצביו הנפשיים זהים בסוגם למצבים הנפשיים שלי – אינו חייב, לפי הפונקציונליזם, להיות **כפיל פיזיקלי** שלי; אפילו מוחו אינו חייב להיות זהה למוחי מבחינה פיזיקלית (המערכת שממשת את מצביו הנפשיים עשויה להיות שונה משלי באופן כזה שכלל לא נתייחס אליה כאל מוח); למעלה מכך, מי שחולק עמי תכונה נפשית כלשהי (או תכונות נפשיות כלשהן) אינו חייב, לפי השקפה זו, לחלוק עמי באופן ייחודי תכונה פיזיקלית כלשהי. זאת משום שזהות פיזיקלית, ואפילו חלקית ביותר, איננה הכרחית לזהות פונקציונלית כלשהי.

נזכיר שתי השלכות של דברים אלו בדבר אי המהותיות של המוח לנפש, שהראשונה שבהן משמעותית במיוחד. ראשית, באופן מפתיע, אנו רואים שקיים **מכנה משותף בין ההשקפה הפונקציונליסטית לבין ההשקפה הדואליסטית**: הרעיון שהמוח (או כל מערכת חומרית ספציפית שהיא) אינו מהותי לנפש הוא רעיון מרכזי של הדואליזם. הפונקציונליזם מראה שגם מטריאליסטים יכולים לקבל רעיון זה!

---

♦ מה ההבדל בין הטענה שהנפש אינה מהותית למוח כשהיא נטענת בהקשר דואליסטי לבין טענה זו כשהיא נטענת בהקשר פונקציונליסטי?

---

את ההשלכה השנייה נציג באמצעות שאלה.

### שאלה

הדברים האמורים בדבר אי-מהותיות המוח לנפש רלוונטיים לשאלת הקשר בין פונקציונליזם לבין **העיקרון ההופכי לעקרון הנסמכות** שבו דנו ביחידה 4-5. מה אפשר ללמוד מהאמור כאן על קשר זה?

כזכור, לפי העיקרון ההופכי לעקרון הנסמכות, זהות בתכונות נפשיות כלשהן מחייבת זהות בתכונות פיזיקליות מסוימות. אנו רואים שההשקפה הפונקציונליסטית איננה מתיישבת עם עיקרון זה. אין היא פוסלת את האפשרות של זהות בתכונות נפשיות ללא זהות בתכונות פיזיקליות כלשהן. בנקודה זאת מתנגשת ההשקפה הפונקציונליסטית בהשקפה של פיזיקליזם-סוג, המחייבת עיקרון זה.

אפשר להיווכח באי-התיישבותן של ההשקפה הפונקציונליסטית ושל העיקרון ההופכי לעקרון הנסמכות גם באופן הבא: ההשקפה הפונקציונליסטית מאפשרת ריבוי מימושים, ואילו, כפי



שראינו ביחידה הקודמת, העיקרון ההופכי לעקרון הנסמכות אינו מאפשר ריבוי מימושים (כזכור, עיקרון זה שולל את האפשרות של קיומם של חייזרים שמוחם עשוי מסיליקון ואשר יש להם מצבים נפשיים מן הסוגים שיש לנו, ככאבים, מחשבות, ורצונות).

אם כן, זהות פיזיקלית אינה הכרחית לזהות פונקציונלית, וכך אינה הכרחית לזהות נפשית לפי ההשקפה הפונקציונליסטית, אבל ההפך אינו נכון. דהיינו, זהות פיזיקלית **מספיקה** לזהות פונקציונלית, וכך היא מספיקה לזהות נפשית לפי ההשקפה הפונקציונליסטית. שתי מערכות פיזיקליות הזהות ביניהן לחלוטין מבחינה פיזיקלית – מערכות שלהן בדיוק אותן תכונות פיזיקליות – תהיינה זהות ביניהן לחלוטין מבחינה פונקציונלית: הזהות הפיזיקלית בין המערכות משמעה, בין היתר, שמתבצעים בהן אותם תהליכים סיבתיים. כך, לכל מצב ממצביה של המערכת האחת יש מצב מקביל במערכת השנייה – מקביל במובן זה שהוא כרוך באותה רשת של קשרים סיבתיים עם מצביה האחרים של המערכת. במילים אחרות, מצביהן של המערכות הללו מממשים את אותן פונקציות – ממלאים את אותם תפקידים סיבתיים. אם כן, זהות פיזיקלית בין מערכות מחייבת זהות פונקציונלית ביניהן, ופירוש הדבר הוא שלפי הפונקציונליזם הפיזיקליסטי, זהות פיזיקלית בין יצורים נפשיים מחייבת זהות נפשית ביניהם. אנו יכולים אפוא לומר שהפונקציונליזם הפיזיקליסטי מחייב את עקרון הנסמכות של הנפשי על הפיזיקלי.

הפונקציונליזם אינו מתיישב עם העיקרון ההופכי לעיקרון הנסמכות של הנפשי על הפיזיקלי.  
הפונקציונליזם הפיזיקליסטי מחייב את עקרון הנסמכות של הנפשי על הפיזיקלי.

### 3. פונקציונליזם ומודל המחשב של הנפש

אחת הדרכים לעמוד על טיבה של ההשקפה הפונקציונליסטית היא להציגה במונחי מודל המחשב של הנפש. הצעתו של מודל זה היא בעיני רבים מן התרומות החשובות ביותר של ההשקפה הפונקציונליסטית, ואחד מסודות קסמה העיקריים. מודל זה מבהיר היבטים של הפונקציונליזם שכבר הצגנו, וחושף היבטים נוספים, וכך מתקבלת תמונה שלמה יותר של השקפה זו.

נחשוב על המחשב הביתי המוכר לנו. נניח שבזמן מסוים הוא מבצע פעולה מסוימת: חישוב מתמטי מסובך, משחק שחמט, או כל פעולה אחרת מן הסוג שמחשבים מבצעים לעתים. כשהמחשב מבצע פעולה מעין זו, מתרחשים בו תהליכים פיזיקליים שונים (בעיקר תהליכים חשמליים). למעשה, אנו יכולים לתאר את כל פעולותיו של המחשב באמצעות תיאור התהליכים הפיזיקליים הללו. תיאור פיזיקלי כזה עשוי להיות תיאור מלא וממצה של המחשב, מצביו ותהליכיו, תיאור שאינו מחמיץ שום עובדה פנימית ("אינטרנסית"), על אודות המחשב.

אולם אנו יכולים לתאר את פעולת המחשב גם באופן אחר, על ידי תיאור שאינו כולל כל מונח פיזיקלי. כלומר, אנו יכולים לתאר את המחשב **במונחי התכנית** שעל פיה הוא פועל. למה הכוונה? נניח שהמשימה של המחשב היא לחשב את המכפלה  $4 \times 3$ . ניתן להעלות על הדעת תכניות שונות לביצוע חישוב זה. על פי תכנית אחת (נקרא לה "תכנית א"), המחשב יבצע את פעולת החיבור  $3+3+3+3$ . על פי תכנית אחרת ("תכנית ב"), המחשב יבצע תחילה את פעולת החיבור  $3+3$ , וכשיקבל את הסכום 6, יבצע את פעולת החיבור  $6+6$ . כשמחשב מבצע את תכנית א, הוא מבצע פעולות פיזיקליות מסוימות, אבל כשאנו אומרים שהוא מבצע את תכנית א – כשאנו אומרים שהוא מחשב את המכפלה  $4 \times 3$  ומפרטים את שלבי הביצוע של חישוב זה באופן שעשינו זאת זה עתה (המחשב מחבר את המספר 3 ארבע פעמים) – לא תיארו את פעולתו במונחים פיזיקליים אלא במונחי התכנית של פעולתו. בלשון אחרת, תיארו את **החישוב** שהמחשב מבצע. החישוב מבוצע על ידי המחשב באופן פיזיקלי, על ידי תהליכים פיזיקליים שונים, אבל המושג של **חישוב** כמו המושג של **תכנית** אינם מושגים פיזיקליים. באופן דומה, ניתן גם לתאר את המצב הספציפי שהמחשב נמצא בו במונחי התכנית – ניתן לומר באיזה שלב של התכנית נמצא המחשב (למשל, הוא בשלב של חיבור  $3+3$  בתכנית ב), ואמירה כזו אינה כרוכה בתיאור פיזיקלי של מצבו של המחשב.

אם כן, אנו יכולים לתאר את פעולת המחשב ואת מצביו הן במונחים פיזיקליים והן במונחי התכנית או התכניות שהוא מבצע (או, כפי שאומרים לעתים, התכנית או התכניות שהוא "מריץ"). תיאור רכיביו, מצביו ופעולותיו של המחשב במונחים פיזיקליים הוא תיאור היבטים שונים של **החומרה** של המחשב (באנגלית: hardware). תיאור מצביו ופעולותיו של

המחשב במונחי התכנית שהוא מריץ הוא תיאור היבטים שונים של **התוכנה** של המחשב (באנגלית: software).

אם כן,

**אפשר לתאר מחשבים ופעולותיהם הן ברמת החומרה והן ברמת התוכנה.**

לייחס למחשב חומרה מסוימת ותוכנה מסוימת פירושו לתאר אותו **ברמות תיאור שונות**. אין למחשב חומרה ותוכנה באותו מובן שיש לו מעבד וכרטיס זיכרון. איש אינו חושב שהמחשב נמצא במצבים פיזיקליים שונים **ובנוסף לכך** במצבי תוכנה שונים. אלא שהימצאותו במצב פיזיקלי מסוים בשעת פעולתו היא גם הימצאותו במצב תוכנה מסוים, למשל, בשלב מסוים של חישוב מתמטי. **ייחוס של מצבי חומרה ומצבי תוכנה למחשב אינו כרוך בתפיסה דואליסטית של מחשבים**. החומרה והתוכנה של מחשב הם מאפיינים שונים של אותה מערכת ומערכת זו היא מערכת פיזיקלית טהורה.

#### שאלה

אמרנו שהימצאותו של מחשב במצב פיזיקלי מסוים בשעת פעולתו היא גם הימצאותו במצב תוכנה מסוים. האם, לדעתכם, גם ההפך נכון, כלומר, האם הימצאותו של מחשב במצב תוכנה מסוים היא גם הימצאותו במצב חומרה מסוים, במצב פיזיקלי מסוים?

התשובה לשאלה זו שלילית. הימצאותו של מחשב במצב תוכנה מסוים (למשל, בשלב מסוים של חישוב מתמטי) אינה הימצאותו במצב פיזיקלי מסוים, משום שמחשבים שונים – מחשבים הנבדלים זה מזה מבחינה פיזיקלית – עשויים להריץ או לממש את אותה תוכנה. תוכנה מסוימת אינה מחייבת חומרה מסוימת, ולכן הימצאותו של מחשב במצב תוכנה מסוים אמנם ממומשת על ידי היותו במצב פיזיקלי מסוים (מצב חומרה מסוים), אבל היא אינה מחייבת את היותו במצב פיזיקלי זה. המחשבים המודרניים הראשונים היו שונים מאוד בהרכבם הפיזיקלי מהמחשבים המוכרים לנו כיום – המחשבים הראשונים היו בנויים בעיקר משפופרות ריק, ואילו המחשבים של היום בנויים משבבים אלקטרוניים; אולם באופן עקרוני ייתכנו מחשבי שפופרות ריק המסוגלים לבצע אותם סוגי חישובים שמבצעים מחשבי שבבים אלקטרוניים.

#### שאלה

האם תופעה זו – מימושן של תוכנות על ידי חומרות שונות זו מזו – מזכירה לכם תופעה אחרת כלשהי? במילים אחרות, באמצעות איזה מונח שהכרנו ניתן לתאר תופעה זו?

תופעה זו מזכירה את תופעת **ריבוי-המימושים** של המצבים הנפשיים והתכונות הנפשיות. אנו יכולים לומר **שתוכנות ניתנות לריבוי מימושים**. מי שהשתכנעו שקיים ריבוי מימושים של

מצבים נפשיים ותכונות נפשיות, יכולים לראות דמיון מעניין בין מאפייני תוכנה של מחשבים לבין מאפיינים נפשיים. קיומו של דמיון זה עורר אצל פילוסופים שונים מחשבות באשר לטיבו של היסוד הנפשי; אנו נעסוק עוד רבות בדמיון זה.

### **תוכנות ניתנות לריבוי מימושים: מחשבים שקיימים ביניהם הבדלים ניכרים ברמת החומרה – הבדלים פיזיקליים ניכרים – עשויים לממש תוכנות זהות.**

כפי שראינו ביחידה 4-5, יש החולקים על קיומה של תופעת ריבוי-המימושים באשר ליסוד הנפשי, אבל דומה שקשה לחלוק על קיומה של תופעת ריבוי-המימושים באשר לתוכנה. אנו מכירים מקרים של מחשבים השונים אלו מאלו במידה רבה מאוד מבחינה חומרית, ואשר מממשים אותן תוכנות. אנו גם מסוגלים להבין תופעה זו. מה מסביר את תופעת הניתנות של תוכנות לריבוי מימושים? מה מסביר את העובדה שאותן תכניות מחשב ואותם חישובים ניתנים להרצה על מחשבים שונים מאוד מבחינה פיזיקלית?

ההסבר לתופעה זו נעוץ בהיותם של מאפייני התוכנה **מאפיינים מופשטים**. הקטגוריות של **תוכנה, תכנית מחשב, חישוב, הימצאות במצב מסוים בתכנית**, וכדומה, אינן קטגוריות פיזיקליות אלא קטגוריות מופשטות. ניתן להבין אותן, באופן עקרוני, מבלי לדעת דבר על המערכות הפיזיקליות שעשויות לבצע חישובים ולהריץ תכניות מחשב. תיאור של תכנית מחשב מסוימת אינו תיאור פיזיקלי: הוא תיאור של המעברים בין מצבים שונים, כאשר המצבים הללו מאופיינים במונחים מתמטיים, לוגיים, וכדומה. המאפיינים הללו הם מאפיינים מופשטים באופן דומה לזה בו תכונות גאומטריות הן תכונות מופשטות: התכונה של היות משולש, התכונה של היות משולש שווה-שוקיים, התכונה של היות זווית ישרה, ותכונות נוספות מעין אלו, אינן תלויות כלל בצבע של האובייקטים שיש להן תכונות כאלו, בהרכבם הכימי, ובמיקומם המרחבי, למרות שלכל אובייקט שיש לו תכונות גאומטריות יש צבע, הרכב כימי, ומיקום מרחבי. **למימושים** של תכונות גאומטריות יש מאפיינים כאלו, אך לא לצורות הגאומטריות עצמן, ולכן הן ניתנות לריבוי מימושים; כלומר, צורה גאומטרית מסוימת עשויה לאפיין אובייקטים בעלי צבעים שונים ומשונים, בעלי הרכבים כימיים שונים ומשונים, וכדומה. כשצורה גאומטרית מסוימת היא המושא של הוכחה גאומטרית, אנו מפשיטים מהמימושים השונים שלה את המאפיינים הללו – אנו מתעלמים מהם. באופן דומה, כשאנו מתייחסים למחשב (שהוא אובייקט פיזיקלי) כאל אמצעי חישוב, כאל מערכת המריצה תכנית מסוימת, אנו מפשיטים אותו מכל מאפייניו החומריים – מכל המאפיינים הכרוכים ב**מימוש** החישוב והתכנית.<sup>23</sup> המושגים של **תכנית, חישוב**, ודומיהם אינם מושגים הכרוכים במימוש – הם, כאמור, מושגים מופשטים. בתור שכאלה – בהיותם מושגים מופשטים האדישים לשאלת המימוש הפיזיקלי הספציפי – הם ניתנים לריבוי מימושים, למימוש על ידי אובייקטים פיזיקליים שונים. להלן נסיף ונבהיר, באמצעות דוגמה ששיחקה

<sup>23</sup> אנלוגיה זו שבין מאפייני התוכנה לבין תכונות גאומטריות לקוחה מג'ון הייל. ראו J. Heil, *Philosophy of Mind: a Contemporary Introduction*. London: Routledge, 2004, pp. 90-91.

תפקיד חשוב בהתפתחות ההשקפה הפונקציונליסטית, את משמעות היותם של מושגי התוכנה מושגים מופשטים.

**שימו לב:** אמנם התוכנה מגיע למחשבים שלנו בצורה מוחשית – על פי רוב על-גבי תקליטורים – אולם אין לזהות את התוכנה – המופשטת – עם האמצעי ("המדיום") הפיזיקלי הנושא אותה. עובדת היותה של התוכנה מופשטת מתיישבת כמובן עם העובדה שהיא ממומשת באופן פיזיקלי, על ידי אובייקטים פיזיקליים, וכדי שאובייקט פיזיקלי כמו מחשב יממש תוכנה כלשהי ודאי שיש להשפיע עליו **באופן פיזיקלי** (למשל, על ידי הכנסת התקליטור המתאים).

לאור מה שאמרנו עד עתה באשר למחשב, חומרה ותוכנה, וריבוי מימושים, כבר ניתן לשער מה הקשר בין דיוננו הנוכחי לבין פילוסופיה של הנפש. אחת הדרכים החשובות להצגת ההשקפה הפונקציונליסטית היא להתוות אנלוגיה בין יצורים נפשיים לבין מחשבים. הרעיון שהמוח הוא מחשב נקרה כמעט לכל מי ששמע על מחשבים – אין ספק שבמובן חשוב המוח הוא (בין היתר) מכונת חישוב, מכונה המבצעת חישובים שונים, תהליכי היסק, וכדומה.<sup>24</sup> אולם האנלוגיה שמתווים הפונקציונליסטים בין יצורים נפשיים לבין מחשבים היא ספציפית הרבה יותר. ראינו כי מחשבים ופעולותיהם ניתנים לתיאור בשתי רמות – רמת החומרה, ורמה מופשטת יותר – רמת התוכנה. על פי הפונקציונליסטים, גם יצורים נפשיים ניתנים לתיאור בשתי רמות כאלו. אנו, ובאופן ספציפי – המוח שלנו, ניתנים כמובן לתיאור פיזיקלי, אולם ניתן גם לתאר אותנו ברמה אחרת, מופשטת יותר. תיאור נפשי אינו אלא תיאור מופשט יותר של יצור נפשי – הוא תיאור התוכנה שלו. **הנפש היא התוכנה של הגוף, ובמובן חשוב (במקרה של בני-אדם ויצורים נפשיים אחרים) הנפש היא התוכנה של המוח.** בהקשר זה נטבעה הסיסמה הפונקציונליסטית הידועה כי:

**היחס בין הנפש למוח הוא כמו היחס בין תוכנת מחשב לבין החומרה שלו.**<sup>25</sup>

שימו לב, שססמה פונקציונליסטית ידועה זו מנוסחת במונחי **היחסים בין הנפש למוח**. כך, היא מבטאת את העובדה שהרעיון הפונקציונליסטי מהווה דרך להתמודד עם **בעיית הגוף והנפש**. לפי גישה זו, מאפיינים פיזיקליים ומאפיינים נפשיים שונים אלו מאלו שוני ניכר, אולם השוני אינו מתבטא בכך שהגוף והנפש הם שני עצמים נפרדים ושונים, או בכך שמצבים

<sup>24</sup> אנו אומרים שאין ספק שהמוח הוא, **בין היתר**, מכונת חישוב – מדוע "בין היתר"? משום שכמובן, מצבי המוח קשורים לא רק בפעולות חישוביות (במובן המקובל של "חישוביות") כמו חישובים מתמטיים, היסקים לוגיים, וכדומה. המוח קשור גם בתפיסות, רגשות, כאבים, ועוד.

<sup>25</sup> "The mind is to the brain as the program is to the hardware". ככל הידוע לנו, מי שניסח את הרעיון הפונקציונליסטי במילים אלו הוא הפסיכולוג הקוגניטיבי פיליפ ג'ונסון-לרד. ראו P.N. Johnson-Laird, *The Computer and the Mind*. Cambridge (Mass.): Harvard University Press, 1988.

פיזיקליים ומצבים נפשיים הם מצבים משני סוגים שונים (העשויים להיות קשורים אלו עם אלו בקשרים סיבתיים). כשם שהמחשב אינו נמצא במצבים פיזיקליים מסוימים **ובנוסף לכך** במצבי תוכנה מסוימים, כך יצור נפשי אינו נמצא במצבים פיזיקליים שונים **ובנוסף לכך** במצבים נפשיים. **מאפיינים פיזיקליים ומאפיינים נפשיים של יצור נפשי הם מאפיינים של אותה מערכת של מצבים**; זאת, בדיוק כשם שמאפייני חומרה ומאפייני תוכנה של מחשב הם מאפיינים של אותה מערכת של מצבים. ההבדל היסודי בין המאפיינים הנפשיים לבין המאפיינים הפיזיקליים הוא שהמאפיינים הנפשיים הם מאפיינים השייכים לקטגוריה מופשטת יותר. כך מאפשר מודל המחשב לתת פשר חדש ליחסי הנפש והמוח.

הסיסמה הפונקציונליסטית שהצגנו לעיל מבטאת היבט חשוב מאוד של מודל המחשב של הנפש, אולם אין היא מבטאת את הרעיון שביסוד מודל זה במלוא עוצמתו. הפונקציונליסטים שפנו למודל המחשב של הנפש התכוונו לטעון לא רק שהיחס בין הנפש למוח הוא כמו היחס בין תוכנת מחשב לבין החומרה שלו; על פיהם, היחס בין הנפש למוח הוא **כמו** היחס בין תוכנת מחשב לבין החומרה שלו משום **שהנפש היא אמנם תוכנה**. הנפש היא תוכנה, שבמקרה האנושי (ובמקרים נוספים) מממש אותה המוח.

בהקשר למודל המחשב של הנפש, יש המעלים את השאלה **מיהו/י המתכנת/ת** של תוכנת הנפש. מי תכנת/ה את המוח האנושי, למשל? עלינו להדגיש, שהטענה שהנפש היא תוכנה אין משמעה שהיא תוכנתה על ידי מתכנת/ת כלשהו/י. טענה זו מאפיינת את היסוד הנפשי כיסוד מופשט מסוג מסוים (דהיינו תוכנה) של אובייקטים פיזיקליים מסוימים, כמו המוח. הטענה שלאובייקט מסוים יש מאפיין כזה – שיש לו תוכנה – היא טענה על תכונות (הוויות) של האובייקט ולא על המקור ההיסטורי שלו ושל תכונותיו אלו. קבלת ההשקפה הפונקציונליסטית איננה מפחיתה כהוא זה מכוחו של הרעיון כי המוח והמאפיינים הנפשיים עוצבו על ידי תהליכי האבולוציה.

את הטענה שהנפש היא תוכנה נוסף ונבהיר לאורכה של יחידה זו. עתה נחדד את הרלוונטיות שלה לבעיית הגוף והנפש. אם נפתח מארז של מחשב בשעת פעולתו, ונתבונן בתהליכיו השונים – ולו באמצעות המכשור המשוכלל ביותר – לא נראה את תוכנתו, לא נראה כל מאפיין של התוכנה. אך, כמובן, יהיה זה כשל לוגי להסיק מעובדה זו שאין למחשב ולפעולתו מאפייני תוכנה. אנו יודעים היטב שמאפיינים אלו קיימים.

## שאלה

הטענה בדבר כשל לוגי זה עשויה להיות רלוונטית לבעיית הגוף והנפש, ובמיוחד לטיעון דואליסטי מסוים אותו פגשנו בשלב מוקדם. התוכלו להעלות בדעתכם/ באיזה טיעון מדובר? וכיצד רלוונטית הטענה הנדונה לבעיית הגוף והנפש?

היזכרו בטיעון הדואליסטי של לייבניץ, שבו דנו ביחידה 4-5. טיעון זה הניח כי בהתבוננו במכונה בעת פעולתה, לא נוכל לראות בה מצבים נפשיים (ובאופן ספציפי, התנסויות). מסקנת הטיעון הייתה כי המצבים הנפשיים אינם יכולים להיות מצביה של מכונה, בין אם מכונה מעשה ידי-אדם ובין אם מכונה כמו המוח. המצבים הנפשיים אינם יכולים להיות מצביו של אובייקט פיזיקלי כלשהו. דוגמת המחשב והיחסים בין חומרה לתוכנה מבהירים כי מן הטענה שכשאנו מתבוננים באובייקטים פיזיקליים – ואפילו בכל האובייקטים הפיזיקליים (ולו באמצעות המכשור האידאלי) – לא ניתן להבחין במאפיינים מסוימים, אין אנו רשאים להסיק שמאפיינים אלו אינם נמצאים "שם", דהיינו, שאין הם מאפייניהם של אובייקטים פיזיקליים. ייתכן שאלה הם מאפיינים מופשטים של האובייקטים הללו (מאפיינים שאינם ניתנים לתפיסה חושית). זוהי גם עמדתם של הפונקציונליסטים באשר לשאלת הגוף והנפש. על פיהם, אמנם איננו יכולים לראות את המאפיינים הנפשיים שכשאנו מתבוננים במוח; אולם בדיוק כשם שמן העובדה שכשאנו מתבוננים במחשב ובפעולתו איננו יכולים לראות את מאפייני התוכנה אין להסיק שלמחשב ולפעולתו אין מאפייני תוכנה, כך, מן העובדה שכשאנו מתבוננים במוח ובפעולתו איננו רואים את המאפיינים הנפשיים, אין להסיק כי מאפיינים אלו אינם נמצאים "שם", כלומר שמאפיינים אלו אינם מאפיינים של המוח ופעולתו. בשני המקרים, שני סוגי המאפיינים הם מאפיינים של אותה מערכת עצמה, מאפיינים השייכים לרמות שונות. המאפיינים הנפשיים טוענים, אם כן, הפונקציונליסטים, הם מאפייני התוכנה של יצורים נפשיים; במקרה האנושי ובמקרים של בעלי-חיים רבים אחרים, תוכנה זו היא התוכנה שמריץ המוח.

לא כל הפונקציונליסטים משתמשים במונחים של מחשבים ותוכנה לתיאור עמדתם, אולם יסוד חשוב של מודל הנפש כתוכנת המחשב של הגוף או של המוח הולם את עצם הרעיון הפונקציונליסטי, על כל גרסותיו. יסוד זה הוא תפיסתם של המאפיינים הנפשיים כמאפיינים מופשטים של מצבים פיזיקליים, מאפיינים השייכים לרמה אחרת מן הרמה של המאפיינים הפיזיקליים אף על פי שהמאפיינים מן הסוגים הללו הם מאפיינים של אותן מערכות של מצבים. קבלת האפשרות של ריבוי-המימושים של המאפיינים הנפשיים, אפשרות המקבילה לאפשרות ריבוי-המימושים של התוכנה, גם היא מאפיין מהותי של הפונקציונליזם על כל גרסותיו. הגרסות הפונקציונליסטיות השמות דגש מיוחד על מודל המחשב מכונות לעתים **"פונקציונליזם-מכונה"** ("Machine Functionalism"),<sup>26</sup> ולעתים **"פונקציונליזם חישובי"** ("Computational Functionalism").

<sup>26</sup> בעבר כונו מחשבים "מכונות חושבות".

### 3.1 מודל המחשב וקשרים סיבתיים – תמיהות אחדות

בהקשר זה עשויה להתעורר תמיהה. תחילה הצגנו את ההשקפה הפונקציונליסטית כהשקפה שעל פיה מצבים נפשיים מאופיינים באמצעות קשריהם הסיבתיים עם גירויים תפיסתיים, תגובות התנהגותיות, ומצבים נפשיים אחרים. קשרים אלו הם המשייכים מצב נפשי לסוגו הנפשי – הם אלו המקנים לו את מאפייניו הנפשיים הספציפיים. אולם בתיאור מודל המחשב של הנפש, לא הזכרנו עד עתה את קשרי המצבים הנפשיים למצבים אחרים, ומן הסתם לא דיברנו על קשרים סיבתיים. על מיקומה של הטענה בדבר הקשרים בין מצבים נפשיים לבין מצבים אחרים במודל המחשב נעמוד בהקדם. אולם שאלה שדורשת מענה היא, האם הטענה שעל פיה מאפיינים נפשיים הם **מאפיינים מופשטים** של יצורים נפשיים ופעולתם מתיישבת עם הטענה שמאפיינים אלו מוקנים למצבים נפשיים על ידי **קשריהם הסיבתיים** עם מצבים אחרים. לכאורה קיים קושי באשר להתיישבותן של שתי הטענות זו עם זו, משום שדומה כי קשרים סיבתיים הם מאפיינים מאוד ממשיים ולא מופשטים. העובדה שמצבים נפשיים **גורמים** את התרחשותם של מצבים אחרים, והעובדה שהם **נגרמים** על ידי התרחשותם של מצבים אחרים, אלו הן עובדות ממשיות ביותר. כיצד, אם כן, ניתן ליישב את רעיון מופשטותם של המאפיינים הנפשיים על פי מודל המחשב של הנפש עם הרעיון הפונקציונליסטי שעל פיו נקבעים המאפיינים הנפשיים על ידי קשרים סיבתיים?

**תמיהה א': כיצד ניתן ליישב את רעיון מופשטותם של המאפיינים הנפשיים על פי מודל המחשב של הנפש עם הרעיון הפונקציונליסטי שעל פיו נקבעים המאפיינים הנפשיים על ידי קשרים סיבתיים?**

נעבור עוד כברת דרך מסוימת לפני שננסה להסיר תמיהה זו מן הדרך. עתה נשים לב לנקודה אחרת, שנראה בהקדם כי היא קשורה לתמיהה זו (קל יותר להבין נקודה זו כשזוכרים שאנו מדברים על גרסה פיזיקליסטית של פונקציונליזם, המניחה שהמצבים הנפשיים הם מצבים פיזיקליים). לפי ההשקפה הפונקציונליסטית, כפילים נפשיים שקולים זה לזה לחלוטין בכוחותיהם הסיבתיים (ליתר דיוק, הם שקולים בכוחות הסיבתיים הרלוונטיים למצביהם הנפשיים – הכוחות הסיבתיים שבעטיים נגרמים מצבים נפשיים ושבועטיים הם גורמים את התרחשותם של מצבים אחרים). יחד עם זאת, הטענה שכפילים נפשיים עשויים להיות שונים לחלוטין במבניהם החומריים (וגם באותם היבטים של מבניהם החומריים הרלוונטיים למצביהם הנפשיים) מהותית להשקפה זו. זוהי תופעת ריבוי המימושים.

#### שאלה

מה המסקנה הנובעת מצירוף הטענות שכפילים נפשיים שקולים בכוחותיהם הסיבתיים ושכפילים נפשיים עשויים להיות שונים במבניהם החומריים, ומדוע מסקנה זו דורשת הסבר?



משתי הטענות הללו נובעת המסקנה שמבנים פיזיקליים שונים לחלוטין – נניח, המוח שלי ומוח הסיליקון של חיזר כלשהו מהכוכב מאדים – עשויים להיות שקולים לחלוטין בכוחותיהם הסיבתיים הרלוונטיים למצביהם הנפשיים. מסקנה זו דורשת הסבר, משום שדומה שהתכונות הפיזיקליות של אובייקטים פיזיקליים הן אלו שאחראיות לכוחותיהם הסיבתיים. אם כך, מדוע מבנים שאין ביניהם כל דומות פיזיקלית משמעותית – המבנים הפיזיקליים בהם ממומשים המצבים הנפשיים – יהיו שקולים בכוחותיהם הסיבתיים? כיצד זה שמצבים נפשיים מאותו סוג שקולים בכוחותיהם הסיבתיים גם אם מימושיהם הפיזיקליים עשויים להיות שונים במידה ניכרת?

**תמיהה ב: כיצד ניתן להסביר את העובדה שמצבים נפשיים השייכים לאותו סוג נפשי שקולים אלו לאלו בכוחותיהם הסיבתיים אף אם הם אך ממומשים באופנים פיזיקליים שונים?**

כאן עשוי להיכנס לתמונה מודל המחשב של הנפש. התשובה שהוא מספק לתמיהה זו היא: מבנים פיזיקליים שונים עשויים להיות בעלי כוחות סיבתיים זהים אם הם מהווים מימושים של אותן תוכנות. אנו יודעים שמחשבים בעלי מבנים פיזיקליים שונים – מחשבים שונים-חומרה – עשויים להיות שקולים בכוחותיהם הסיבתיים, וזאת כאשר הם מממשים את אותה תוכנה. אם כך, ההנחה שהנפש היא התוכנה של הגוף (או המוח) מסבירה את הטענה שלה מחויב הפונקציונליזם כי מבנים חומריים שונים מסוימים – המבנים החומריים של כפילים נפשיים שאינם כפילים פיזיקליים, ובפרט, המצבים הפיזיקליים המממשים את מצביהם הנפשיים – שקולים בכוחותיהם הסיבתיים. הטענה הפונקציונליסטית מוסברת על ידי הטענה שהנפש היא תוכנת הגוף, משום שאם הנפש היא תוכנת הגוף, הרי שהמבנים הפיזיקליים השונים המממשים את המצבים הנפשיים של כפילים נפשיים הם מימושים של אותה תוכנה; והם אכן שקולים בכוחותיהם הסיבתיים בדיוק כשם שמחשבים שונים חומרה המממשים אותה תוכנה שקולים בכוחותיהם הסיבתיים. כך, נראה שגם בנקודה זו הרעיון הראשוני של ההשקפה הפונקציונליסטית, שעל פיו מצבים נפשיים מאופיינים על ידי קשריהם הסיבתיים מתחבר עם מודל המחשב של הנפש. אך **טבעי הוא לדוֹגלים בהשקפה הפונקציונליסטית לקבל את מודל המחשב, העוזר להם להסביר היבט חשוב של השקפתם, ואין פלא אפוא שהוא נחשב כחלק מהותי מהשקפה זו.**

**מענה לתמיהה ב:** אם הנפש היא תוכנת הגוף הרי שהמבנים הפיזיקליים השונים המממשים את המצבים הנפשיים של כפילים נפשיים הם מימושים של אותה תוכנה; והם אכן שקולים בכוחותיהם הסיבתיים בדיוק כשם שמחשבים שונים חומרה המממשים את אותה תוכנה שקולים בכוחותיהם הסיבתיים.

מענה זה, כפי שנראה בהמשך, עדיין אינו שלם. אך עתה נזכיר כי בהקשר זה ניתן לחשוב על נימוק בזכות ההשקפה הפונקציונליסטית. למעשה, סביר שהטענה שכפילים נפשיים – יצורים נפשיים הִזֵּהים זה לזה בכל מאפייניהם הנפשיים – שקולים בכוחות הסיבתיים הרלוונטיים

למצביהם הנפשיים, תהיה מקובלת על הכול, לא רק על פונקציונליסטים. אם, לדוגמה, אצל שנינו עולה ברגע זה החשש שטיל כימי ינחת בסביבתנו בקרוב, ואנו נמצאים באותם תנאים וזהים זה לזה בכל מאפייניו הנפשיים האחרים (לשנינו אותן האמנות, אנו אינטליגנטיים במידה שווה, סדרי העדיפויות שלנו זהים, וכדומה), הרי שחשש זה, יחד עם מצבים נפשיים נוספים שאנו חולקים, יוליך את שנינו לאותם מצבים נפשיים אחרים (החלטה לנסוע לאזור אחר בארץ, או החלטה להתחפר במקלט, וכדומה), ולבסוף לאותן התנהגויות (נסיעה לאזור אחר בארץ, או התחפרות במקלט, וכדומה). אולם מי שמקבלים הנחה זאת בדבר השקילות בכוחות הסיבתיים של כפילים נפשיים, ומקבלים גם את אפשרות ריבוי-המימושים של המצבים הנפשיים, ניצבים בפני התמיהה השנייה שהעלינו לעיל. הם ניצבים בפני הצורך להסביר כיצד זה שהמבנים החומריים של כפילים נפשיים שאינם כפילים פיזיקליים – ובפרט, המצבים הפיזיקליים המממשים את מצביהם הנפשיים – שקולים בכוחותיהם הסיבתיים. לפונקציונליסטים, כפי שראינו, יש מענה לתמיהה זו, מענה הנסמך על מודל המחשב של הנפש המהווה חלק מהותי של ההשקפה הפונקציונליסטית. אם הנפש היא התוכנה של הגוף, כפי שגורס מודל זה, הרי שהמבנים הפיזיקליים השונים המממשים את המצבים הנפשיים הם מימושים של אותה תוכנה; והם שקולים בכוחותיהם הסיבתיים בדיוק כשם שמחשבים שוני חומרה המממשים אותה תוכנה שקולים בכוחותיהם הסיבתיים. כך, כשאנו מסבירים את ההשקפה הפונקציונליסטית במונחי הרעיון שהנפש היא התוכנה של הגוף (או של המוח), אנו רואים את כוחה ההסברי של השקפה זו. היא מסבירה נתון שכל מי שמקבלים את אפשרות ריבוי המימושים של המצבים הנפשיים נדרשים להסבירו. אנו נוכחים אפוא בקיומו של נימוק בזכות ההשקפה הפונקציונליסטית. נדגיש את עיקרי נימוק זה:

- (1) מצבים נפשיים ניתנים לריבוי-מימושים פיזיקלי.
- (2) כפילים נפשיים שקולים בכוחות הסיבתיים הרלוונטיים למצביהם הנפשיים [זהו הנתון האמור].
- (3) המבנים החומריים של כפילים נפשיים שאינם כפילים פיזיקליים שקולים בכוחותיהם הסיבתיים הרלוונטיים למצביהם הנפשיים [נובע מ-(1) ו-(2)].
- (4) טענה (3) דורשת הסבר: כיצד זה שמבנים חומריים שאין ביניהם כל דומות פיזיקלית משמעותית יהיו שקולים בכוחותיהם הסיבתיים?
- (5) הרעיון הפונקציונליסטי בדבר הנפש כתוכנת הגוף מסביר זאת: כיוון שהנפש היא התוכנה של הגוף, הרי שהמבנים הפיזיקליים השונים המממשים את המצבים הנפשיים הם מימושים של אותה תוכנה; והם שקולים בכוחותיהם הסיבתיים בדיוק כשם שמחשבים שוני חומרה המממשים אותה תוכנה שקולים בכוחותיהם הסיבתיים.

נראה, אם כן, שיש לנו טעם טוב בזכות קבלת הרעיון הפונקציונליסטי, שכן ההנחה שהוא אמיתי תהווה הסבר טוב לתופעה חשובה הדורשת הסבר.

מובן שאין זה **טעם מכריע** בזכות ההשקפה הפונקציונליסטית: ראשית, משום שטעם זה תלוי בהנחה בדבר ריבוי-המימושים של המצבים הנפשיים, ומי שדוחים הנחה זו (שכפי שראינו

ביחידה 4-5 אינה מקובלת ללא כל עוררין) ידחו כמובן טעם זה; שנית, משום שאיננו יכולים לשלול את אפשרות קיומם של הסברים טובים אחרים לתופעה הנדונה. עם זאת, דומה שבעבור מי שמתרשמים שקיים (או לפחות שאפשרי) ריבוי מימושים של המצבים הנפשיים, קיים כאן טעם רציני בזכות הפונקציונליזם, טעם המטה את כף המאזניים לטובת הפונקציונליזם. טעם זה יינגף אם אכן יוצעו הסברים טובים לא פחות לתופעה הנדונה, או אם יועלו ביקורות חזקות כנגד הפונקציונליזם. בהמשך יחידה זו נציג ביקורות שונות שהועלו כנגד הפונקציונליזם.

חשוב גם לעמוד על-כך שההסבר הפונקציונליסטי של הטענה כי המבנים החומריים של כפילים נפשיים שאינם כפילים פיזיקליים שקולים בכוחותיהם הסיבתיים, אינו שלם, שכן הוא מעורר שאלה. הסבר זה אמור להראות מדוע מתקיימת שקילות בכוחות הסיבתיים של כפילים נפשיים, על ידי הסתמכות על שקילות במאפייני התוכנה שלהם. אולם מאפייני התוכנה הם, כאמור, מאפיינים מופשטים, ואילו כוחות סיבתיים הם מאפיינים ממשיים ביותר – בעיניים של כוחות אלו מביאה התרחשותם של אירועים מסוימים להתרחשותם של אירועים אחרים. כיצד יכולה שקילות במאפיינים מופשטים כלשהם להסביר שקילות בכוחות סיבתיים? מאפיינים מופשטים נראים חסרי כל רלוונטיות לכוחות סיבתיים ממשיים של אובייקטים, ולא ברור אפוא כיצד יוכלו להסבירם.

עתה ניתן לראות את הקשר בין ההסבר הנדון של השקילות הסיבתית בין כפילים נפשיים שאינם כפילים פיזיקליים (ההסבר לתמיהה ב), לבין התמיהה הראשונה מבין שתי התמיהות שהעלנו, התמיהה בדבר יישוב רעיון מופשטותם של המאפיינים הנפשיים על פי מודל המחשב של הנפש (המאפיין, כאמור, את כל הגרסות של ההשקפה הפונקציונליסטית) עם הרעיון הפונקציונליסטי הבסיסי בדבר היקבעותם של המאפיינים הנפשיים על ידי קשרים סיבתיים. השאלה היא אותה שאלה: מה למאפייני תוכנה מופשטים ולכוחות סיבתיים?

התשובה לשאלה זו תעלה מהבהרת טבעם של שני היסודות הרלוונטיים – המופשטות, והכוחות הסיבתיים. ננסה להבהיר באופן מדויק יותר את הרעיון הפונקציונליסטי – הבא לידי ביטוי בולט במודל המחשב של הנפש – שעל פיו תכונות נפשיות הן תכונות מופשטות; הבהרה זו, כפי שנראה, תוביל גם להבהרת מהותו של המושג "כוחות סיבתיים" כפי שמשמשים בו בהקשרה של ההשקפה הפונקציונליסטית. הבהרתם של שני יסודות אלו היא בעלת חשיבות רבה להבנת התמונה הפונקציונליסטית בדבר טבעם של המצבים הנפשיים, גם ללא קשר לתרומתה להבהרת התמיהה וההסבר שהצגנו בפסקאות האחרונות.

השאלה "מה למאפייני תוכנה מופשטים ולכוחות סיבתיים?" קשורה הדוקות לנקודה הזו: ראינו כי כפילים נפשיים שקולים בכוחות הסיבתיים הרלוונטיים למצביהם הנפשיים. אולם הדוגמה הבאה מציבה בעיה בפני קביעה זו: נניח שאני וכפיל נפשי שלי בעל מוח סיליקון חולקים מצב נפשי – החשש שינחת טיל כימי בסביבתנו הקרובה, ושחשש זה מוליך את שנינו לעוד מצב נפשי משותף – להחלטה להתחפר במקלט (זכרו שבהיותנו כפילים נפשיים, אנו

זהים זה לזה גם בכל מאפייניו הנפשיים האחרים). בדיוק כשם שהחשש הנדון ממומש אצל שנינו על ידי מצבים פיזיקליים שונים לחלוטין זה מזה, כך גם ההחלטה שחשש זה הוליד אליה. כלומר, החשש שלי והחשש המקביל של כפילי בעל מוח הסיליקון **גורמים התרחשות מצבים פיזיקליים שונים לחלוטין זה מזה** (שהרי מוחותינו שונים לחלוטין זה מזה מבחינת הרכבם הכימי-פיזיקלי). במובן זה, מצבים נפשיים זהים **אינם שקולים** בכוחותיהם הסיבתיים (כאמור, הם גורמים מצבים שונים – מצבים הנבדלים אלו מאלו בתכונותיהם הפיזיקליות), וכפילים נפשיים **אינם שקולים** בכוחותיהם הסיבתיים הרלוונטיים למצביהם הנפשיים. אולם ראינו כי כפילים נפשיים **שקולים** בכוחות הסיבתיים הרלוונטיים למצביהם הנפשיים. האמנם לפנינו סתירה? אין סתירה, כפי שנראה בהמשך, בדיוק משום שהמובן של הביטוי "כוחות סיבתיים" כפי שמשמשים בו בהקשרה של ההשקפה הפונקציונליסטית הוא מיוחד. נוכל לעמוד על מיוחדות זו לאחר שנציג מודל קלאסי של מחשב ששיחק תפקיד חשוב בהתפתחות ההשקפה הפונקציונליסטית.

## 3.2 פונקציונליזם ומכונת טיורינג

### 3.2.1 מכונת טיורינג ופעולתה

מכונת טיורינג (Turing Machine) היא דגם מופשט של מחשב, שהרעיון ביסודו פותח על ידי המתמטיקאי והלוגיקן הבריטי אלן טיורינג (Alan Turing), שחי במחצית הראשונה של המאה ה-20.<sup>27</sup> רעיון זה נחשב כאחד ההישגים האינטלקטואליים החשובים ביותר בתקופה המודרנית – הוא נחשב כרעיון שעומד ביסוד פעולתו של המחשב המודרני. רעיון זה השפיע רבות על פיתוחה של ההשקפה הפונקציונליסטית, בעיקר בראשית ימיה, והוא שיחק תפקיד חשוב במיוחד בגרסה הראשונה של הפונקציונליזם שפיתח הילארי פטנאם, שכאמור רבים רואים בו את אבי הפונקציונליזם. לעתים מכונה גרסה זו "פונקציונליזם של מכונת טיורינג" ("Turing-Machine Functionalism"). הצגת הרעיון של מכונת טיורינג עשויה לעזור לנו להבין מושגים ורעיונות שונים המשחקים תפקיד במודל המחשב של הנפש בכלל (ולאו דווקא במודל המחשב הספציפי המתייחס למכונת טיורינג).

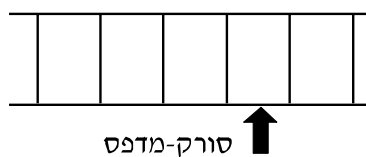
מהי מכונת טיורינג?<sup>28</sup> כאמור, מכונת טיורינג היא דגם מופשט של מחשב שהמבנה הבסיסי שלו הוא זה:<sup>29</sup> למכונה זו סרט המחולק למשבצות או "תאים", ו"סורק/מדפס" – אמצעי

<sup>27</sup> במלחמת העולם השנייה מילא טיורינג תפקיד מפתח בפיצוח הצופן שבו השתמשו הכוחות הגרמניים (צופן ה"אניגמה"). טיורינג שם קץ לחייו בשנת 1954, בהיותו בן 42, בעקבות רדיפתו בשל היותו הומוסקסואל.

<sup>28</sup> הרעיון של מכונת טיורינג הוצג על ידי טיורינג במאמרו "On Computable Numbers, with an Application to the Entscheidungsproblem", *Proceedings of the London Mathematical Society, Series 2*, 1936, pp. 230-265.

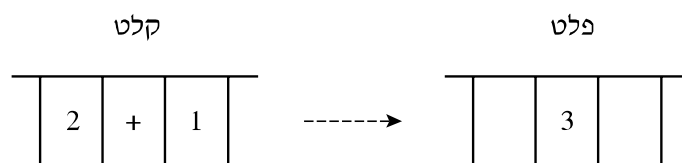
<sup>29</sup> הסבר הרעיון של מכונת טיורינג כרוך בהכרח בפרטים טכניים אחדים, אולם גם אם בתחילה אלו ייראו לכם מסובכים, התאזרו נא בסבלנות, שכן הדוגמאות שתוצגנה בהמשך תבהרנה את הדברים.

אשר "קורא" את הסימנים הכתובים בתאי הסרט, מדפיס בהם סימנים, או מוחק סימנים קיימים. הסורק/מדפס עובר בין תאי הסרט, בהתאם לכללים מסוימים שמייד נסבירם, ו"קורא" את תוכנו של התא, כלומר מזהה איזה סימן מודפס בתא (אם מודפס בו סימן). לאחר קריאת הסימן, פועל הסורק/מדפס פעולה מסוימת או פעולות מסוימות בהתאם לכללים. הפעולות האפשריות הן מחיקת סימן, הדפסת סימן, מעבר לתא אחר של הסרט (באחד משני הכיוונים האפשריים – ימינה או שמאלה), או עצירה, ופעולות מסוג נוסף שמייד נציין.



### תרשים 7. מכונת טיורינג

לפני תחילת פעולת המכונה מוזנים לה נתונים, דהיינו מודפסים סימנים שונים על חלק מתאי הסרט (למשל, הנתונים " $2+1$ "), ונתונים אלו קובעים את פעולת החישוב שעל המכונה לבצע. נתונים אלו – אוסף הסימנים המודפסים על תאי הסרט לפני תחילת פעולת הסורק/מדפס – מכונים ה"קלט" (input) של המכונה. אוסף הסימנים המודפסים על תאי הסרט לאחר תום פעולותיה של נקרא מכונה ה"פלט" (output) של המכונה. זוהי תוצאת החישוב, הנתונים שהמכונה מפיקה בתגובה לקלט. למשל, המכונה עשויה לבצע חישוב מתמטי של חיבור: הקלט שלה יהיה שני מספרים (נניח, 2 ו-1) וסימן החיבור, והפלט שלה צריך להיות מספר אחר, המספר המהווה את סכום שני המספרים. אם הכללים שעל פיהם פועלת המכונה הם כללים נכונים, המכונה תפיק את הפלט הנכון, כלומר את הסכום האמיתי של שני המספרים המהווים את הקלט (במקרה זה, 3).



### תרשים 8. קלט ופלט של חישוב במכונת טיורינג (תיאור ראשוני)<sup>30</sup>

באיור זה מצוינים הקלט והפלט בספרות רגילות. אולם מכונת טיורינג אמורה לפעול על סימנים פשוטים יותר: מספרים (ולמעשה נתונים כלשהם) מצוינים על-גבי הסרט כקבוצות

<sup>30</sup> שימו לב, שהמכונה מוחקת את הקלט ובמקומו מדפיסה את הפלט; היא אינה שומרת את הקלט (על פי פיתוחים אחדים של הרעיון של מכונת טיורינג, למכונה שני סרטים, ואז היא אינה מוחקת את הפלט).

של יחידות בודדות. דהיינו, המספר 3 מודפס כ-"1 1 1", המספר 4 כ-"1 1 1 1", וכדומה (את הספרה "1" ניתן למעשה להחליף בכל סימן אחר, ובלבד שנחליפה באותו סימן בכל התאים שבהם היא מופיעה, ושסימן זה אינו ממלא במכונה כל תפקיד אחר. הדבר נכון לגבי כל סימן מסימני הקלט והפלט). כמו כן, הן קבוצת סימני הקלט והן קבוצת סימני הפלט תחומות משני צדדיהן על ידי סימן כלשהו, נניח "#". אם כן, הצגה יותר מדויקת של הקלט והפלט של החישוב הנדון ( $2+1$ ) ייראו כך:



### תרשים 9. קלט ופלט של חישוב במכונת טיורינג

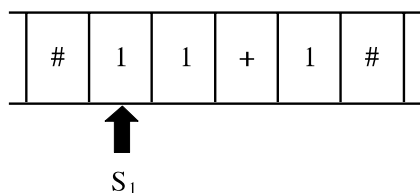
כיצד מבצעת המכונה את החישוב? חשוב להבהיר שכאשר אנו שואלים זאת איננו מתכוונים לשאול על טיבו של המנגנון הפיזיקלי העומד ביסוד פעולת המכונה, למשל, איזה סוג מנוע "מסיע" את הסורק/מדפס מתא לתא, או מה הלבוש הפיזיקלי שלובשות הוראות הכללים שלהן מצייתת המכונה. בלשון אחרת, אין לנו עניין ברמת החומרה. שאלתנו היא: **מהם הכללים** שעל פיהם פועל הסורק/מדפס? פעולה על **פי אלו כללים** תוליד את המכונה מן הפלט שבצד שמאל אל הקלט שבצד ימין? שאלתנו זו היא ברמת התוכנה.

לפני שנשיב לשאלתנו זו, נציין שלסורק/מדפס יש גם **מצבים פנימיים** אפשריים אחדים. המצבים הפנימיים הללו של הסורק/מדפס – המכונים על פי רוב "**מצבי מכונה**" – הם מצבים המשנים את אופן התגובה של הסורק/מדפס לקלט שאותו הוא מזהה. כלומר, סדרת הפעולות שעל הסורק/מדפס לבצע בשלב מסוים נקבעת על ידי שני גורמים: על פי תוכן התא שבו נמצא הסורק/מדפס באותו שלב (כלומר, על פי הסימן המודפס בתא זה), ועל פי מצבו הפנימי של הסורק/מדפס באותו שלב. במילים אחרות, הסורק/מדפס אינו מגיב בהכרח באותם אופנים כשהוא מזהה את אותו סימן על-גבי הסרט. ייתכן שהכללים יקבעו שאם הוא מזהה בתא הנסרק את הסימן "1" למשל, והוא נמצא במצב פנימי מסוים עליו לבצע סדרת פעולות מסוימת; ואם הוא מזהה בתא הנסרק את הסימן "1" אבל הוא נמצא במצב פנימי אחר, עליו לבצע סדרת פעולות אחרת. מצבו הפנימי של הסורק/מדפס בכל שלב מכתוב את אופני התגובה שלו לסימנים שהוא מזהה. נמחיש זאת באמצעות דוגמת החישוב של  $2+1$ .

עלינו, כאמור, לקבוע את הכללים שיאפשרו למכונה לעבור מן המצב המופיע בצד שמאל של האיור (הפלט) אל המצב המופיע בצד ימין (הקלט). במילים אחרות, עלינו "לתכנת" חישוב של פעולת חיבור בעבור מכונת טיורינג. כפי שנראה, ייתכנו תכניות חיבור שונות.

במצב ההתחלי הסורק/מדפס נמצא במצב מכונה (מצב פנימי) שנצינו כ- $S_1$ , והוא קורא את התא הראשון שמימין לסימן "#". **(כל חישוב של פעולת טיורינג מתחיל כך)**. האיור שלפניכם מראה את המצב ההתחלי של הסרט והסורק/מדפס:<sup>31</sup>

### מצב 1:



הכלל הראשון שמעניין אותנו צריך אפוא להורות לסורק/מדפס מה לעשות כאשר הוא נמצא במצב מכונה  $S_1$  ומזהה את הספרה "1" בתא שהוא סורק. ובכן, מתכנתת מסוימת הציעה להתחיל את החישוב בכלל הבא:<sup>32</sup>

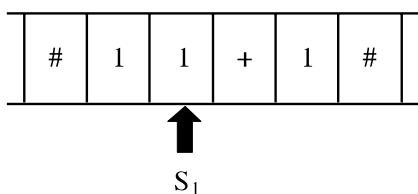
- (א) כשהסורק/מדפס נמצא במצב מכונה  $S_1$  ומזהה את הסימן "1", עליו:
- I. לא לשנות את תוכן התא (כלומר, להשאיר בתא את הסימן "1").
  - II. לנוע צעד אחד ימינה.
  - III. להישאר במצבו הפנימי (מצב מכונה)  $S_1$ .

### שאלה

התוכלו לאייר את מצב הסרט והסורק/מדפס לאחר ביצוע כלל (א) במצב 1?

לאחר ביצועו של כלל (א), נקבל מצב דומה מאוד למצב ההתחלי. כאמור, אין שינוי בתוכן התא שנסרק בשלב הראשון (עדיין מופיע בתא זה הסימן "1"), ואין גם שינוי במצב המכונה (שנותר  $S_1$ ). השינוי היחיד הוא שהסורק/מדפס קורא בשלב הנוכחי את התא הבא מימין. כלומר, המצב המתקבל הוא זה:

### מצב 2:

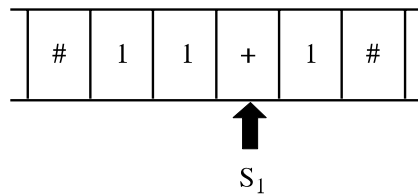


<sup>31</sup> שימו לב שמצב הסורק/מדפס מורכב הן מהמצב הפנימי שלו והן מתוכן התא שהסורק/מדפס קורא.  
<sup>32</sup> איננו נכנסים להסבר הגיונם של הכללים. הנקודה החשובה היא שאפשר יהיה להיווכח שהכללים מספקים את התוצאות הנכונות.

**שאלה**

האם בשלב זה יש צורך בכלל פעולה חדש, או שמא הכלל הרלוונטי לפעולתו הבאה של הסורק/מדפס הוא שוב כלל (א)?

מאחר שכללי הפעולה תלויים בשני גורמים – תוכן התא הנסרק ומצבו הפנימי של הסורק/מדפס (מצב-המכונה) – ושני הגורמים הללו לא השתנו (הסורק/מדפס אמנם קורא תא אחר, אולם תוכנו של התא האחר זהה לתוכנו של התא הקודם, כלומר "1"), הרי שכלל הפעולה הרלוונטי בשלב זה הוא שוב כלל (א), הכלל המורה לסורק/מדפס מה לעשות כשהוא נמצא במצב פנימי  $S_1$  ומזהה את הספרה "1". הפעלתו של כלל (א) בשלב הנוכחי תוביל אותנו למצב הבא (אנא, בדקו בעצמכם!):

**מצב 3:**

ההבדל היחיד לעומת השלב הקודם הוא שהסורק/מדפס קורא עתה תא אחר. כיוון שבתא זה מופיע סימן אחר, סימן ה-"+", כלל (א) אינו רלוונטי (כאמור, כלל זה מורה לסורק/מדפס מה לעשות כשהוא נמצא במצב פנימי  $S_1$  ומזהה את הספרה "1"). נדרש אפוא כלל פעולה חדש. המתכנתת הציעה את הכלל הבא:

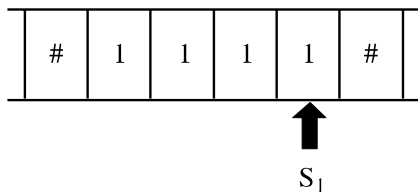
(ב) כשהסורק/מדפס נמצא במצב מכונה  $S_1$  ומזהה את הסימן "+", עליו:

I. להדפיס בתא את הסימן "1" במקום הסימן "+".

II. לנוע צעד אחד ימינה.

III. להישאר במצבו הפנימי (מצב מכונה)  $S_1$ .

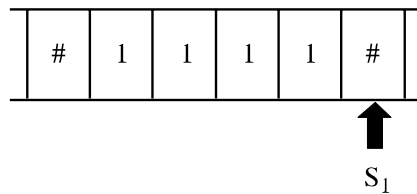
עתה יתקבל המצב הבא (אנא, בדקו בעצמכם!):

**מצב 4:**

כיוון שהסורק/מדפס שוב נמצא במצב פנימי  $S_1$  ומזהה את הסימן "1", הכלל הרלוונטי לפעולתו בשלב זה הוא, שוב, כלל (א). נפעילו ונקבל את המצב הבא:



### מצב 5:

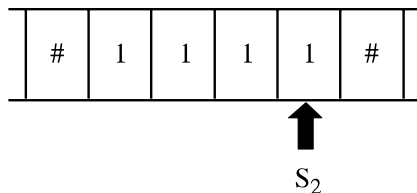


עתה קורא הסורק/מדפס סימן אחר – הסימן "#". לפיכך נדרש כלל חדש, כלל שיורה לו מה לעשות כשהוא נמצא במצב פנימי  $S_1$  ומזהה את הסימן "#". הכלל שהציעה המתכנתת הוא זה:

- (ג) כשהסורק/מדפס נמצא במצב מכונה  $S_1$  ומזהה את הסימן "#", עליו:
- I. לא לשנות את תוכן התא (כלומר, להשאיר בתא את הסימן "#").
  - II. לנוע צעד אחד שמאלה.
  - III. לעבור למצב פנימי (מצב מכונה)  $S_2$ .

נפעיל את כלל (ג) ונקבל את המצב הבא:

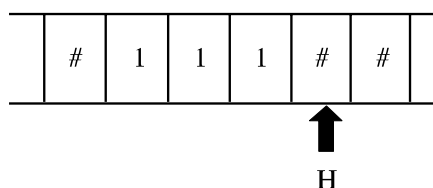
### מצב 6:



מאחר שהסורק/מדפס נמצא עתה במצב פנימי שונה ( $S_2$ ), אנו זקוקים לכלל חדש. המתכנתת הציעה את כלל (ד):

- (ד) כשהסורק/מדפס נמצא במצב מכונה  $S_2$  ומזהה את הסימן "1", עליו:
- I. להדפיס בתא את הסימן "# במקום הסימן "1".
  - II. לעצור את תנועתו.
- נפעיל את כלל (ד) ונקבל את המצב הזה:

### מצב 7:



הסורק/מדפס עצר (האות "H" מסמנת את העצירה) – פעולת מכונת הטיורינג הסתיימה. התוצאה הסופית, הפלט, היא "1 1 1", כלומר 3. הגענו לתוצאה הנכונה – החישוב הצליח.<sup>33</sup>

ארבעת הכללים שעל פיהם התבצע החישוב מהווים **תכנית** – תכנית לחיבור מספרים – של מכונת טיורינג. לתכנית כזו – לרשימת כללים מעין זו המכתיבה לאמצעי חישוב אילו פעולות לבצע בכל שלב ושלב בהתאם לנתונים שבהם הוא נתקל בשלב זה – קוראים לעתים "אלגוריתם".

ייתכן שחלק מן הצעדים שתכנית זו הכתיבה באשר לחישוב שהדגמנו ייראו לכם מוזרים – למשל, התנועה ימינה עד קצה שרשרת הסימנים ואז החזרה שמאלה ומחיקת סימן שנכתב בשלב קודם, אולם אם תנסו לבצע את פעולת החיבור הזו בדרכים אחרות תגלו שכדי להגיע לתוצאה הרצויה נדרש "טריק" מסוג כלשהו.<sup>34</sup> על כל פנים, אין זה חשוב לעניינו אם זוהי תכנית חכמה במיוחד או לא. ניתן לראות שזו תכנית חיבור יעילה (היא תנפיק תוצאות נכונות לכל פעולות החיבור), וכך להבין את הרעיון שביסוד פעולתה של מכונת טיורינג – כיצד היא מבצעת חישובים.

כאשר נתונה לנו תכנית כלשהי של מכונת טיורינג, כל שעלינו לעשות כדי לדעת כיצד יתבצע חישוב מסוים הוא לזהות בכל שלב איזה כלל יש להפעיל, ולהפעילו. לשם זיהוי הכלל עלינו לדעת שני נתונים: את המצב הפנימי ( $S_1$ , או  $S_2$ , וכדומה) באותו שלב, ואת תוכן התא שהסורק/מדפס קורא באותו שלב. לאחר שזיהינו את הכלל הרלוונטי, עלינו למלא אחר הוראותיו: למחוק ולהדפיס את הסימנים בתא, כפי שהכלל מורה; להזיז את הסורק/מדפס ימינה או שמאלה, או לעצור את תנועתו; להעביר את הסורק/מדפס למצב פנימי אחר או להשאירו באותו מצב.

כפי שאמרנו, אותה פעולת חיבור ניתנת לביצוע בדרכים שונות, כלומר בהתאם לתכניות שונות.

## שאלה

התוכלו לתכנת תכנית לחישוב פעולות חיבור השונה מן התכנית שהוצגה לעיל?

<sup>33</sup> אפשר להתעלם מן העובדה שמופיעים בצד ימין שני סימני "#". כאמור, סימנים אלו מיועדים לתחם את קבוצות סימני הקלט והפלט ואין כל חשיבות לכמות הסימנים הללו המופיעים בכל קצה.

<sup>34</sup> ההסבר למהלך של תנועה ימינה ואז חזרה שמאלה ומחיקת סימן שנכתב בשלב קודם, הוא שהמכונה כותבת "1" במקום הסימן "+", כך, אנו מקבלים את הסימן "1" פעם אחת "יותר מידי" (הן בעבור כל היחידות המרכיבות את שני המספרים שאנו מחברים והן עבור הסימן "+"); החזרה אחורנית ומחיקת "1" מתקנת זאת.

הנה תכנית אחרת לחישוב הסכום  $2+1$  :

(א) כשהסורק/מדפס נמצא במצב מכונה  $S_1$  ומזהה את הסימן "1", עליו :

I. להדפיס בתא את הסימן "#" במקום הסימן "1".

II. לנוע צעד אחד ימינה.

III. לעבור למצב פנימי (מצב מכונה)  $S_2$ .

(ב) כשהסורק/מדפס נמצא במצב מכונה  $S_2$  ומזהה את הסימן "1", עליו :

I. לא לשנות את תוכן התא (כלומר, להשאיר בתא את הסימן "1").

II. לנוע צעד אחד ימינה.

III. להישאר במצבו הפנימי (מצב מכונה)  $S_2$ .

(ג) כשהסורק/מדפס נמצא במצב מכונה  $S_2$  ומזהה את הסימן "+", עליו :

I. להדפיס בתא את הסימן "1" במקום הסימן "+".

II. לנוע צעד אחד ימינה.

III. להישאר במצבו הפנימי (מצב מכונה)  $S_2$ .

(ד) כשהסורק/מדפס נמצא במצב מכונה  $S_2$  ומזהה את הסימן "#", עליו :

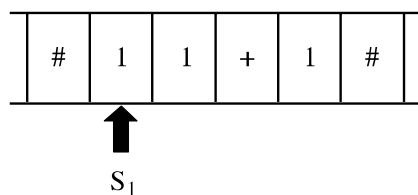
I. לא לשנות את תוכן התא (כלומר, להשאיר בתא את הסימן "#").

II. לעצור את תנועתו.

## שאלה

התוכלו להראות, שלב אחר שלב, כיצד תבצע מכונת טיורינג את פעולת החיבור  $2+1$  על פי תכנית זו?<sup>35</sup>

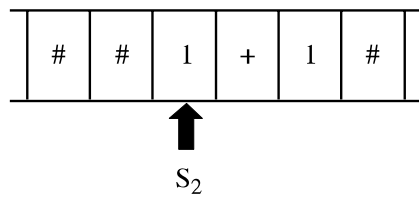
## מצב 1:



<sup>35</sup> אין זה מענייננו בקורס זה לפתח אצל הקוראים מיומנות בביצוע חישובי מכונת טיורינג (וודאי שלא לפתח אצל הקוראים מיומנות בתכנות מכונת טיורינג; הנקודה החשובה היא הבנת עקרון-הפעולה של המכונה. אם אינך מצליח/ה לבצע את החישובים שאנו מדגימים בעצמך, אין בכך כדי לרמוז שלא תוכל/י להבין את נושאה של יחידה זו.

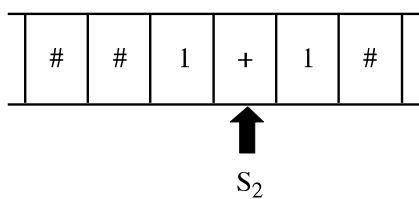
נפעיל את כלל (א), ונקבל:

**מצב 2:**



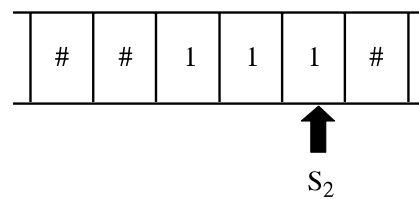
נפעיל את כלל (ב), ונקבל:

**מצב 3:**



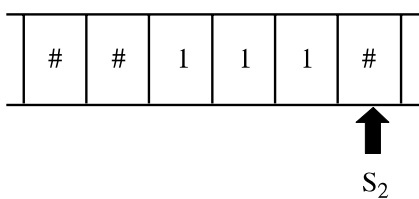
נפעיל את כלל (ג), ונקבל:

**מצב 4:**



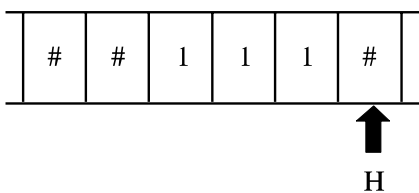
נפעיל את כלל (ב), ונקבל:

**מצב 5:**



נפעיל את כלל (ד), ונקבל:

**מצב 6:**



הסורק/מדפס עצר – פעולת המכונה הגיעה לסיומה. התוצאה הסופית, הפלט, היא שוב "1 1 1", כלומר 3. הגענו לתוצאה הנכונה גם בדרך החישוב החלופית.

ראינו, אם כן, שמכונות טיורינג יכולות לבצע את אותם חישובים בדרכים שונות – על פי אלגוריתמים שונים או תכניות שונות, דהיינו קבוצות שונות של כללים. מכונות טיורינג יכולות להגיע באופן שיטתי מאותם קלטים לאותם פלטים בדרכים שונות. בהנחה ששתי תכניות חיבור המספרים שהצגנו הן אמנם תכניות חיבור יעילות (תכניות המבטיחות את קבלת התוצאות הנכונות באשר לכל פעולות החיבור), אנו יכולים לומר שאלה הן **"תכניות שקולות קלט-פלט"**. **תכניות שונות של מכונת טיורינג – תכניות הפועלות על פי כללים שונים ומבצעות צעדים שונים בהתאם לכללים – עשויות אפוא להיות תכניות שקולות קלט-פלט** (כלומר, כאלו שמפיקות אותו פלט בכל מקרה מוזנות באותו קלט). זהות מוחלטת בין תכניות מכונת טיורינג מחייבת שקילות קלט-פלט, אולם שקילות קלט-פלט אינה מספיקה לשם זהות מוחלטת. שתי תכניות של מכונת טיורינג הן זהות אם, ורק אם, הן מורכבות מאותם כללים בדיוק; וכך, אם שתי מכונות טיורינג ממשיות מבצעות את אותו חישוב על פי אותה תכנית, על פי אותו אלגוריתם, הן תגענה מאותם קלטים לאותם פלטים באמצעות אותם צעדים.

בשל תפקיד מהותי זה של הצעדים באפיון תכנית של מכונת טיורינג, אי אפשר לאפיין תכניות של מכונת טיורינג באמצעות דגמי הקלט-פלט הכרוכים בהן, או, במילים אחרות, באמצעות דגמי גירוי-תגובה. במינוח המוכר לנו בפילוסופיה של הנפש, פירוש הדבר שאי אפשר לאפיין תכנית של מכונת טיורינג אפיון התנהגותני, אפיון ביהביוריסטי (זכרו: אפיון ביהביוריסטי של מצב נפשי הוא אפיונו במונחי דגמי הקלט-פלט, או הגירוי-תגובה, שהוא כרוך בהם).

הדגמנו את ביצועה של פעולת חישוב מסוג אחד על ידי מכונת טיורינג – פעולת החיבור של מספרים. אולם מכונות טיורינג מסוגלות לבצע פעולות חישוב מסוגים רבים, הן פעולות מתמטיות אחרות והן פעולות שאינן מתמטיות. אחת ההשלכות החשובות של הרעיון של מכונת טיורינג היא שכל חישוב (דהיינו הגעה מקלט מסוים לפלט מסוים בהתאם לכללים כלשהם) שבאופן עקרוני ניתן לבצעו, כל חישוב שניתן לביצוע על ידי מחשב כלשהו, ניתן לביצוע על ידי מכונת טיורינג מן הסוג המתאים (מכונת חיבור, מכונת חיסור, וכדומה).<sup>36</sup> לא נרחיב כעת בנקודה זו, שיש לה חשיבות רבה באשר לענף המתמטי של חישוביות, אולם בהמשך נראה את הרלוונטיות שלה לענייננו.

אנו אומרים שמכונות טיורינג מסוגלות לבצע פעולות מסוגים שונים, ואף ראינו שמכונות טיורינג מסוגלות לבצע את אותן פעולות חישוב בדרכים שונות – באמצעות תכניות שונות. האם ייתכן שאותה מכונה פיזיקלית תתפקד הן כמכונת טיורינג לחיבור מספרים והן כמכונת

<sup>36</sup> על השלכה זו עמדו, ללא קשר ביניהם, הן טיורינג עצמו והן המתמטיקאי והלוגיקן אלונזו צ'רץ' (Alonzo Church). היא מכונה "תזת צ'רץ'-טיורינג" (לעתים היא מיוחסת רק לצ'רץ', ומכונה "התזה של צ'רץ'"). הניסוח המדויק של תזה זו והסברה חורגים ממטרותינו בקורס זה.

טיורינג לחילוק מספרים, למשל? מובן שניתן לשנות באופן מכני את תפקודה של מכונה – לדוגמה, להעביר אותה מתפקוד כמכונת חיבור לתפקוד כמכונת חיסור, או מתפקוד כמכונת חיבור לפי תכנית אחת לתפקוד כמכונת חיבור לפי תכנית שנייה – למשל באמצעות מתג כלשהו. שינוי כזה יתבטא בשינוי בכללים שהמכונה פועלת על פיהם כשהיא "קוראת" קלט מסוים. אולם האם עשויה מכונת טיורינג לבצע משימות חישוב מסוגים שונים מבלי שנשנה את תפקודה באופן מכני?

ובכן, טיורינג פיתח דגם של מכונה הקרויה "**מכונת טיורינג אוניברסלית**", מכונת טיורינג שניתן להציג לה, כחלק מהקלט – דהיינו באמצעות סימנים על-גבי הסרט – **את הכללים** שעל פיהם עליה לפעול כשאנו מציגים לה משימת חישוב מסוימת (למשל חיבור). יש דרך להציג למכונה על-גבי הסרט הן את משימת החישוב המבוקשת והן את כללי החישוב הרלוונטיים למשימה זו (כללי חיבור כשמדובר בפעולת חיבור, כללי חילוק כשמדובר בפעולת חילוק, וכדומה). וכך, אותה מכונת טיורינג אוניברסלית תוכל לבצע משימות חישוב שונות מבלי שיעשה בה שינוי תפקוד כלשהו: המכונה "תקרא" מה מודפס על הסרט בכל מקרה (הכללים והנתונים) ו"תדע" אלו צעדים עליה לבצע – כלומר, הנתונים שתסרוק יקבעו מה תבצע – בכל שלב. פירוש הדבר הוא ש**מכונת טיורינג אוניברסלית היא מחשב רב תכליתי**. אין היא מוגבלת למשימות חישוב מסוג אחד בלבד.

הזכרנו קודם את הקביעה כי כל חישוב שבאופן עקרוני ניתן לביצוע, כל חישוב שניתן לביצוע על ידי מחשב כלשהו, ניתן לביצוע על ידי מכונת טיורינג מסוימת. אנו יכולים לומר ש**מכונת טיורינג אוניברסלית יכולה לבצע כל חישוב – כל מעבר בין קלט לפלט הנשלט על ידי חוקיות או כללים כלשהם – שניתן באופן עקרוני לביצוע**. במילים אחרות, מכונת טיורינג אוניברסלית יכולה לבצע כל חישוב שניתן לביצוע על ידי מחשב כלשהו. אין צורך במכונות שונות, שכל אחת מהן מסוגלת לבצע חישובים מסוג מסוים או מקבוצה מוגבלת של סוגי חישובים. מכונת טיורינג אוניברסלית יכולה אפוא להוות דגם לכל מחשב שהוא – היא דגם של מחשב רב תכליתי במובן החזק ביותר.

## 3.2.2 מימושים פיזיקליים של מכונת טיורינג, ואפיון מצבי תוכנה

### 3.2.2.1 המימוש של תוכנה ומצביה

שימו לב, שעד כה דברנו על מכונת טיורינג ברמת התוכנה בלבד. כלומר, לא התייחסנו כלל לשאלות בדבר אופן המימוש הפיזיקלי של מכונה זו. ניתן לראות שמכונת טיורינג עשויה ללבוש צורות פיזיקליות שונות. למשל, הסרט עשוי להיות סרט נייר שעליו מוטבעים הסימנים בדיו, כאשר בסורק/מדפס מותקן קורא אופטי הקורא את סימני הדיו (כדוגמת הקורא האופטי המזהה את קוד המחיר שעל-גבי מוצרים בחנויות). אולם הסרט עשוי גם להיות סרט מגנטי, שהסימנים "רשומים" בו באופנים חשמליים ומזוהים על ידי סורק מתאים (במקרה

כזה, מובן שהסימנים השונים לא "ייראו" כפי שהם נראים על-גבי סרט-נייר. סימנים כמו "1", "+" וכדומה יהיו מאפיינים חשמליים של הסרט המגנטי). וניתן, כמובן, לחשוב על צורות פיזיקליות נוספות בהן מכונת טיורינג ניתנת למימוש, הן באשר לאמצעי לציון הקלט והפלט והן באשר למצבים הפנימיים.<sup>37</sup> אנו רואים, אם כן, שמכונת טיורינג ניתנת לריבוי מימושים. אין זה מקרי. לומר שאובייקט מסוים הוא מכונת טיורינג משמעו לתאר אותו ברמת התוכנה, ברמה "החשובית"; לא ברמת הרכבו הפיזיקלי ותכונותיו הפיזיקליות, אלא ברמת פעולות החישוב שהוא מסוגל לבצע והכללים על פיהם הוא מסוגל לפעול. המושג של מכונת טיורינג הוא מושג מופשט.

למעשה, התיאור הסטנדרטי של מכונת טיורינג (שאותו הצגנו גם אנחנו) אינו תיאור ברמה המופשטת, שכן, בין היתר, זהו תיאור של סרט, המחולק לתאים במימד אחד (לאורכו), שעליו מופיעים סימני הקלט והפלט. בכך נכנס התיאור במידה כלשהי לרמת המימוש הפיזיקלי: עצם הרעיון של מכונת טיורינג אינו מחייב שהאמצעי לציון הקלט והפלט ילבש צורה של סרט כזה. ניתן היה לדמיין מימוש של מכונת טיורינג שבו הקלט והפלט מופיעים כהבהובי נורות בצבעים שונים, כצלילים שונים (אם כי תכנון מכונה עם קלט/פלט צלילי ידרוש יצירתיות גדולה יותר), ועוד.

נניח שאנו דנים במכונות טיורינג לא אוניברסליות, המבצעות פעולות מסוג אחד בלבד (מכונת חיבור, מכונת חילוק, וכדומה), ועל פי תכנית אחת בלבד. נניח שנבנה שתי מכונות כאלו בפועל. שתי מכונות כאלו עשויות להיות שונות מאוד זו מזו מבחינה פיזיקלית, ועדיין להיות **שקולות מבחינה חישובית**, כלומר, כאלו המבצעות בדיוק את אותן פעולות (נניח ששתיהן מכונות המבצעות פעולות חיבור), ועל פי אותה תכנית (נניח, אחת משתי תכניות החיבור שהצגנו). במקרה כזה נאמר שמדובר באותה מכונת טיורינג, או בשתי מכונות טיורינג מאותו סוג. במילים אחרות, קריטריון הסיווג, קריטריון האינדיווידואציה של מכונת טיורינג ניתן במונחי פעולות-החישוב שהמכונה מבצעת ותכניות הביצוע של הפעולות הללו; קריטריון זה הוא קריטריון חישובי. שתי מכונות חישוב הן אותה מכונת טיורינג (הן שייכות לאותו סוג מכונות טיורינג), אם, ורק אם, הן מבצעות את אותם חישובים ובאותן דרכים. כך, אנו מזהים מכונת טיורינג לפי התכנית שלה – לפי הכללים הקובעים את כללי פעולתה. וכזכור, תכנית של פעולת מכונת טיורינג איננה ניתנת לאפיון ביהיוריסטי, במונחים של דגמי קלט-פלט. זהותה של תכנית כזו נקבעת גם על פי שלבי הפעולה הפנימיים שלה. ניתן אפוא לומר, שמכונת **טיורינג מאופיינת אפיון פונקציונליסטי**. אנו נוסיף ונסביר טענה זו בהמשך.

<sup>37</sup> שתי הצורות האפשריות שהזכרנו למימוש מכונת טיורינג, וצורות נוספות, מוצגות על ידי ג' קים. ראו: J. Kim, *Philosophy of Mind*, 1996, Boulder: Westview Press, pp. 85-86.

קיימת דרך מקוצרת להציג תכניות מכונת טיורינג – באמצעות "טבלת מכונה".

	$S_1$	$S_2$
1		
+		
#		

### תרשים 10. טבלת מכונה

הטורים (האנכיים) בטבלת המכונה מציינים את מצבי המכונה, והשורות (האופקיות) מציינות את הסימנים העשויים להופיע על-גבי הסרט. כך, ההוראות שתיכתבנה במשבצת השמאלית העליונה מתייחסות לסיטואציה שבה המכונה נמצאת במצב פנימי  $S_1$  ומזהה את הסימן "1" בתא הנסרק; ההוראות שתיכתבנה במשבצת השמאלית השנייה מלמעלה מתייחסות לסיטואציה שבה המכונה נמצאת במצב פנימי  $S_1$  ומזהה את הסימן "+"; בתא הנסרק, וכדומה. כיצד תיכתבנה ההוראות? ניקח לדוגמה את כלל (א) של תכנית חיבור המספרים השנייה שהצגנו.

(א) כשהסורק/מדפס נמצא במצב פנימי  $S_1$  ומזהה את הסימן "1", עליו:

I. להדפיס בתא את הסימן "#"; במקום הסימן "1".

III. לנוע צעד אחד ימינה.

VI. לעבור למצב פנימי (מצב מכונה)  $S_2$ .

הוראות אלו תיכתבנה כך:  $\#RS_2$ .

הציון "#"; משמעו הוראה לסורק/מדפס להדפיס "#"; בתא הנסקר במקום הסימן המצוי בו; הציון "R" משמעו הוראה לסורק/מדפס לזוז אל התא הנמצא מימין לתא הנוכחי (קיצור של "Right"); הציון  $S_2$  משמעו הוראה לסורק/מדפס להימצא במצב פנימי  $S_2$ . אם הוראת העצירה תצוין על-ידי "H" (קיצור של "Halt"), טבלת-המכונה של מכונת טיורינג הפועלת על פי תכנית חיבור זו תראה כך:



	$S_1$	$S_2$
1	#R $S_2$	1R $S_2$
+		1R $S_2$
#		#H

**תרשים 11.** טבלת מכונה למכונת חיבור

אמרנו שמכונת טיורינג מזוהה על פי תכניתה. ובכן, דרך חלופית להביע רעיון זה הוא לומר שמכונת טיורינג מזוהה על פי טבלת המכונה שלה.

### 3.2.2.2 אפיונו של מצב תוכנה (מצב חישובי)

מה פירוש הטענה שאמצעי פיזיקלי כלשהו מממש תכנית כלשהי? כיצד ממומשת מכונת טיורינג באופן פיזיקלי? מה הקשר בין התהליכים הפיזיקליים של פעולתה הממשית של המכונה הממשית, הפיזיקלית, לבין השלבים המופשטים של תכנית זו או אחרת? באיזה מובן מצב מסוים של המכונה הפיזיקלית מהווה מימוש של מצב מופשט מסוים של תכנית של מכונת טיורינג?

ובכן, אולי יקל עלינו להבין את רעיון המימוש הפיזיקלי של מאפייני תוכנה אם, ראשית, נעמוד על-כך שאין פשר להיותו של מצב מסוים של המכונה הפיזיקלית מימוש של מצב זה או אחר בתכנית של מכונת טיורינג, כאשר הוא נדון בפני עצמו ובמנותק מיחסיו עם מצבים אחרים. זאת משום שזהותו של מצב בתכנית נקבעת על פי יחסיו עם מצבים אחרים. נפנה לדוגמה. נחשוב על תכנית חיבור המספרים הראשונה מבין השתיים שהצגנו לעיל; ונניח שאנו רוצים לדעת איזה מצב פיזיקלי יהווה מימוש של המצב הפנימי (מצב המכונה)  $S_1$ . ובכן, מה אנו יודעים על מצב מכונה זה? מה שאנו יודעים על מצב זה נתון לנו על ידי הכללים של התכנית הנדונה.

לכאורה, הכללים הרלוונטיים הם הכללים בהם מצוין מצב  $S_1$ , כלומר כללים (א), (ב) ו-(ג). אנו יכולים, אם כן, לאפיין את המצב הפנימי  $S_1$  בהתאם לכללים אלו, באופן הזה:

$S_1$  הוא המצב אשר כאשר המכונה נמצאת בו, אז –

אם הסורק/מדפס מזהה את הסימן "1" בתא הנסרק, הוא משאיר אותו, נע צעד אחד ימינה, ונשאר במצבו; ואם הסורק/מדפס מזהה את הסימן "+" בתא הנסרק, הוא מדפיס בתא "1" במקומו, נע צעד אחד ימינה, ונשאר במצבו; ואם הסורק/מדפס מזהה את הסימן "#" בתא הנסרק, הוא משאיר אותו, נע צעד אחד ימינה, ועובר למצב  $S_2$ .

זהו אפיון יחסותי של  $S_1$ , אפיונו במונחי קשריו עם פעולות ומצבים אחרים. אולם כיוון שכלל (ד) לא בא לידי ביטוי באפיון זה, זהו אפיון חלקי בלבד של  $S_1$ .

## שאלה

כיצד זה ש"היעדרו" של כלל (ד) מאפיון זה עושה אפיון זה לאפיון חלקי בלבד של  $S_1$ , כאשר כלל (ד) כלל אינו מתייחס ל- $S_1$ ?

ובכן, אפיון ממצה של  $S_1$  במונחי קשריו למצבים אחרים חייב להיות כרוך באפיון ממצה של המצבים האחרים הקשורים עמו. אולם המצבים האחרים המצוינים באפיון שהצגנו עתה מצוינים גם הם רק על ידי קשריהם שלהם למצבים אחרים; וכדי שאפיונם של המצבים האחרים הללו יהיה אפיון מלא, עליו לציין את כל קשריהם למצבים אחרים. אולם האפיון שהצגנו אינו עומד בתנאי זה, משום שחלק מקשריהם של המצבים האחרים הללו מצוינים על ידי כלל (ד), שאפיון זה לא התייחס אליו. אפיון ממצה של  $S_1$  חייב אפוא לכלול גם את הקשרים המפורטים על ידי כללים שאינם מזכירים את  $S_1$ ; במקרה הנוכחי זהו כלל (ד), המפרט את קשרי המצבים של המכונה כאשר היא נמצאת במצב  $S_2$ . כלומר, כדי שהאפיון הנ"ל של  $S_1$  יהיה ממצה עלינו להוסיף לו את התוספת הזו:

ו- $S_2$  הוא המצב אשר כאשר המכונה נמצאת בו, אז אם הסורק/מדפס מזהה את הסימן "1", הוא מדפיס "#". בתא במקומו, ועוצר את תנועתו.

אנו רואים שאפיון ממצה של מצב מכונה כלשהו חייב למעשה לפרט את כל קשרי המצבים הקבועים בתכנית – על פי כל הכללים של התכנית. נוכל להביע רעיון זה בדרך אחרת על ידי כך שנאמר כי אפיון המצבים של מכונת טיורינג הוא אפיון הוליסטי. אם כן, מצב של מכונת טיורינג מאופיין על ידי פירוט רשת הקשרים כולה הקבועה בתכנית, ו"תפקידו" ברשת קשרים זו, דהיינו לאילו מצבים עוברת המכונה כאשר היא נמצאת במצב זה ומתקיימים תנאים אחרים, ובהתקיים אילו תנאים ומאלה מצבים עוברת המכונה להיות במצב זה. נקודה זו אינה חלה רק על מצבים של מכונות טיורינג, אלא על מצבים של תוכנות, מצבים חישוביים, בכלל. כך,

מצב חישובי, מצב מסוים של תוכנה, מאופיין על ידי פירוט רשת הקשרים כולה הקבועה בתכנית, ו"תפקידו" או "מיקומו" ברשת קשרים זו.

ואם כן,

כשהפונקציונליסטים מאפיינים מצב נפשי באמצעות מיקומו ברשת הקשרים שבינו לבין קלטים חושיים, פלטים התנהגותיים, ומצבים נפשיים אחרים, הם מאפיינים אותו בתור המצב החישובי שהוא (מצב התוכנה שהוא).

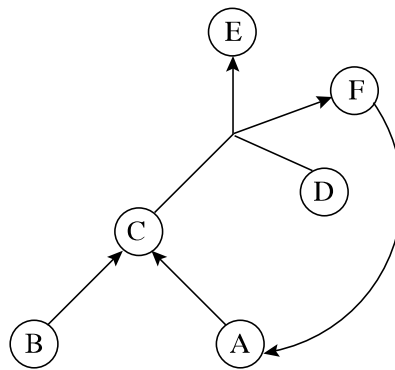
### 3.2.2.3 המימוש הפיזיקלי של מצבים חישוביים

עתה אנו יכולים לחזור לשאלת המימוש הפיזיקלי של מצבי מכונת טיורינג, ומצבי תוכנה בכלל. תכונותיו הפנימיות (הפיזיקליות) של המצב המממש אינן חשובות כלל. חשובות רק תכונותיו היחסיות, דהיינו היחסים שהוא מקיים עם המצבים המממשים את המצבים האחרים בתכנית של המכונה. כך, היותו של מצב פיזיקלי מסוים מימושו של מצב מסוים בתכנית מסוימת (של מכונת טיורינג או של כל מחשב שהוא) תלויה, ראשית, בהיותה של מערכת המצבים של המכונה הפיזיקלית מימושה של תכנית זו. מצב פיזיקלי מסוים של המכונה הממשית יהווה מימושו של מצב (מופשט) בתכנית אם, ורק אם, המצב הפיזיקלי מקיים עם מצבים אחרים של המכונה הפיזיקלית "מערכת קשרים זהה" לזו שמקיים מצב מופשט מסוים בתכנית עם מצבים מופשטים אחרים בתכנית.

מדוע מופיע הביטוי "מערכת קשרים זהה" במרכאות? משום שעולה הקושיה: **כיצד יכולה מערכת קשרים בין מצבים פיזיקליים ממשיים להיות זהה למערכת קשרים בין מצבים חישוביים מופשטים?** נוכל לפתור קושייה זו אם נשים לב שמתקיימים **מעברים** בין המצבים המופשטים של תכניות מכונת טיורינג. התכנית מורכבת מכללי מעבר – כללים הקובעים שהמכונה תעבור ממצבים מסוימים למצבים אחרים. והלא מעברים בין מצבים פיזיקליים, מצבים ממשיים, הם דבר שבשגרה: **קשרים סיבתיים** כרוכים במעברים כאלו. מערכת פיזיקלית שבה התרחשותו של מצב א **גורמת** את התרחשותו של מצב ב היא מערכת העוברת ממצב א למצב ב. תיאורה של מערכת של קשרים סיבתיים במונחים של מעברים בלבד אינו אלא הפשטה, התעלמות, מהמנגנון הסיבתי שאחראי למעברים. במילים אחרות, תיאור כזה הוא תיאור של אותם קשרים ברמה מופשטת יותר.<sup>38</sup>

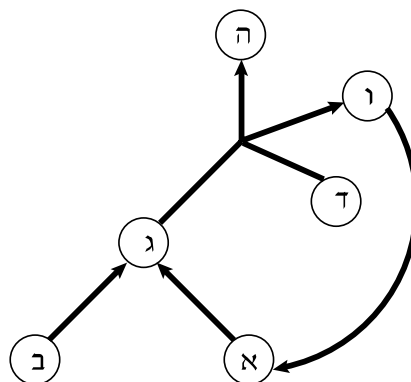
אם כן, מה שנדרש ממערכת פיזיקלית כדי שתהווה מימוש של מכונת טיורינג מסוימת (מכונת טיורינג בעלת תכנית מסוימת), או של תכנית מחשב כלשהי, הוא שיתקיימו בה הקשרים הסיבתיים ההולמים. מה פירוש "הולמים"? פירוש הדבר הוא שהם כרוכים באותם מעברים בין מצבים כמו מצבי המכונה המופשטים; פירוש הדבר הוא שרשת הקשרים הסיבתית ורשת הקשרים המופשטת, החישובית, תהיינה בעלות **אותו מבנה**. הרעיון של זהות במבנה בין רשת של קשרים סיבתיים בין מצבים ממשיים לבין רשת של קשרים בין מצבי-תוכנה מופשטים ניתן להדגמה פשוטה (התבוננו באיורים שלפניכם): נניח שתכנית חישוב מסוימת היא כזו, שהיא כרוכה במעבר ממצבים A ו-B למצב C, ובמעבר מצרף המצבים C ו-D לצמד המצבים E ו-F, ובמעבר מ-F ל-A (כל המצבים הללו הם מצבי תכנית מופשטים, אשר אפיונם אינו כרוך בשום אפיון פיזיקלי; הנתונים היחידים המאפיינים את המצבים הללו הם המעברים הללו).

<sup>38</sup> אין אנו נכנסים כאן לשאלה הניתוח הפילוסופי של סיבתיות: מהם התנאים המתקיימים כאשר התרחשות של מצב א גורמת את התרחשותו של מצב ב. זו שאלה פילוסופית חשובה ביותר, אולם העיסוק בה אינו נדרש למטרותנו הנוכחיות.



תרשים 12. רשת קשרים חישובית

רשת זו זהה במבנה שלה לרשת הקשרים של המכונה הפיזיקלית הבאה: התרחשותם של מצביה הפיזיקליים א ו-ב גורמת את התרחשותו של מצב פיזיקלי ג, אשר התרחשותו והתרחשותו של מצב נוסף ד גורמת את התרחשותם של המצבים ה ו-ו, והתרחשותו של ו גורמת את התרחשותו של מצב א (הפעם מדובר במצבים המאופיינים על פי תכונותיהם הפיזיקליות. לדוגמה, מצב א' עשוי להיות היותו של רכיב מסוים של המכונה בעל טמפרטורה מסוימת; מצב ב' עשוי להיות היותו של מצב מסוים של המכונה בעל מטען חשמלי מסוים, וכדומה).



תרשים 13. רשת קשרים סיבתית

מכונה זו מהווה מימוש פיזיקלי של תכנית זו. נתנו אפוא פשר לרעיון בדבר מימושה הפיזיקלי של תכנית.

ממתן פשר לרעיון בדבר מימושה הפיזיקלי של תכנית צריך לפסוע רק צעד קטן כדי לתת פשר גם לרעיון בדבר היותן של שתי מערכות פיזיקליות שונות מימושים של אותה תכנית. שתי מערכות פיזיקליות שלרשתות קשריהן הסיבתיים אותו מבנה, במובן שהסברנו עתה, מהוות

מימושים של אותה תכנית, אף אם מצביהם הפיזיקליים שונים אלו מאלו. מה שחשוב הוא לא טיבם הפיזיקלי של המצבים, אלא הקשרים ביניהם. מתן פשר להיותן של מערכות פיזיקליות מימושים של אותה תכנית חישוב הוא, למעשה, מתן פשר לאפשרות **ריבוי המימושים** של תכניות חישוב.

לאחר שנתנו פשר לרעיון בדבר מימושן הפיזיקלי של תכניות חישוב, אנו יכולים לתת פשר לרעיון בדבר **מימוש הפיזיקלי של מצב מסוים בתכנית**. מצב פיזיקלי מסוים של מערכת פיזיקלית מסוימת מממש מצב מופשט מסוים של תכנית חישוב – מצב חישובי מסוים – אם, ורק אם, (1) המערכת הפיזיקלית שאלה הוא שייך מהווה מימוש של תכנית החישוב שאלה שייך המצב המופשט, כלומר למערכת הפיזיקלית רשת קשרים סיבתית בעלת מבנה זהה לרשת הקשרים של התכנית; ו-(2) המצב הפיזיקלי תופס ברשת הקשרים הסיבתית את אותו מקום שתופס המצב המופשט ברשת הקשרים המופשטת, החישובית. בדוגמה דלעיל, מצב A מממש את המצב החישובי A, מצב B מממש את המצב החישובי B, וכדומה.

מאפיון זה נגזר האפיון של היותם של שני מצבים פיזיקליים **מימושים של אותו מצב חישובי**: שני מצבים פיזיקליים מממשים אותו מצב חישובי אם, ורק אם, (1) רשתות הקשרים הסיבתיים של המערכות הפיזיקליות להן שייך כל אחד מהם זהות במבניהן; ו-(2) הם תופסים מקומות זהים ברשתות הללו.

שימו לב שתנאי (1) של אפיון זה אינו דורש שהמערכות הפיזיקליות המדוברות תהיינה זהות בתכונות הפיזיקליות שלהן: אפשרי ריבוי מימושים של תכניות וכך אפשרי גם ריבוי מימושים של מצבים חישוביים. אולם תנאי זה משמעו ששני מצבים פיזיקליים עשויים להיות מימושים של אותו מצב חישובי רק אם המערכות הפיזיקליות שלהם **מממשות את אותה תכנית**. כלומר, אין משמעות לזהות חישובית של שני מצבים פיזיקליים אלא במסגרת תכנית נתונה. מצבים פיזיקליים זהים מבחינה חישובית – מצבים פיזיקליים המהווים מימושים של אותו מצב חישובי – הם מצבים הממלאים את אותו תפקיד ברשת הקשרים שבה הם כרוכים; אלו הם מצבים הקשורים באותם קשרים למצבים אחרים; פירושו של דבר, שרשתות הקשרים שלהם זהות במבניהן, והן תופסות את אותו מקום ברשתות הללו. אין שחר לדבר על "אותו מקום" או "אותו תפקיד" ברשת אלא אם כן הרשתות זהות במבניהן. לדרישה זו חשיבות באשר לפונקציונליזם ומודל המחשב של הנפש, כפי שנראה בהמשך.

נסכם את הרעיונות שאלהם הגענו:

(א<sub>1</sub>) **מערכת פיזיקלית מהווה מימוש של תכנית מסוימת** (מכונת טיורינג או תכנית אחרת) אם, ורק אם, הקשרים הסיבתיים שבין מצביה כרוכים באותם מעברים בין מצבים כמו מצבי המכונה המופשטים; במילים אחרות, אם רשת הקשרים הסיבתית של המערכת הפיזיקלית ורשת הקשרים המופשטת, החישובית, של התכנית, הן בעלות **אותו מבנה**.

(א<sub>2</sub>) שתי מערכות פיזיקליות שונות מהוות מימושים של אותה תכנית אם, ורק אם, הרשתות של קשריהן הסיבתיים הן בעלות אותו מבנה.

(ב<sub>1</sub>) **מצב פיזיקלי מסוים של מערכת פיזיקלית מסוימת מממש מצב מסוים של תכנית מסוימת (מצב חישובי), אם, ורק אם, (1) המערכת הפיזיקלית שאליה הוא שייך מהווה מימוש של תכנית החישוב אליה שייך המצב המופשט, כלומר למערכת הפיזיקלית רשת קשרים סיבתית בעלת מבנה זהה לרשת הקשרים של התכנית; ו-(2) המצב הפיזיקלי תופס ברשת הקשרים הסיבתית את אותו מקום שתופס המצב המופשט ברשת הקשרים המופשטת של התכנית.**

(ב<sub>2</sub>) **שני מצבים פיזיקליים מממשים אותו מצב חישובי, אם, ורק אם, (1) רשתות הקשרים הסיבתיים של המערכות הפיזיקליות שלהן שייך כל אחד מהם זהות במבניהן; ו-(2) הם תופסים מקומות זהים ברשתות הללו.**

הדיון ברעיון של מכונת טיורינג ובהשלכותיו השונות הוליד אותנו לרעיונות אלו, אך כפי שניתן לראות, הם חלים על תוכנות ומימושיהן הפיזיקליים באשר הם, ולא רק על מכונת טיורינג. למעשה, הם חלים גם על מימושים **שאינם פיזיקליים** של תוכנות (נקודה זו תהיה חשובה בעיני מי שדוגלים בגרסה דואליסטית של פונקציונליזם). ניתן לפיכך להכליל את הרעיון האחרון:

(ב<sub>2</sub>\*) **שני מצבים מממשים את אותו מצב חישובי, אם, ורק אם, (1) רשתות הקשרים הסיבתיים של המערכות שלהן שייך כל אחד מהם זהות במבניהן; ו-(2) הם תופסים מקומות זהים ברשתות הללו.**

רעיונות אלו מקנים תוכן ממשי לטענות המרכזיות של מודל המחשב של הנפש וכך לעצם ההשקפה הפונקציונליסטית. עתה אפשר להבין באופן הולם יותר את הרעיון שהנפש היא תוכנת הגוף, ואת קריטריון האינדיווידואציה הפונקציונליסטי. לפי הפונקציונליזם, סוגו הנפשי של מצב מכונן על ידי הפונקציה שלו בחיים הפסיכולוגיים של הפרט, וכך שני מצבים נפשיים הם זהים בסוגם אם הם מממשים את אותה פונקציה. מימוש אותה פונקציה משמע מימוש אותו מצב חישובי, ועתה ראינו את פשרו המדויק של הרעיון בדבר מימוש אותו מצב חישובי. בסעיף הבא נראה כיצד הרעיונות שנדונו לעיל מאפשרים לנו לספק מענה לתמיהות שהועלו בשלב קודם.

### 3.2.3 זהות בקשרים סיבתיים, ומענה לתמיהות

כאשר מדובר במימושים (פיזיקליים) של תכניות, המעברים בין המצבים מתגלמים, כאמור, בקשרים סיבתיים. שתי מערכות פיזיקליות שונות מהוות מימושים של אותה תכנית אם, ורק אם, הרשתות של קשריהן הסיבתיים הן בעלות אותו מבנה. לעתים אומרים שבשתי מערכות כאלו – שתי מערכות הנבדלות זו מזו מבחינה פיזיקלית, אך שרשתות הקשרים הסיבתיים

שלהן זהות במבנה (ושלפי הקריטריון האמור מממשות את אותה תכנית) **מתקיימים אותם קשרים סיבתיים**. כיוון שהרשתות זהות במבניהן, כל מצב מן המצבים של אחת המערכות זהה מבחינה חישובית למצב מסוים של המערכת האחרת, כלומר שני מצבים כאלו "תופסים אותו מקום" ברשתות. שתי רשתות הקשרים הסיבתיים הן אפוא רשתות הקושרות ביניהם מצבים זהים זה לזה לפי קריטריון מסוים – קריטריון חישובי – ובמובן זה (שנכנה אותו "המובן החלש של זהות סיבתית") ניתן לומר שבשתי המערכות ישנם אותם קשרים סיבתיים.

אבל במובן אחר, אין מתקיימים במערכות כאלו אותם קשרים סיבתיים. במובן חשוב, זהות בקשרים סיבתיים בין שתי מערכות שונות מצבים מחייבת זהות **בכוחות הסיבתיים** של המצבים בשתי המערכות הללו. אולם קשה לומר שלשני מצבים פיזיקליים השונים זה מזה באופן ניכר מבחינה פיזיקלית – כמו מצב מסוים של מוח אנושי ומצב מסוים של מוח סיליקון – יש אותם כוחות סיבתיים, שהרי (לפחות על פי רוב) מצבים כאלו יגרמו מצבים השונים זה מזה מבחינה פיזיקלית באופן ניכר (למשל, כשמצב מסוים של המוח האנושי גורם למצב אחר של מוח זה, ודבר מקביל מתרחש במוח הסיליקון). במילים אחרות, לשני מצבים יש את אותם כוחות סיבתיים זהים רק אם המצבים הפיזיקליים שהם גורמים זהים זה לזה מבחינה פיזיקלית, ובמקרה של המוח האנושי ומוח הסיליקון המצבים הנגרמים הם שונים מאוד מבחינה פיזיקלית. במקרה כזה מתקיימת זהות בתכונות מופשטות בלבד בין המצבים הנגרמים (תפיסת אותו מקום ברשת). במובן חשוב, כוחות סיבתיים הם מאפיינים ממשיים ו"זהות בכוחות סיבתיים" משמעה היכולת לגרום למצבים זהים במאפייניהם הממשיים (ואם מדובר במערכות פיזיקליות, חייבת להיות זהות במאפיינים פיזיקליים בין המצבים הנגרמים). אין די בזהות במאפיינים מופשטים-חישוביים בלבד. במובן שעליו אנו מדברים עתה (שנכנה אותו "המובן החזק של זהות סיבתית"), הקשרים הסיבתיים בשתי המערכות הנדונות **אינם** זהים אלו לאלו.

נקודה זו קשורה הדוקות להיותם של מאפייני תוכנה מאפיינים מופשטים, ולהיותם של המושגים של **מצב בתכנית**, או **תפיסת אותו מצב בתכנית**, או **זהות חישובית בין מצבים**, וכדומה, מושגים מופשטים. כשאנו מאפיינים מצב כמצב חישובי מסוים, אנו מאפיינים אותו רק במונחי תכנית שאליה הוא שייך ומיקומו בתכנית זו; דהיינו, אנו מאפיינים אותו במונחי קשריו עם מצבים אחרים, שגם הם מאופיינים רק במונחי קשריהם למצבים אחרים. אנו מאפיינים את המצב באמצעות מיקומו במערכת קשרים **מופשטת** – אנו "מפשטים" אותו מתכונות פיזיקליות או תכונות פנימיות (תכונות שיש לו ללא תלות בקשריו עם מצבים אחרים) כלשהן; ואנו מפשיטים אותו גם מקשרים סיבתיים, במובן החזק של הביטוי "קשרים סיבתיים", מובן על פיו זהותו של קשר סיבתי נקבעת על פי התכונות הפנימיות של המצבים אותם הוא קושר (כשמדובר במצבים פיזיקליים, אלו הן תכונות פיזיקליות).

כזכור, בשלב מוקדם העלנו את התמיהה ("תמיהה א"), כיצד ניתן ליישב את רעיון **מופשטותם של המאפיינים הנפשיים** על פי מודל המחשב של הנפש – המהווה חלק מהותי של ההשקפה הפונקציונליסטית, עם הרעיון הפונקציונליסטי בדבר היקבעותם של המאפיינים

הנפשיים על ידי **קשרים סיבתיים**, שהם **מאפיינים ממשיים** של מצבים. מה למאפייני תוכנה מופשטים ולכוחות סיבתיים?

העלנו תמיהה נוספת באותו הקשר ("תמיהה ב'): כיצד עשויים כפילים נפשיים השונים במבניהם הפיזיקליים להיות שקולים בכוחותיהם הסיבתיים? הצגנו את המענה לתמיהה זו שמספק מודל המחשב של הנפש, לפיו מבנים פיזיקליים שונים עשויים להיות בעלי כוחות סיבתיים זהים אם הם מימושים של אותן תוכנות. אולם, שאלנו, כיצד יכולה שקילות במאפיינים מופשטים כלשהם להסביר שקילות בכוחות סיבתיים? מאפיינים מופשטים נראים חסרי כל רלוונטיות לכוחות סיבתיים ממשיים של אובייקטים, ולא ברור אפוא כיצד יוכלו להסבירם. כך, גם תמיהה ב' העלתה את השאלה הכללית: מה למאפייני תוכנה מופשטים ולכוחות סיבתיים?

עתה ביכולתנו להסיר תמיהות אלו מדרכנו. בשתי מערכות פיזיקליות המממשות את אותה תכנית והנבדלות זו מזו מבחינה פיזיקלית מתקיימים אותם קשרים סיבתיים רק במובן החלש: המערכות הללו זהות **במבנה** הרשתות הסיבתיות שלהן. זהות מבנית מעין זו, כפי שראינו, היא זהות במאפיינים מופשטים בלבד; שהרי זהות מבנית כזו מתקיימת בין מערכת פיזיקלית (כמו מחשב) **לבין תכנית המחשב המופשטת** שהיא מממשת – כשמחשב מממש תכנית מסוימת, רשת הקשרים הסיבתית של פעולתו זהה במבנה שלה לרשת הקשרים המופשטת של התכנית. וכך, כפילים נפשיים הנבדלים זה מזה מבחינה פיזיקלית שקולים בכוחותיהם הסיבתיים רק במובן מוגבל וחלש זה – במובן זה שרשתות הקשרים הסיבתיים של פעולותיהם זהות במבניהן, דהיינו במאפיינים מופשטים שלהן. אם כך, אין קושי ברעיון של הסבר שקילות סיבתית באמצעות הנחת שקילות במאפיינים מופשטים, משום שהשקילות הסיבתית בה מדובר היא שקילות במאפיינים מופשטים בלבד.

כך ניתן גם ליישב את רעיון מופשטותם של המאפיינים הנפשיים על פי מודל המחשב של הנפש, עם הרעיון הפונקציונליסטי בדבר היקבעות זהותם של מצבים נפשיים על ידי קשרים סיבתיים בינם לבין מצבים נפשיים אחרים, קלטים ופלטים. את הרעיון הפונקציונליסטי הזה יש לסייג: במובן חשוב – במובן החזק של "קשרים סיבתיים" – אין זה נכון לומר שזהותם של מצבים נפשיים נקבעת לפי רעיון זה על ידי קשרים סיבתיים. הדבר נכון רק במובן החלש של "קשרים סיבתיים"; כלומר, על פי הרעיון הפונקציונליסטי, זהותו של מצב נפשי נקבעת על פי רשת הקשרים המופשטת המקבילה – דהיינו רשת הקשרים בעלת המבנה הזהה – לרשת הקשרים הסיבתית שבה הוא כרוך (לרשת הקשרים הסיבתיים בינו לבין מצבים נפשיים אחרים, גירויים חושיים ותגובות התנהגותיות). פירוש הדבר, למעשה, שזהותו של מצב נפשי נקבעת על פי מיקומו בתכנית אותה מממשת פעילותו הנפשית של בעליו. וכך, **הרעיון הפונקציונליסטי הבסיסי, על פי פירושו זה, שקול לרעיון של מודל המחשב של הנפש**. ניתן לומר, שמודל המחשב של הנפש מעניק פשר לקביעה הפונקציונליסטית הבסיסית, המנוסחת במונחי קשרים סיבתיים. עבור כל מי שמקבלים את רעיון ריבוי-המימושים של היסוד הנפשי.



סביר להניח שמרבית הפילוסופים הפונקציונליסטים אמנם מתכוונים לפשר זה של הקביעה שזהותם של מצבים נפשיים נקבעת על ידי קשריהם הסיבתיים. זאת משום שמרבית הפילוסופים הפונקציונליסטים מקבלים את תופעת ריבוי-המימושים של המצבים הנפשיים (אצל רבים מהם, ההתרשמות מתופעה זו הייתה מניע חזק בזכות הרעיון הפונקציונליסטי), ותופעת ריבוי המימושים אינה מתיישבת עם הפשר החזק יותר לטענה זו – עם הפשר על פיו שני מצבים זהים זה לזה בקשריהם הסיבתיים, אם המצבים שעמם קשור כל אחד מהם זהים זה לזה בתכונותיהם הפנימיות (תכונות פיזיקליות – במקרה בו מדובר במצבים פיזיקליים). לשם נוחות, נמשיך לדבר על ההשקפה הפונקציונליסטית כהשקפה על פיה זהותם של מצבים נפשיים נקבעת על פי קשריהם הסיבתיים עם מצבים אחרים (מצבים נפשיים אחרים, קלטים חושיים ופלטת התנהגותיים), מבלי שנטרח ונבהיר בכל פעם שאת הביטוי "קשרים סיבתיים" יש להבין בהקשר זה באופן חלש יותר, כפי שראינו.

על פי ההשקפה הפונקציונליסטית, שני מצבים נפשיים שונים שייכים לאותו סוג נפשי אם, ורק אם, מתקיימת זהות בין הקשרים הסיבתיים שכל אחד מהם מקיים עם מצבים נפשיים אחרים, קלט חושי ופלט התנהגותי, ולו **במובן החלש** ("המופשט") של "זהות בקשרים סיבתיים".

### 3.2.4 הנפש כמכונת טיורינג

הדיון במכונת טיורינג שימש אותנו כדי להבהיר מושגים אחדים בהם כרוך מודל המחשב של הנפש, ומאפיינים אחדים של הגישה הפונקציונליסטית. אולם לדיון זה חשיבות נוספת לעניינינו. הרעיון שהנפש היא תוכנת מחשב – שהיא, במקרה שלנו ושל יצורים נפשיים רבים אחרים, התוכנה שבהתאם לה פועל המוח – לבש בתחילת דרכו של הפונקציונליזם את הצורה הספציפית של ראיית הנפש כמכונת טיורינג, וראיית מצבינו הנפשיים כמצבי מכונת טיורינג. מי שהציע את המודל הפונקציונליסטי של מכונת טיורינג כמודל לנפש היה הילארי פטנאם.<sup>39</sup> הרעיון בראשיתו היה, שיצור נפשי הוא יצור המממש מכונת טיורינג מסוימת, מכונת טיורינג בעלת רמת מורכבות גבוהה ביותר, בעלת מספר מצבי מכונה עצום, המגיבה למגוון רחב של קלטים (גירויים חושיים) והמנפיקה מגוון רחב של פלטים (התנהגויות). ניתן לתאר את

<sup>39</sup> במאמרו "Minds and Machines" מ-1960, ו-"Robots: Machines or Artificially Created Life" מ-1964 התווה פטנאם אנלוגיה בין הנפש לבין מכונת טיורינג. במאמרו "The Mental Life of some Machines" מ-1967 העלה פטנאם את הטענה החזקה יותר שעל פיה מצבים נפשיים הם מצבים חישוביים של מכונת טיורינג (שלושת המאמרים הללו פורסמו אחר-כך בקובץ מאמרו של פטנאם: *Mind, Language, and Reality: Philosophical Papers, Vol. 2*, Cambridge: Cambridge University Press, 1975.)

ראו סקירה על התפתחות מחשבתו של פטנאם באשר למודל המחשב של הנפש במאמרו של אורון שגריר: O. Shagrir, "The Rise and Fall of Computational Functionalism", in Y. Ben-Menahem (ed.), *Hilary Putnam*. Cambridge: Cambridge University Press, 2005.

ייתכן מאוד שטיורינג עצמו חשב מחשבות דומות, אולם כשהציג את הרעיון של מכונת טיורינג לא דן במפורש בהשלכותיו באשר לפילוסופיה של הנפש. לרעיון אחר של טיורינג – הרעיון של "מבחן טיורינג" – תפקיד ישיר יותר בפילוסופיה הנפש, ונדון בו בהמשך.

היחסים המורכבים שבהם כרוכים מצבינו הנפשיים – בינם לבין עצמם ובינם לבין קלטים חושיים ופלטטים התנהגותיים – במונחים של תכנית של מכונת טיורינג. מצבים נפשיים מסוימים – כאב, ההאמנה שיורד גשם, וכדומה – הם מימושים של מצבים פנימיים מסוימים של מכונת טיורינג.

אם כן, לפי רעיון זה, יצורים נפשיים מממשים מכונת טיורינג – ניתן לתאר את תפקודם הנפשי על פי כלליה של תכנית מכונת טיורינג מורכבת מאוד. ועובדה זו – העובדה שהם מממשים תכנית זו (ולא, למשל, העובדה שהם **בעלי מבנה חומרי מסוים** המממש אותה) – היא שעושה אותם ליצורים נפשיים. זאת באשר להיותו של יצור נפשי. ובאשר למצבים נפשיים ספציפיים, להימצא במצב נפשי מסוים משמע להימצא במצב המממש מצב מכונה (מצב פנימי) מסוים של מכונת טיורינג. את המצבים הללו ניתן לאפיין באמצעות כללי המעבר החלים עליהם (כשם שאפינו בשלב קודם את המצב  $S_1$  של מכונת חיבור המספרים). מכונת טיורינג הרלוונטית במקרה זה היא מכונת טיורינג מורכבת מאוד, ומן הסתם היא מאופיינת על ידי כללים רבים מאוד.

למען הדיוק, עלינו להזכיר שפטנאם טען כי הנפש אינה מכונת טיורינג רגילה אלא **"מכונת טיורינג הסתברותית"**. הכללים השולטים במצבי מכונת טיורינג (רגילה) הם כללים **דטרמיניסטיים** – דהיינו, כללים המכתיבים לחלוטין לאלו מצבים תעבור המכונה בכל שלב ושלב של פעולתה – ואילו הכללים השולטים במצבינו הנפשיים הם כללים **הסתברותיים** בלבד: הם קובעים בכל שלב שהמכונה תעבור למצבים מסוימים באחוז מסוים של המקרים, ולמצבים אחרים באחוז מסוים של המקרים. אנו, כפי שטען פטנאם, הננו "אוטומט הסתברותי" ("probabilistic automaton"). טענה זו, כך חשב, משקפת באופן נאמן יותר את מציאות חיינו הנפשיים.

קיימת אנלוגיה בין אפיונם של מצבים נפשיים על פי הרעיון הפונקציונליסטי לבין אפיונם של מצבים פנימיים של מכונת טיורינג: כשם שעל פי הפונקציונליסטים, מה שמקנה למצב נפשי את זהותו הנפשית הספציפית הוא התפקיד הסיבתי שהוא ממלא ביחס לגירויים חושיים, תגובות התנהגותיות, ומצבים נפשיים אחרים, כך, מה שמקנה למצב פנימי של מכונת טיורינג את זהותו אלו הם קשריו לנתונים שהמכונה קוראת, לנתונים שהיא מדפיסה, ולמצבים פנימיים אחרים. הפנייה למצבים פנימיים אחרים היא משמעותית: כשם שלא ניתן לאפיין האמנה, למשל, במונחים של התנהגות או דגמי התנהגות בלבד, מבלי להתייחס לרציה (זוהי, כאמור ביחידה 3 [סעיף 4.3], נקודה שבה האפיון הביהביוריסטי של היסוד הנפשי נראה לוקה בחסר), כך, לפחות על פי רוב, אי אפשר לאפיין מצב פנימי של מכונת טיורינג במונחי הפלט שהמכונה מפיקה כאשר היא נמצאת במצב זה בהינתן קלט מסוים, מבלי להתייחס למצבי מכונה אחרים (ראינו לעיל, כי אפיון ממצה של  $S_1$  בתכנית חיבור המספרים חייב לכלול התייחסות לכל כללי התכנית – גם לכללים המפרטים את קשרי המצבים של המכונה כאשר היא נמצאת במצבים פנימיים אחרים). העמידה על אנלוגיה זו בין מצביה של מכונת טיורינג לבין מצבים נפשיים, בצד אנלוגיות אחרות – למשל, אפשרות ריבוי המימושים החלה על

מצבי מכונת טיורינג ובעיני רבים גם על מצבים נפשיים – הולכה לרעיון של מכונת טיורינג כמודל של הנפש.<sup>40</sup> רעיון זה יוצק תוכן ממשי ברעיון הפונקציונליסטי: כשמנהירים את המושג של אפיון פונקציונלי של מערכת או מצב במונחי מכונת טיורינג, נראה שזהו מושג ברור, שניתן לתת לו אפיון לוגי-מתמטי מדויק.

הרעיון של הנפש כמכונת טיורינג אין משמעו שאנו מבצעים פעולות חיבור, למשל, בדרך שבה מכונת טיורינג מבצעת חיבור. במילים אחרות, הרעיון לא היה שהנפש היא תכנית הדומה לתכניות שאפיינו לעיל, או שמישהו אפיון אי-פעם. הנפש, כאמור, היא תכנית מורכבת לאין שיעור. מורכבות זו באה לידי ביטוי בכך שהיא מגיבה לטווח רחב מאוד של קלטים (כדוגמה, חשבו על מספר החושים שלנו ועל מגוון הגירויים הרחב, וכך מגוון המידע הרחב, שכל-אחד מהם מספק), בכך שטווח הפלטים – ההתנהגויות – שהיא עשויה להפיק, אפילו בתגובה לקלט מסוים, הוא עצום, ובכך שמספר סוגי המצבים הנפשיים הוא עצום. אין אפוא ספק, שתכנית מכונת טיורינג של הנפש תצטרך לכלול מספר רב מאוד של כללים, ולא תדמה בפרטיה לכל תכנית חישוב מוכרת. אך הרעיון הוא שתכנית כלשהי של מכונת טיורינג מאפיינת את חיינו הנפשיים. באופן עקרוני, חיינו הנפשיים פועלים כמכונת טיורינג, "מכונת טיורינג רב תכליתית" ("A general purpose Turing Machine"). במינוח שבו השתמשנו קודם, על פי רעיון זה הנפש היא **מכונת טיורינג אוניברסלית**.

אולם דומה שאנו שונים מתכניות מכונת הטיורינג שהצגנו לא רק ברמת המורכבות. אחד הקשיים שהתגלו בראיית הנפש כמכונת טיורינג עניינו דווקא בקשר שבין מצב נפשי לבין מצבים אחרים, קשר שכפי שראינו לפני פסקות אחדות, יש לו מקבילה במכונת טיורינג. לצד מקבילה זו, נראה שמצוי כאן גם הבדל משמעותי: במקרה הנפשי, האמנה ורצייה עשויות לחבור יחד כדי להוליך להתנהגות, אולם במקרה של מכונת טיורינג הסיטואציה המקבילה אינה אפשרית.

## שאלה

התוכלו לחשוב מדוע לא ייתכן ששני מצבים פנימיים של מכונת טיורינג יחברו יחדיו כדי להוליך לפלט מסוים?

ובכן, אין זה אפשרי משום **שמכונת טיורינג מאופיינת על ידי מצב-פנימי אחד בלבד בכל שלב**. ניתן היה אולי לחשוב שהנפש היא אוסף של מכונות טיורינג – של תכניות מכונת טיורינג שונות – ומצבים נפשיים שונים המתרחשים בו-זמנית הם מצבים פנימיים של המכונות השונות – אולם רעיון זה אינו יכול להסביר כיצד שני מצבים נפשיים, כמו האמנה ורצייה, **משתפים פעולה** באותו תהליך סיבתי. זו דוגמה אחת לבעייתיות בראיית מצבים נפשיים כמצבים פנימיים של מכונת-טיורינג.

<sup>40</sup> פטנאם אף טען (במאמרו הנזכר "Minds and Machines") שגם את האסימטריה בין ידע של גוף שלישי לידע של גוף ראשון (ראו יחידה 1) ניתן להסביר באמצעות המודל של מכונת טיורינג. לא נוכל לדון כאן ברעיון זה.

ואמנם, מלכתחילה הכירו הדוגלים ברעיון שהנפש היא מכונת טיורינג בעלת רמת מורכבות הולמת כי מכונת טיורינג, בדיוק כפי שאופיינה במקור, אינה יכולה לשמש כמודל לנפש, וכי יש לתקן את אפיוניה בדרכים שונות. הרעיון של מכונת טיורינג **הסתברותית** שהזכרנו לעיל הוא דרך תיקון אחת; הרעיון של הפרדת אמצעי הקלט-פלט (הסרט, בתיאור המסורתי) לשני אמצעים – אחד לקלט ואחד לפלט – הוא דרך תיקון אחרת; הרעיון של ראיית הנפש כמורכבת ממכונות טיורינג רבות הקשורות אלו באלו הוא דרך תיקון נוספת, והוצעו תיקונים נוספים. לא ניכנס כאן לפרטים. ספק אם ניתן להתגבר על הקושי שצוין לעיל, או על קשיים נוספים, באמצעות תיקונים כלשהם אשר יישמרו מובן משמעותי של הרעיון שהנפש היא מכונת טיורינג, מובן שיבדיל את הרעיון הזה מהרעיון הפונקציונליסטי הכללי יותר שהנפש היא תוכנה (כלשהי) של מחשב.

קשיים נוספים שדומה כי ניצבים בפני הרעיון שהנפש היא מכונת טיורינג עניינם בהסבר **זיכרון ויכולת למידה**: אין זה ברור כיצד יכולה מכונת טיורינג להשתמש במידע שקלטה בעבר או בתוצאות חישובים שכבר ביצעה. אחת הבעיות בהקשר זה היא זו: האופן שבו נתונים כלשהם עשויים להישמר במכונת טיורינג הוא על ידי השארת סימנים על-גבי הסרט – **הסרט הוא אמצעי הזיכרון האפשרי היחיד**. ובכן, נניח שאכן תתאפשר שמירת נתונים מפעולות קודמות של המכונה על-גבי הסרט, ובמובן זה הנתונים "יזכרו" (כפי שראינו, מכונת טיורינג מוחקת את סימני הקלט וכותבת במקומם את סימני הפלט – דהיינו היא אינה שומרת את נתוני הקלט; אולם דומה שאין בעיה עקרונית בתכנון מכונה שלא תמחק את סימני הקלט). עדיין, לא ברור איך תשתמש המכונה בנתונים שהיא "זוכרת". שכן, כיצד תדע המכונה לאלו נתוני זיכרון לפנות כדי להשתמש בהם בעת ביצוע משימת-חישוב מסוימת? כיצד תדע תוכנו של איזה תא רלוונטי למשימתה? אין זה ברור אילו כללים יוכלו להפנות את המכונה אל הנתונים הרלוונטיים.<sup>41</sup>

הרעיון שהנפש היא מכונת טיורינג אמנם נחשב כבעייתי, עם זאת חשיבותו רבה, שכן הוא היווה דוגמה לרעיון הפונקציונליסטי בדבר מודל המחשב של הנפש, רעיון שאותו אמצו פילוסופים רבים.<sup>42</sup> אנו מקווים שהדיון במכונת טיורינג תרם להבהרת הבטים שונים של רעיון

<sup>41</sup> בעיה ספציפית זו היא דוגמה של בעיה כללית ועקרונית, "בעיית הרלוונטיות", המעסיקה רבות חוקרים במדעים הקוגניטיביים ובבינה מלאכותית: כיצד מתבצעת ברירת פריטי המידע הרלוונטיים לביצוע משימות קוגניטיביות מתוך המספר העצום של פריטי המידע שברשותנו? אלו כללים עשויים בכלל להנחות את בחירת המידע הרלוונטי לביצוע משימה קוגניטיבית מסוימת? הדיון בבעיה זו חורג מנושאי קורס זה.

<sup>42</sup> במאמרו "Philosophy and Our Mental Life" בתוך: *Mind, Language, and Reality: Philosophical Papers*, Vol. 2, Cambridge: Cambridge University Press, 1975, יצא פטנאם עצמו כנגד הרעיון של הנפש כמכונת טיורינג, אולם עדיין החזיק בהשקפה הפונקציונליסטית. כאמור, בשלב מאוחר יותר נטש פטנאם אף

זה, ובמיוחד להבהרתה של הטענה שהיחס בין הנפש לגוף הוא כיחס בין תוכנת מחשב לחומרתו, ולקשר בינה לבין עצם הרעיון הפונקציונליסטי. רבים רואים ברעיון זה את אחד הרעיונות החשובים והמעניינים שהוצעו אי-פעם בפילוסופיה של הנפש.

מודל המחשב של הנפש אינו רק השקפה פילוסופית. למעשה, מודל זה, בצורה זו או אחרת שלו, הוא המודל העומד ביסודו של שדה המחקר המקיף של הנפש בעשרות השנים האחרונות, שדה מחקר הקרוי "המדע הקוגניטיבי".<sup>43</sup> כשם שהביהיוריזם הפילוסופי והביהיוריזם הפסיכולוגי צעדו זה בצד זה, כך צעדו זה בצד זה הפונקציונליזם והמדע הקוגניטיבי. שניהם, בהתאמה, דחו את הרעיונות של הביהיוריזם הפילוסופי והביהיוריזם הפסיכולוגי: הפונקציונליזם התבסס על ההכרה שלא ניתן לאפיין מצבים נפשיים במונחי קלט ופלט בלבד, ובמקביל התבסס המדע הקוגניטיבי על ההכרה שבשל מורכבותה של החוקיות הפסיכולוגית המאפיינת את התנהגויותיהם של בני-אדם ויצורים אחרים, הפסיכולוגיה שלהם אינה יכולה להילמד על ידי התחקות אחר דגמי קלט ופלט (גירוי ותגובה).<sup>44</sup> המדע הקוגניטיבי מקנה אפוא מקום מרכזי בחקר הנפש למצבים פנימיים ולחוקיות השוררת הן ביניהם, והן ביניהם לבין קלטים חושיים ופלטם התנהגותיים. אל המצבים הפנימיים מתיחסים במדע הקוגניטיבי בראש ובראשונה כאל מצבים חישוביים, מצבים של תוכנת מחשב. כלומר, חוקי הפסיכולוגיה בהם עניינו של המדע הקוגניטיבי הם כאלו, שהמצבים הפנימיים כרוכים בהם לא בשל תכונותיהם הפיזיקליות, אלא בשל תכונותיהן המופשטות-החישוביות, דהיינו מיקומם ברשתות הקשרים המכוננות את החיים הנפשיים.<sup>45</sup>

השקפה זו. הביקורות שהעלה כנגדה מופיעות, בין היתר, בספרו בשני המאמרים הללו: *Representation and Reality*, Cambridge (Mass.): Cambridge University Press, 1988 ו-"Why Functionalism didn't Work", in his *Words and Life: Hilary Putnam* (edited by Conant), Cambridge (Mass.): Harvard University Press, 1994; "Functionalism: Cognitive Science or Science Fiction", in *The Future of the Cognitive Revolution*, edited by D. Martel, Johnson C. Erneling and Erneling, Oxford: Oxford University Press, 1997.

<sup>43</sup> המדע הקוגניטיבי הוא שדה מחקר רב-תחומי. תחומי המחקר העיקריים הנוטלים בו חלק הם פסיכולוגיה קוגניטיבית, מדעי המוח, בלשנות, פילוסופיה, ובינה מלאכותית.

<sup>44</sup> ביחידה 3 בסעיף 4.4, הזכרנו בהקשר זה את ביקורתו של הבלשן חומסקי על הביהיוריזם הפסיכולוגי של סקינר. ביקורת זו היוותה תמריץ חשוב לפיתוח המדע הקוגניטיבי.

<sup>45</sup> לתכונותיהן הפיזיקליות של מערכות נפשיות יש בהחלט תפקיד חשוב במדע הקוגניטיבי. אחד הטעמים לכך הוא שזכור התכונות הפיזיקליות של מערכת רלוונטיות לשאלה אילו מצבים נפשיים ממומשים בה, וקיימים טעמים נוספים. אולם עיקר עניינו של המדע הקוגניטיבי הוא בהסבר חיי הנפש באמצעות אפיונים חישוביים.

## 4. שתי גרסאות של פונקציונליזם

הרעיון העומד בבסיסה של ההשקפה הפונקציונליסטית הוא שיצור נפשי מאופיין באמצעות רשת מסוג מסוים של קשרים, ומצבים נפשיים מאופיינים על ידי מיקומם, או תפקידם, ברשת מסוג מסוים של קשרים – רשת של קשרים השוררים בין המצבים הנפשיים עצמם, ביניהם לבין קלטים חושיים, וביניהם לבין פלטים התנהגותיים. כפי שצינו, **הפסיכולוגיה** היא שמספקת את תיאורה של רשת הקשרים הזו. תיאור הקשרים בין מצבים נפשיים, קלטים חושיים ופלטים התנהגותיים הוא, למעשה, תיאורן של הסדירויות שעל פיהן מתנהלים המעברים בין מצבים נפשיים מסוימים (ולעתים, קלטים חושיים מסוימים) לבין מצבים נפשיים אחרים (ולעתים, פלטים התנהגותיים מסוימים); והתיאוריה הפסיכולוגית היא זו המתארת את הסדירויות הללו. לפי ההשקפה הפונקציונליסטית, אם כן, מצבים נפשיים מאופיינים על פי תפקידיהם בחייו הנפשיים של היצור הנפשי **כפי שתפקיד זה מתואר על ידי התיאוריה הפסיכולוגית החלה עליו**.

כזכור, את התיאוריה הפסיכולוגית השוינו במובן זה **לעלילה הבלשית** שמתאר לואיס, שבאמצעותה אופיינו שלוש הדמויות.

אולם כעת עולה השאלה הבאה: **מהי, לפי ההשקפה הפונקציונליסטית, אותה תיאוריה פסיכולוגית האמורה להוות את הבסיס לאפיוניהם של המצבים הנפשיים?** הדרך הטובה ביותר להבין את משמעותה המדויקת של שאלה זו היא לבחון את התשובות השונות שהוצעו לה. לשאלה זו הוצעו שתי תשובות עיקריות. כפי שנראה, יש לתשובות אלו חשיבות רבה, שכן הן עומדות בבסיסן של שתי גרסאות שונות מאוד זו מזו של התיאוריה הפונקציונליסטית – פונקציונליזם עממי ופונקציונליזם אמפירי.

### 4.1 פונקציונליזם עממי (או אנליטי)

#### 4.1.1 הפסיכולוגיה העממית

לפי הגרסה של ההשקפה הפונקציונליסטית הקרויה "פונקציונליזם עממי" (folk functionalism), או "פונקציונליזם אנליטי" (analytical functionalism), התיאוריה

הפסיכולוגיה האמורה להוות את הבסיס לאפיוניהם של המצבים הנפשיים היא **הפסיכולוגיה העממית** (Folk Psychology).<sup>46</sup> למה הכוונה ב"פסיכולוגיה עממית"?

לעתים קרובות **אנו מייחסים לזולת (כמו גם לעצמנו) מצבים נפשיים שונים**: האמנות, רציות, פחדים, תקוות, כאבים, רגשות, זיכרונות, ועוד ועוד. לעתים אנו מייחסים לזולת מצבים נפשיים אחדים על-סמך מצבים נפשיים אחרים שאנו סוברים שהוא נמצא בהם ("מאחר שאלי חושב שבני אדם מהימן, הוא ודאי מאמין שבעוד חודש יחזיר לו את ההלוואה"); לעתים על-סמך התנהגותו ("רן מתלבש יפה כשלימור אמורה להיות בסביבה ברור שהוא מאוהב בה"); לעתים על סמך מצב הסביבה בה הוא נמצא – הגירויים החושיים שאנו מניחים כי סביבתו מספקת לו ("הגר עמדה קרוב מאוד לנמר. היא בוודאי מבוהלת"), ולעתים על סמך צירוף של מספר גורמים כאלו. לעתים קרובות **אנו חוזים את התנהגות הזולת** על-סמך המצבים הנפשיים שאנו מייחסים לו ("מאחר ששרגא רוצה לקבל את המשרה ומאמין שלגלית השפעה על הבוס, תראה איך הוא יתחנף אליה"), ולעתים **אנו מסבירים את התנהגות הזולת** על-סמך המצבים הנפשיים שאנו מייחסים לו ("אתי שדדה את הבנק כי רצתה לחלץ את אחיה מהמצוקה הכלכלית שאליה נקלע").

כולנו עושים דברים כאלו באופן שגרתי ויומיומי – ייחוסים, חיזויים והסברים כאלו הם חלק בלתי-נפרד מחיינו ומיחסינו עם הזולת. במובן מסוים, אפשר אפוא לומר כי **כולנו עוסקים בפסיכולוגיה**. כמעט כל אדם עושה זאת, אף אם לא למד ולו שיעור אחד בפסיכולוגיה, ולא קרא ולו עמוד אחד בספר פסיכולוגיה, ואף אם אין לו מושג מהי פסיכולוגיה. במובן זה, כולנו עוסקים בפסיכולוגיה **עממית**. הפסיכולוגיה העממית היא העיסוק האינטואיטיבי והטבעי בייחוס מצבים נפשיים לזולת באופנים שהזכרנו עתה, בחיזוי והסבר התנהגויות הזולת על-סמך מצביו הנפשיים, וכדומה.

מקובל למדי להניח (וכך מניחים הדוגלים בפונקציונליזם עממי) כי ביסוד הייחוסים, החיזויים וההסברים הללו עומדים כללים שונים. לפניכם דוגמאות אחדות של כללים מסוג הכללים המוזכרים לעתים בהקשר זה:

<sup>46</sup> שמות אחרים, פחות רווחים, של גרסה זו של פונקציונליזם (או של גרסאות הדומות לה), הם "פונקציונליזם של השכל הישר" (Common-sense functionalism), "פונקציונליזם א-פריורי" (a priori functionalism), ו"פונקציונליזם מושגי" (conceptual functionalism).  
בין הכתבים המגנים על גרסה זו ומפתחים אותה ניתן להזכיר את הללו:

D. Lewis, "Psychophysical and Theoretical Identifications", *Australasian Journal of Philosophy* 50 (1972), pp. 249-258; D. Lewis, "Mad Pain and Martian Pain", *Readings in Philosophy of Psychology*, Vol. 1, ed. N. Block, Cambridge (Mass.): Harvard University Press, 1980; D. Lewis, "Lewis, David: Reduction of Mind", *A Companion to the Philosophy of Mind*, Oxford: Blackwell, 1994; S. Shoemaker, *Identity, Cause and Mind*, Oxford: Oxford University Press, 1984; F. Jackson and P. Petit, "Folk Belief and Commonplace Belief", *Mind and Language* 8, (1993), pp. 298-305; D. Braddon-Mitchell and F. Jackson, *Philosophy of Mind and Cognition*, Oxford: Blackwell, 1996, and "The Divide and Conquer Path to Analytic Functionalism", *Philosophical Topics* 26 (1999), pp. 71-89.

- (1) בני אדם נוטים לחוש כאב כשחלק מגופם ניזוק באופן משמעותי.
- (2) בני אדם החשים כאב חזק נוטים לרצות להיפטר מהכאב.
- (3) בני אדם הנתקפים באופן פתאומי בכאב חזק נוטים לגנוח.
- (4) בני אדם החשים צמא נוטים לרצות לשתות.
- (5) כשבני אדם עצבניים, הם נוטים להיות חסרי-סבלנות, ולהזעיק פנים.
- (6) בני אדם המשיגים מטרה החשובה להם, מבלי לשלם עבורה מחיר היקר בעיניהם, נוטים לשמוח.
- (7) לבני אדם הסובלים מדיכאון יש נטייה להיות פסימיים.
- (8) כשבני אדם שמערכת הראייה שלהם תקינה עומדים לאור יום במרחק של עשרה מטר מפיל, הם רואים את הפיל, ומאמינים שלפניהם פיל.
- (9) בני אדם המאמינים שאם התכנית הכלכלית שהממשלה תציע תפגע באופן משמעותי בחלשים, אין לבצעה, והמשתכנעים שהתכנית הכלכלית שהממשלה הציעה תפגע באופן משמעותי בחלשים, יטו לחשוב שאין לבצע את התכנית הכלכלית.
- (10) בני אדם הרוצים ב-P ומאמינים ש-Q יהיה אמצעי יעיל להשגת P, ואשר אין להם רציות מנוגדות ואין להם אסטרטגיות חלופיות להשגת P, נוטים לנסות להשיג את Q.

אלו הן דוגמאות של כללי הפסיכולוגיה העממית, כללים שמקובל למדי לחשוב (וכך מניחים הדוגלים בפונקציונליזם עממי) כי ידיעה מסוימת שלהם עומדת בבסיס ייחסי המצבים הנפשיים. כללים אלו מהווים את **התיאוריה של הפסיכולוגיה העממית**, את **גוף הידע** העומד ביסוד הייחוסים, החיזויים וההסברים הללו, והמאפשר אותם.

**שימו לב:** אנו מבחינים כאן בין **העיסוק** (או "הפרקטיקה") של הפסיכולוגיה העממית, לבין **התיאוריה** של הפסיכולוגיה העממית. העיסוק בפסיכולוגיה העממית הוא עצם הייחוס של מצבים נפשיים באופנים שהזכרנו, עיסוק שקשה להכחיש כי כולנו עוסקים בו. התיאוריה של הפסיכולוגיה העממית היא גוף הידע העומד ביסוד עיסוקנו בפסיכולוגיה עממית – אוסף הכללים והעקרונות שידעתם מאפשרת לנו לייחס לזולת מצבים נפשיים באופנים הנזכרים. כאמור, מקובל למדי לחשוב כי ביסוד עיסוקנו בפסיכולוגיה עממית עומדת ידיעה של כללים ועקרונות (כמו אלו שהדגמנו לעיל); במילים אחרות, מקובל למדי לחשוב שאמנם קיימת תיאוריה של פסיכולוגיה עממית, תיאוריה שבה מחזיק כל אחד מאתנו. יש החולקים על הטענה שביסוד העיסוק בפסיכולוגיה עממית אמנם עומדת תיאוריה (על הטענה שאמנם כולנו, העוסקים בפרקטיקה הנדונה, יודעים כללים ועקרונות כדוגמת אלו שהדגמנו, ומיישמים אותם כשאנו מייחסים לזולת מצבים נפשיים). נתייחס לכך בהמשך. אנשי



הפונקציונליזם העממי דוגלים בטענה שאמנם ביסוד העיסוק בפסיכולוגיה עממית עומדת תיאוריה. מעתה, כשנשתמש בביטוי "פסיכולוגיה עממית", כוונתנו לציין את ה**תיאוריה** שביסוד ייחוס המצבים הנפשיים.

כאשר מדובר בתיאוריה של הפסיכולוגיה העממית, חשוב להעיר כי אין זו תיאוריה במלוא מובן המילה. גוף הידע העומד ביסוד העיסוק בפסיכולוגיה עממית אינו גוף ידע שיטתי, מאורגן ומדויק, אשר חלקיו מתלכדים אלו עם אלו. במילים אחרות, התיאוריה של הפסיכולוגיה העממית, כפי שמעיד שמה, היא "תיאוריה" עממית ולא תיאוריה מדעית מסודרת ושיטתית. הפסיכולוגיה העממית היא התפיסה של השכל הישר בדבר הנפש ואופן פעילותה. ידיעתם של כללי התיאוריה הזו ועקרונותיה נרכשת בשלבים המוקדמים של חינוך, ויש הטוענים שידעיתם של חלק מהם טבועה בשכלנו מלידה.

באופן דומה, כאשר מדובר על כך שאנו **יודעים** את הכללים והעקרונות של התיאוריה – הכללים והעקרונות שידעיתם מאפשרת לנו לייחס לזולת ייחוסים מן הסוגים האמורים, הכוונה איננה שאנו חושבים במפורש ובמודע על הכללים והעקרונות הללו בכל מקרה בו אנו מייחסים לזולת מצבים נפשיים. ברוב המקרים אנו מחילים את הכללים והעקרונות באופן בלתי-מודע. אנו נרחיב בעניין זה בהמשך.

אם כן, בדברנו על הפסיכולוגיה העממית כוונתנו לגוף של ידע בסיסי (במובן של "ידע" שעוד זקוק להבהרה) שאותו (או את עיקריו) אנו חולקים – ידע שעניינו במצבים נפשיים, בקשרים בין מצבים נפשיים לבין מצבים הנפשיים אחרים, וביניהם לבין גירויים חיצוניים והתנהגויות שונות. המדובר במצבור של הנחות מקובלות (platitudes) – הנחות שאנו רואים כמובנות מאליהן – על אודות, למשל, כאבים, האמנות, רציות, רגשות וכוונות; הנחות על קשרים הדדיים בין מצבים כאלו, על קשרים ביניהם לבין סביבתו של היצור הנפשי הנמצא במצבים אלו, וביניהם לבין התנהגויותיו. אנו יודעים, במידה זו או אחרת, מה גורם לאנשים לשמוח ומה גורם להם להתעצב; אנו יודעים שהאמנות מסוימות מוליכות לגיבוש האמנות אחרות; שרציות מסוימות, כשהן חוברות להאמנות מסוימות, מוליכות לרציות אחרות, ועוד ועוד.

כאמור, על פי תפיסה פילוסופית רווחת (המקובלת, בין היתר, על הדוגלים בפונקציונליזם העממי), ידע זה עומד בבסיס יכולותינו ליחס לבני אדם מצבים נפשיים שונים, ולחזות ולהסביר את התנהגותם ומצביהם הנפשיים. כך, על פי תפיסה זו, יכולתי לחזות שאורי היחף, שהרגע דרך על מסמר, יחוש כאב, ירצה להיפטר מהכאב ויגנח, מבוססת על כך שהחלתי את ההכללות המבוטאות בכללים 1-3 (ברשימה דלעיל) על המקרה הנדון; יכולתי להסביר את העובדה שדינה פתחה את מגרת הממתקים כתוצאה של העובדה שרצתה לאכול שוקולד, ושהאמינה שחפיסת שוקולד תמצא במגרת הממתקים, מבוססת על החלת ההכללה המבוטאת בכלל 10 על המקרה הנדון, וכדומה.

הפסיכולוגיה העממית איננה התיאוריה העממית היחידה. כולנו יודעים, גם אם לא למדנו ולו שיעור אחד בפיזיקה, שאם נרפה מן התפוח שבידנו הוא יפול; ושאם נבעט בכדור בחוזקה הוא ינוע בכיוון הבעיטה במהירות, וכדומה. במילים אחרות, אנו יודעים "פיזיקה עממית".

#### 4.1.2 הפונקציונליזם העממי

כפי שחזרנו ואמרנו, לפי ההשקפה הפונקציונליסטית מצבים נפשיים מאופיינים על ידי מיקומם, או תפקידם, ברשת של קשרים – רשת של קשרים השוררים בין המצבים הנפשיים עצמם, ביניהם לבין קלטים חושיים, וביניהם לבין פלטים התנהגותיים. הפסיכולוגיה העממית מתארת רשת קשרים כזו, ולפי הפונקציונליזם העממי זוהי הרשת הרלוונטית לאפיונים נפשיים. כלומר, לפי השקפת הפונקציונליזם העממי, המיקום באותה רשת של קשרים (בין מצבים נפשיים לבין עצמם וביניהם לבין קלטים חושיים ופלטים התנהגותיים) אשר מתארת הפסיכולוגיה העממית, הוא המאפיין את המצבים הנפשיים וקובע את סוגיהם. הקשרים המתקיימים על פיטענות הפסיכולוגיה העממית הם המאפיינים את המצבים הנפשיים, לפי גישה זו. בהמשך נציג את הגישה הפונקציונליסטית החלופית, שעל פיה ייחוס הפסיכולוגיה המדעית – תיאוריה שאותה רובנו איננו מכירים – הם המאפיינים את המצבים הנפשיים. אולם לפי הגישה בה אנו דנים עתה – הפונקציונליזם העממי – "תיאוריה" המוכרת כמעט לכל אחד ואחד מאתנו עומדת בבסיס אפיוני המצבים הנפשיים.

לפי הפונקציונליזם העממי, מצבים נפשיים מאופיינים על ידי מיקומם ברשת הקשרים המתוארת על ידי הפסיכולוגיה העממית, קשרים אשר שוררים בין המצבים הנפשיים עצמם, ביניהם לבין קלטים חושיים, וביניהם לבין פלטים התנהגותיים.

העובדה שאפיוני הפסיכולוגיה העממית מוכרים כמעט לכולנו היא מהותית להגיון שביסוד הפונקציונליזם העממי, שכן עמדה זו היא **עמדה אנליטית**. על פיה, הטענות בדבר האפיונים הפונקציונליים של מצבים נפשיים **קובעות את משמעות המונחים הנפשיים**. במילים אחרות, הזהויות בין מצבים נפשיים לבין מצבים פונקציונליים ובין תכונות נפשיות לבין תכונות פונקציונליות, לפי גישה זו, הן **זהויות אנליטיות**, זהויות הנובעות ממשמעות המונחים. כך, למשל, משמעות המונח "כאב" היא מצב פנימי הקשור עם מצבים נפשיים, קלטים ופלטים, באופן שמתואר על ידי כללים 1-3 (מרשימת הכללים שהובאה לעיל) ועל ידי כללים דומים רבים אחרים: משמעות המונח "כאב" היא מצב פנימי שבני אדם נוטים להימצא בו כשחלק מגופם ניזוק באופן משמעותי, שבני אדם הנמצאים בו נוטים לרצות להיפטר ממנו, וכדומה. טענות אלו ואחרות מכוננות את מושג הכאב שלנו. מי שאינם מבינים שאם אורי חש כאב, אז אורי נמצא במצב פנימי שאנשים הנמצאים בו נוטים לרצות להיפטר ממנו, אינם מבינים **מה פירוש הדבר** שאורי חש כאב. הם לא סתם לוקים בחוסר ידיעה

מסוימת על אודות כאבים (כדוגמת הידיעה בדבר הקשר בין כאבים לבין בליעת גלולות אקמול) – הם אינם מבינים את משמעות המונח "כאב", הם אינם שולטים במושג הכאב. ובאופן דומה, מי שאינם מבינים שאם דינה רוצה לאכול שוקולד, ומאמינה שתמצא חפיסת שוקולד במגרת הממתקים, אז (בהנחה שאין לה רציות מנוגדות, ואין לה דרכים יעילות אחרות לשים את ידה על חפיסת שוקולד, וכדומה) דינה תטה לפתוח את מגרת הממתקים, אינם מבינים מה **פירוש הדבר** שדינה רוצה לאכול שוקולד ומאמינה שתמצא חפיסת שוקולד במגרת הממתקים.

נשוב ונרענן את זיכרוננו באשר למושגים "**אנליטי**" ו"**סינתטי**". משפטי נושא-נשוא, דהיינו משפטים שצורתם "א הוא ב", הם או אנליטיים או סינתטיים. משפטים כאלו הם **אנליטיים** אם הנשוא שלהם אינו מוסיף דבר על מה שכלול בנושא. לדוגמה, "האב הוא הורה" הוא משפט אנליטי, שכן הנשוא שלו, "הורה", כלול בנושא שלו, "האב" (שהרי משמעות "אב" היא **הורה זכר** – דהיינו גם **הורה** וגם **זכר**). אמתותו של משפט אנליטי אינה נעוצה כלל במצב העניינים בעולם – הוא אמתו אך ורק בשל משמעות המונחים המופיעים בו.

משפטי נושא-נשוא הם סינתטיים, אם הנשוא שלהם מוסיף דבר-מה על מה שכלול בנושא. לדוגמה "התפוח שוקל יותר מ-10 גרם" הוא משפט סינתטי, שכן הנשוא "שוקל יותר מ-10 גרם" מוסיף דבר-מה על מה שכלול בנושא "התפוח" – המונח "תפוח" אינו כולל, כחלק ממשמעותו, **שוקל יותר מ-10 גרם**. אמתותו של משפט סינתטי (אם הוא אמתו) אינה תלויה רק במשמעות המונחים המרכיבים אותו, אלא גם במצב העניינים בעולם. המשפט "התפוח שוקל יותר מ-10 גרם" אמתו, אם הוא אמתו, הן בשל משמעות מונחיו והן בשל כך, שבעובדה בעולם, התפוח שבו מדובר שוקל יותר מ-10 גרם.

דייויד לואיס (David Lewis), הנחשב לדעת הרוב לאביה של עמדת הפונקציונליזם העממי, הבהיר כיצד נגזרת משמעותם של המונחים הנפשיים מטענותיה המרכזיות של הפסיכולוגיה העממית על פי עמדה זו, ומהו "המתכון" לניתוחו של מונח נפשי מסוים על ידי פנייה לטענות אלו:

עניינה של הפסיכולוגיה העממית ביחסים הסיבתיים שבין מצבים נפשיים, גירויים תפיסתיים ותגובות התנהגותיות... כך היא משייכת לכל מצב נפשי תפקיד סיבתי טיפוסי. עכשיו יש לנו מתכון לניתוח. נניח שהצלחנו להפיק את כל העקרונות הכלליים, הידועים במובלע, של הפסיכולוגיה העממית. אם M הוא השם בפסיכולוגיה העממית למצב נפשי, הפסיכולוגיה

העממית תאמר שהמצב M ממלא באופן טיפוסי תפקיד סיבתי מסוים – נקרא לו תפקיד M. אזי אנחנו מנתחים את M כבעל המשמעות 'המצב אשר באופן טיפוסי ממלא את תפקיד M'.<sup>47</sup>

המתכון פשוט למדי. משמעותו של ביטוי נפשי מסוים דוגמת "כאב", "האמנה שהשמים הם כחולים", וכדומה, היא "המצב שבאופן טיפוסי ממלא תפקיד סיבתי-פונקציונלי כזה וכזה", ואילו תיאורו של התפקיד האמור (מה שייכתב במקום המילים "כזה וכזה") ניתן על ידי טענותיה של הפסיכולוגיה העממית באשר למצב הרלוונטי.

על פי הפונקציונליזם העממי, טענות התיאוריה הפסיכולוגית המאפיינות את המצבים הנפשיים קובעות את משמעויות המונחים הנפשיים.

כאמור, רעיון זה קשור הדוקות לרעיון המרכזי של הפונקציונליזם העממי, על פיו התיאוריה הפסיכולוגית האמורה להוות את הבסיס לאפיוניהם של המצבים הנפשיים היא הפסיכולוגיה העממית, תיאוריה המורכבת מטענות שבעיקרן ידועות לכול.

## שאלה

כיצד, לדעתכם, קשורים שני הרעיונות הללו?

שני הרעיונות קשורים הדוקות, כך נראה, משום שהטענות הקובעות את משמעויות המונחים הנפשיים חייבות להיות ידועות למי שמשתמשים במונחים נפשיים ומבינים אותם, דהיינו למרבית בני האדם. לא ייתכן שאנו מבינים את משמעותו של מונח מסוים, משמעות שנקבעת על ידי טענות מסוימות, ואיננו מכירים טענות אלו. כך, עמדה שעל פיה משמעויות המונחים הנפשיים נקבעות על ידי טענות בדבר קשרים בין המצבים הנפשיים לבין עצמם וביניהם לבין קלטים ופלטטים, מחויבת לכך שהטענות הללו ידועות למי שמשתמשים במונחים הנפשיים ומבינים אותם (דהיינו לרובינו). אילו הטענות שלפי ההשקפה הנדונה קובעות את משמעויות המונחים הנפשיים היו טענות הידועות רק לחלק קטן מאתנו – למשל, לפסיכולוגים מומחים, שברשותם ידע אודות המצבים הנפשיים שאינו מצוי ברשות האחרים – התוצאה הייתה שרובינו (מי שאינם פסיכולוגים מומחים) איננו מבינים את משמעויות המונחים הנפשיים. נראה שעמדה בדבר משמעויות המונחים הנפשיים על פיה דוברים רגילים של השפה אינם יודעים את פירוש המונחים "כאב", "רצייה בשוקולד", "שמחה" ו"האמנה שהשמים כחולים" היא עמדה בעייתית. עמדה זו נראית בלתי סבירה, במיוחד לנוכח העובדה שרובינו משתמשים במונחים נפשיים באורח שגרתי ויומיומי, שהרי רובינו מייחסים לזולת מצבים נפשיים. כך, פונקציונליסטים הסוברים שהטענות בדבר קשריהם הסיבתיים של המצבים

<sup>47</sup> D. Lewis, "Lewis, David: Reduction of Mind", in S. Gutenplan (ed.), *A Companion to the Philosophy of Mind*, Oxford: Blackwell, 1994, p. 416.

הנפשיים קובעות את משמעויות המונחים הנפשיים, מחויבים לעמדה שטענות אלו, בעיקרן, ידועות לכול; הם מחויבים לעמדה שהתיאוריה הפסיכולוגית המורכבת מטענות כאלו חייבת להיות תיאוריה שבמובן זה או אחר ידועה לכל. אין היא יכולה להיות תיאוריה פסיכולוגית מדעית (שאינה מוכרת למרבית בני האדם), אלא היא חייבת להיות התיאוריה הפסיכולוגית העממית, זו העומדת ביסוד ייחוס המצבים הנפשיים שבהם עוסקים רובינו באופן שגרתי.

ראינו, למעשה, כי עמדת הפונקציונליזם העממי מהווה דין וחשבון באשר למשמעויות המונחים הנפשיים. במילים אחרות, הפונקציונליזם העממי היא תורה אנליטית. על פי תורה זו, הניתוח הנכון של מונחים נפשיים כמו "כעס", "תקווה", "האמנה שהתרופה החדשה תשפר את ראייתי באופן ניכר", וכדומה, הוא ניתוח פונקציונליסטי – משמעויות המונחים הללו ניתנות על ידי תיאור הפונקציות שבהן כרוכים המצבים הנפשיים שאותם מציינים המונחים, דהיינו על ידי תיאור קשריהם הסיבתיים למצבים אחרים. בשל היותה תורה אנליטית, תורה בדבר משמעויות המונחים הנפשיים, מכונה עמדת הפונקציונליזם העממי גם בשם "פונקציונליזם אנליטי".<sup>48</sup>

אמרנו בשלב מוקדם שההשקפה הפונקציונליסטית נחשבת כיורשתה של ההשקפה הביהביוריסטית. עתה ניתן לראות, שגרסה מסוימת של ההשקפה הפונקציונליסטית – הפונקציונליזם העממי – קרובה במיוחד להשקפה הביהביוריסטית. זאת משום ששתי ההשקפות הן השקפות אנליטיות; שתיהן הן, בין היתר, השקפות בדבר משמעויות המונחים הנפשיים, השקפות המזהות תכונות נפשיות עם תכונות מסוימות, כאשר הזהות הנטענת היא זהות אנליטית ולא זהות עובדתית גרידא (כזכור, אחד מכינוייה של ההשקפה הביהביוריסטית הוא "ביהביורזם אנליטי"). אולם השקפת הפונקציונליזם העממי נחשבת כיורשת של ההשקפה הביהביוריסטית לא רק משום ששתי ההשקפות הללו הן השקפות אנליטיות בדבר משמעויות המונחים הנפשיים, אלא משום ששתיהן מנתחות באופנים דומים את המונחים הנפשיים – כמונחים המציינים פונקציות או במצבים המממשים פונקציות. ויש אף דמיון בין הפונקציות הרלוונטיות בשני המקרים – הן כרוכות בגירויים חושיים ובתגובות התנהגותיות, אלא שהפונקציות שעליהם מדברים הפונקציונליסטים כרוכות ברכיבים מסוג נוסף – במצבים נפשיים אחרים של היצור הנפשי.

## שאלה

הזכרנו תיאוריה נוספת, שניתן לראות כי השקפת הפונקציונליזם העממי קרובה אליה אף יותר מאשר לביהביורזם. מהי השקפה זו?

<sup>48</sup> יש המבחנים בין עמדת "הפונקציונליזם האנליטי", שרעיונה המרכזי הוא כי האפיונים הפונקציונליים של המצבים הנפשיים (ללא קשר לשאלה באיזו תיאוריה, אם בכלל, הם כרוכים) קובעים את משמעויות המונחים הנפשיים, לבין עמדת "הפונקציונליזם העממי", שרעיונה המרכזי הוא כי הטענות הפסיכולוגיות האמורות להוות את הבסיס לאפיוניהם של המצבים הנפשיים הן טענות של הפסיכולוגיה העממית. אולם מאחר שמקובל לחשוב שאם אפיונים פונקציונליים כלשהם קובעים את משמעויות המונחים הנפשיים הרי אפיונים אלו חייבים להיות האפיונים העולים מטענות הפסיכולוגיה העממית, הפילוסופים הדוגלים ברעיון הראשון מקבלים בדרך כלל גם את הרעיון השני, כך שבפועל ההבחנה התיאורטית בין שתי העמדות אינה באה לידי ביטוי.

מדובר בהשקפה של ארמסטרונג (שהיא למעשה גרסה של ההשקפה שכינינו "פיזיקליזם התנהגותני"). גם השקפתו של ארמסטרונג היא, בין היתר, השקפה בדבר משמעות המונחים הנפשיים. גם היא מנתחת את המונחים הנפשיים במונחים פונקציונליים, ובדיוק כמו הפונקציונליזם העממי (ובניגוד לביהביוריסם), על פיה המונחים הללו מציינים **מצבים המממשים פונקציות**. אבל אין זהות בין ההשקפה של ארמסטרונג לבין הפונקציונליזם העממי: הפונקציות, לפי ההשקפה של ארמסטרונג, אמנם דומות לפונקציות בתמונה של הפונקציונליזם העממי – הן כרוכות בגירויים חושיים ובתגובות התנהגותיות, אולם הן אינן זהות – הפונקציות שעליהן מדברים הפונקציונליסטים כרוכות גם בקשרים למצבים נפשיים אחרים של היצור הנפשי.

חשוב מאוד להבהיר כי השקפת הפונקציונליזם העממי איננה רק השקפה בדבר משמעויות המונחים הנפשיים. השקפה זו היא בהחלט גם השקפה בדבר המצבים הנפשיים והתכונות הנפשיות. היא מתחייבת לכך שקיימות **זהויות בין מצבים נפשיים לבין מצבים פונקציונליים מסוימים, ובין תכונות נפשיות לבין תכונות פונקציונליות מסוימות**. ייחודה של השקפה זו כהשקפה אנליטית מתבטא בכך שהזהויות הנפשיות-פונקציונליות על פיה הן **זהויות אנליטיות** – זהויות המתקיימות בשל קשרי המשמעות בין המונחים הנפשיים לבין תיאורי הקשרים הסיבתיים בין המצבים הנפשיים לבין מצבים אחרים – ולא טענות סינתטיות. ראינו ביחידה 4-5 כי לפחות לכאורה קיים קושי באשר לאפשרותה של זהות סינתטית של תכונות. הפונקציונליזם העממי אינו מאוים על ידי קושי זה. בעיני מי שמתרשמים הן כי קושי זה הוא קושי חמור והן כי הרעיון הפונקציונליסטי ביסודו הוא רעיון נכון, לגרסה אנליטית של פונקציונליזם יש יתרון חשוב.

כזכור, על פי ההשקפה הפיזיקליסטית (או תורת הזהות הנפשית-חומרית) על גרסאותיה השונות, הזהות בין מצבים נפשיים לבין מצבים פיזיקליים היא **זהות סינתטית**, זהות שבעובדה. הדוגלים בפונקציונליזם עממי שהם גם פיזיקליסטים – וכאמור, מרבית הפונקציונליסטים הם פיזיקליסטים – מחויבים אפוא לתמונה הכוללת זהויות משני סוגים. התכונות הנפשיות קשורות לאפיונים פונקציונליים מסוימים, לפונקציות מסוימות, בקשרים אנליטיים, ואילו זהו עניין שבעובדה שהפונקציות אשר מאפיינות את המצבים הנפשיים ממומשות על ידי מצבי-מוח (או, במילים אחרות, שמצבי-מוח משחקים את התפקידים הסיבתיים הרלוונטיים). גם על פי הפונקציונליזם העממי אפוא הקשר בין הנפש למוח הוא ביסודו קשר עובדתי. אולם בהמשך נראה כי יש להבין טענה זו באופן מסויג.

יש להדגיש שהדוגלים בפונקציונליזם עממי או אנליטי סבורים שניתוח של השפה – ניתוח משמעויותיהם של המונחים הנפשיים – מצדיק את העמדה הפונקציונליסטית. במילים אחרות, כאשר אנו בוחנים כיצד אנו מבינים מונחים כמו "כאב", "רצייה", וכדומה, אנו מגלים שאנו מבינים אותם כמונחים פונקציונליים. מונחים כמו "מנוע" או "תרמוסטט" ("וסת חום") הם מונחים פונקציונליים – אנו מבינים אותם כמציינים אובייקטים הממלאים פונקציות מסוימות (ולא, למשל, אובייקטים בעלי מבנה פיזיקלי מסוים) – "תרמוסטט" עשוי

לציין אובייקטים בעלי מבנים פיזיקליים שונים לחלוטין, כל עוד אובייקטים אלו ימלאו את הפונקציה של ויסות חום). כך, אנו מגלים שאנו מבינים את המונחים הנפשיים כמונחים המציניים מצבים המשחקים תפקידים סיבתיים מסוימים, מצבים הממלאים פונקציות של קישור בין קלטים חושיים, פלטים התנהגותיים ומצבים נפשיים אחרים. ולפיכך מתקיימות זהויות – זהויות אנליטיות – בין כאב לבין מצב הממלא פונקציה כזו וכזו, בין רצייה לבין מצב הממלא פונקציה כזו וכזו, וכדומה. לשפה מיוחס אפוא בעניין זה תפקיד חשוב ביותר.

ארמסטרונג טען, כזכור, לקשר מושגי בין מונחים נפשיים לבין תיאורים התנהגותיים, תיאורים של דגמי גירוי ותגובה. כשמצרפים לטיעונו זה של ארמסטרונג את הטענה שעל פיה התפקידים הסיבתיים של המצבים הנפשיים חייבים לכלול התייחסויות לא רק לגירויים תפיסתיים ולתגובות התנהגותיות אלא גם למצבים נפשיים אחרים של היצור הנפשי, מה שמתקבל הוא נימוק בזכות הפונקציונליזם העממי.

### 4.1.3 פונקציונליזם עממי וידיעת טענות הפסיכולוגיה העממית

הזכרנו שטענות הפסיכולוגיה העממית אשר קובעות את משמעויות המונחים הנפשיים אמורות להיות ידועות לכולנו. אולם **באיזה מובן עליהן להיות ידועות לנו?** באחד ממאמריו המוקדמים כתב דייויד לואיס את הדברים הללו:

קבצו את כל האמירות המקובלות שאתם יכולים לחשוב עליהן בהקשר של יחסים סיבתיים של מצבים נפשיים, גירויים חושיים ותגובות מוטוריות. כללו רק אמירות מקובלות שהן בבחינת ידע משותף – כל אחד יודע אותן, כל אחד יודע שכל אחד אחר יודע אותן, וכך הלאה.<sup>49</sup>

דברים אלו פורשו כמכילים את הדרישה שהטענות האמורות צריכות להיות ידועות לכל במובן חזק – שביכולתנו לנסח אותן אחת לאחת, ולספק רשימה מוגמרת שלהן שתכלול כל מה שמהותי באשר למצבים נפשיים ולקשריהם למצבים נפשיים אחרים, לקלטים ולפלטים. כלומר, הידע הפסיכולוגי בו מדובר הוא **ידע מפורש**. המובן שבו אנו יודעים את העובדות שעל אודותיהן נסבות הטענות הרלוונטיות דומה למובן שבו אנו יודעים שלונדון היא בירת אנגליה, ששתיים ועוד שתיים הן ארבע, וכדומה. הטענות הללו נגישות לנו, באופן כזה שלו רק נרצה בכך ונחשוב על נושאיהן, נוכל לנסח אותן בצורה מפורשת.

אולם רבים (ובניהם לואיס עצמו, בעבודתו המאוחרת יותר) סבורים שדרישה שכזו – הדרישה שרק טענות פסיכולוגיות מקובלות שאותן אנו יודעים **במפורש** יכללו בגוף הטענות המשחקות תפקיד באפיונם של המצבים הנפשיים – היא דרישה חזקה מדי. חשבו, למשל, על טענה 10 מן הרשימה דלעיל. האומנם סביר להניח שבני אדם רבים היו מגיעים בעצמם

<sup>49</sup> D. Lewis, "Psychophysical and Theoretical Identifications", *Australasian Journal of Philosophy* 50 (1972), p. 256.

לניסוח כלשהו של טענה זו: ספק רב. אולם טענה זו מבטאת עובדה חשובה על אודות האמנות ורציות, עובדה שבדאי מהווה חלק חשוב של האפיון הפונקציונלי של המצבים הנפשיים הללו, ולפיכך דומה שידע על אודותיה נדרש לשם מושגי ההאמנה והרציה. כיצד אפוא יכולים פילוסופים הדוגלים בפונקציונליזם עממי להתמודד עם קושי זה?

ובכן, רבים מן הדוגלים בעמדת הפונקציונליזם העממי מצביעים על העובדה שעמדה זו איננה מחויבת לכך שהידע הפסיכולוגי העממי הוא ידע מפורש. חלקו עשוי להיות ידע **מובלע**. נקרא את דבריהם של דייוויד ברדון-מיצ'ל (David Braddon-Mitchell) ופרנק ג'קסון (Frank Jackson), ממגיניה של העמדה:

נראה לנו, שמה שעל פונקציונליסטים של השכל הישר לעשות זה לפנות לידע **מובלע**. הדגם הקלאסי הוא ידיעתנו את הדקדוק. רובנו מסוגלים לשרבט מספר כללי דקדוק: 'על צורת הפועל להתאים לנושא', 'במרבית המשפטים צריך להיות נושא' וכדומה, אבל רשימת כללים הדומה בגודלה לרשימה המלאה היא מעבר ליכולתנו. למרות זאת, רובנו מסוגלים להבחין בין מה שדקדוקי למה שאינו דקדוקי, והיכולת הזו שלנו אינה עניין של מזל. יש לנו שליטה מובלעת בכללי הדקדוק. תפקיד המומחים לדקדוק הוא התפקיד הלא מבוטל של הפיכת הידע המובלע לידע מפורש. באותו אופן על הפונקציונליסט של השכל הישר לטעון שיש לנו ידע מובלע עשיר ומורכב (בצדו של הידע המפורש שלנו) של הקשרים בין מצבינו הנפשיים לקלטים ולפלטטים שלהם. מה שחושף את הידע המובלע שלנו כשמדובר בדקדוק הוא יכולתנו למיין משפטים כדקדוקיים או כלא דקדוקיים. מה שחושף את הידע המובלע שלנו כשמדובר בנפש הוא יכולתנו לחזות ולהסביר התנהגות בנסיבות שונות במונחים של מצבים נפשיים.<sup>50</sup>

לכולנו יכולות לשוניות מרשימות, הבאות לידי-ביטוי בכך שאנו מסוגלים ליצור ולהבין אינספור משפטים תקינים מבחינה דקדוקית, ולהבחין בין משפטים תקינים לבין משפטים שאינם כאלו. לכאורה נראה כי איננו יודעים את רוב הכללים הדקדוקיים – אין לנו כל יכולת לנסחם (גם לילדים צעירים שאינם יכולים לנסח אף כלל דקדוקי יש יכולות כאלו). אולם אף אם אין לנו ידע מפורש ובר ניסוח של הכללים הללו, עמדה מקובלת היא שיכולותינו הלשוניות מהוות ראיה טובה לכך שידע זה מצוי בידינו בצורה כלשהי, בלתי מפורשת ובלתי מודעת. ביצועינו הלשוניים והבחנותינו בין משפטים תקינים לבין משפטים שאינם תקינים מודרכים על ידי כללי הדקדוק. הרעיון הכללי הוא, שבהינתן מערך של יכולות קוגניטיביות מורכבות, ובהיעדר ידע מפורש של הסובייקטים היכול להסביר יכולות אלו, ישנם טעמים טובים להניח שמה שעומד ביסוד היכולות הללו הוא קיומו של מערך כללים הידועים לסובייקטים באופן מובלע, כללים המדריכים את ביצועיהם במקרים ספציפיים.

הדוגמה הדקדוקית היא אחת מיני רבות. מהלכים מעין אלו מהווים אחת מהאסטרטגיות ההסבריות הרווחות ביותר במדעים הקוגניטיביים – הם משמשים כדי להסביר יכולות קוגניטיביות מסוגים רבים ושונים, כמו, למשל, את יכולתנו לחזות ולהסביר את התנהגותם

<sup>50</sup> D. Braddon-Mitchell, and F. Jackson, *Philosophy of Mind and Cognition*, Oxford: Blackwell, 1996, pp. 55-56.



הפיזיקלית של גופים מסוימים (לדוגמה את יכולתנו לחזות שאם נזרוק בעוצמה כדור אל הקיר הוא יחזור אלינו). כך, ניתן לדבר על "תיאוריה דקדוקית עממית", על "תיאוריה פיזיקלית עממית", וכדומה – על מערכות של כללים הידועים לכולנו על אודות נושאים שונים, גם אם אינם ידועים לנו במפורש ואיננו מסוגלים לנסחם.

ברדון-מיצ'ל וג'קסון ופילוסופים רבים אחרים סוברים שגם במקרה הפסיכולוגי – המקרה של ייחוס מצבים נפשיים לזולת – קיים מערך של יכולות קוגניטיביות מורכבות, שסביר מאוד להסבירן באמצעות הנחתו של גוף ידע מובלע כזה על אודות מערך הקשרים בין מצבים נפשיים, קלטים ופלטרים. היכולות המדוברות הן יכולותינו להסביר ולחזות את התנהגויותיהם של בני אדם על ידי התייחסות למצביהם הנפשיים. דומה שנדרש ידע רב על אודות הסדירויות השולטות במעברים בין מצבים נפשיים, פלטרים וקלטים, כדי שתתאפשר יכולת של חיזוי והסבר של התנהגות בני אדם על-סמך מצביהם הנפשיים, התנהגות מורכבת ביותר הנקבעת על ידי מספר רב של גורמים. אמנם ברוב המקרים אין ביכולתנו לנסח את הכללים המתארים את הסדירויות הללו, אולם עובדה היא שיש לנו יכולות ניבוי וחיזוי מרשימות כאלו, ולדעת רבים עצם קיומן מעיד על כך שאנו אכן יודעים – באופן מובלע – לפחות את רובם של הכללים הרלוונטיים. ברדון-מיצ'ל וג'קסון מביעים הלך מחשבה זה באופן נחרץ ביותר, באומרים שללא ההנחה שברשותנו ידע פסיכולוגי מובלע מורכב כזה – ידע רב, מפורט ומורכב הכולל את ההכללות המרכזיות באשר לסדירויות הללו – קיומן של יכולות אלו ייותר בגדר נס.<sup>51</sup>

כך, נראה שהפונקציונליזם העממי יכול לחמוק מן הקושי הכרוך בכך שאין לנו ידע מפורש של חלק מהקשרים השוררים בין המצבים הנפשיים לבין עצמם ובינם לבין קלטים חושיים ופלטרים התנהגותיים: כמו בתחומים רבים אחרים, יש לנו ידע מובלע על אודות קשרים אלו, ידע הבא לידי ביטוי ביכולות החיזוי וההסבר שלנו את התנהגות הזולת.

קיימת עמדה מתחרה לעמדה זו באשר לשאלה מה עומד ביסוד יכולותינו לחזות ולהסביר את התנהגותם ומצביהם הנפשיים של בני אדם. עמדה חלופית זו, המכונה "תיאוריית הסימולציה", שוללת את הרעיון שמה שעומד ביסוד יכולות אלו הוא החלטה של תיאוריה על אודות תהליכים נפשיים שונים, והמעברים בין מצבים נפשיים, קלטים ופלטרים (התיאוריה שבה דנו עד כה, שבמרכזה רעיון זה, מכונה על פי רוב בשם "תיאוריית התיאוריה"). רעיונה המרכזי של העמדה החלופית הוא שבבואנו לחזות או להסביר את מצביו הנפשיים או את התנהגותו של אחר, אנו "שמים את עצמנו בנעליו" – אנו מדמיינים את עצמנו במצבו, עם נטיותיו הפסיכולוגיות, וכדומה, ונותנים לתהליכים הנפשיים לפעול את פעולתם. תוצאת התהליכים הללו, אם הכול מתנהל כשורה, היא מצבים נפשיים (ובכללם כוונות להתנהג באופנים מסוימים) הדומים לאלו של הסובייקט שברצוננו לחזות או להסביר. בקצרה, משימות החיזוי וההסבר האמורות מבוצעות באמצעות חיקוי או סימולציה של הפסיכולוגיה

Ibid., p. 57.

של האחר, ולא באמצעות החלתה של **תיאוריה** על אודות הפסיכולוגיה שלו.<sup>52</sup> שימו לב, שחסידי הפונקציונליזם העממי אינם חייבים לדחות את תיאורית הסימולציה מכל וכל: הם יכולים לטעון הן שיש לנו ידע מובלע על הסדירויות שבהן מעורבים מצבינו הנפשיים, והן שבפועל אנו חוזים ומסבירים את התנהגויות הזולת ואת מצביו הנפשיים באמצעות סימולציה (או גם באמצעות סימולציה). אולם יש המחזיקים בגרסה חזקה של תיאורית הסימולציה, שעל פיה סימולציה עומדת גם ביסוד עצם הבנתנו את המושגים הנפשיים, ולא רק ביסוד כושר הניבוי וההסבר שלנו את זולתנו. גרסה זו אינה מתיישבת עם הפונקציונליזם העממי.

שאלה חשובה בהקשרו של הדיון בפונקציונליזם העממי עניינה **באמתות הטענות של הפסיכולוגיה העממית**. ידוע לנו כי רבות מהטענות של תיאוריות עממיות אחרות התגלו כשקריות, או לפחות כמאוד לא מדויקות. בכך אין כמובן כדי להפתיע, שהרי תיאוריות עממיות אינן פרי מחקר שיטתי, מקיף ומדויק. אולם אם אמיתות לא מעטות של הפסיכולוגיה העממית הן שקריות, ייתכן שהפונקציונליזם העממי נתקל בבעיה. לפי הפונקציונליזם העממי מונח נפשי מציין את המצב הממלא את התפקידים הסיבתיים המיוחסים לו על ידי טענות הפסיכולוגיה העממית. המונח "כאב", למשל, מציין את המצב אשר נגרם על ידי נזק לרקמות הגוף, נוטה לגרום לרצייה להיפטר ממנו ולתחושת מצוקה, כשהוא חזק בעליו אינו יכול להפסיק להיות מודע לו, וכדומה. עתה, נניח שחלק מן הטענות אשר המונח "כאב" מופיע בהן הן שקריות. נניח, למשל, שבתנאים שונים אנשים מצליחים שלא לחשוב על כאב חזק, שבתנאים שונים הם אינם שואפים להיפטר ממנו, וכדומה (הנחות אלו כלל אינם דמיוניות. ישנם דיווחים רפואיים על תופעות מכל אחד משני הסוגים הללו). פירוש הדבר שאין **מצב נפשי** הממלא את התפקיד הסיבתי הזה. ואם אמנם המונח "כאב" מציין את המצב הממלא תפקיד זה, פירוש הדבר שמונח זה הוא "מונח ריק" – הוא אינו מציין דבר, ובפרט, הוא אינו מציין כל מצב נפשי.

דוגמה תסייע לנו להבין נקודה זאת. לפני מאות בשנים רווחה "תיאוריה עממית" מסוימת בדבר מכשפות. על פי תיאוריה זאת, נשים מסוימות מתאפיינות בתכונות אופי מסוימות ובכוחות מאגיים מסוימים, הן קשורות לגורמי רשע מסוימים, הן עפות בעזרת מטאטא, וכדומה. אם אנו מקבלים את ההשקפה שהמונח "מכשפה" הוגדר על ידי טענותיה של תיאוריה עממית זאת, הרי מאחר שטענות אלו הן שקריות – כל אדם בר-דעת בימינו יודע שאין (ולא היו) נשים בעלות התכונות המתוארות על ידי טענות אלו – המונח "מכשפה" אינו מציין דבר, הוא מונח ריק. באופן דומה, בהנחה הסבירה שרבות מטענות הפסיכולוגיה העממית הן שקריות, נראה שעמדת הפונקציונליזם העממי תקלע לתוצאה שעל פיה לפחות חלק מן המונחים הנפשיים אינם מציינים כל מצבים שהם. נניח, למשל, שאני אומר "כאבים

<sup>52</sup> רבים מהמאמרים החשובים באשר לויכוח בין תיאורית-התיאוריה לבין תיאוריית הסימולציה כלולים בספרים הללו:

M. Davies & T. Stone (eds.), *Folk Psychology*, Oxford & Cambridge (Mass.): Blackwell, 1995.

M. Davies & T. Stone (eds.), *Mental Simulation*, Oxford & Cambridge (Mass.): Blackwell, 1995.

בחלקו התחתון של הגב הם הכאבים הכרוניים הנפוצים ביותר", ושהפסיכולוגיה העממית אמנם טוענת – ובטעות! – שאנשים אינם יכולים להימנע מלחשוב על כאב חזק, שהם תמיד שואפים להיפטר מכאב חזק, וכדומה. מאחר שטענות אלו שקריות, כלומר **אין** מצב נפשי שאפיונים אלו חלים עליו, הרי שבהשמיעי משפט זה לא דיברתי על דבר – דברי חסרי משמעות! זוהי, כמובן, תוצאה בעייתית בעבור הפונקציונליזם העממי, שהרי קשה להכחיש שמשפטי זה הוא בעל משמעות, ושהמונח "כאב" מציין דבר מה.

זאת ועוד: למעשה, התנגדות זו איננה תלויה בהנחה **שכל** טענות הפסיכולוגיה העממית הן שקריות, ואף לא בהנחה **שרובן** שקריות. נניח שרק **טענה אחת** מאוסף טענות הפסיכולוגיה העממית אשר מכילות מונח נפשי מסוים ("פחד", "תקווה", "דיכאון", או כל מונח נפשי אחר) היא שקרית. אם כך, הרי שאין **מצב נפשי** שאוסף טענות זה כולו חל עליו. ומכאן, אם משמעותו של המונח הנפשי אמורה להיקבע על ידי **צירופן של כל טענות הפסיכולוגיה העממית אשר מכילות אותו**, הרי שאפילו שקריותה של טענה אחת מן הטענות הללו משמעה שאין מצב נפשי שמונח זה מציין. כיוון שסביר להניח, באשר למונחים נפשיים רבים, שחלק מהטענות של הפסיכולוגיה העממית שבהן הם מעורבים הן שקריות (שהרי, כאמור, בהיותה "תיאוריה" עממית אין הפסיכולוגיה העממית פרי של מחקר שיטתי ומדויק), קו מחשבה זה מראה שהפונקציונליזם העממי מוליך לתוצאה שמונחים נפשיים רבים הם מונחים ריקים. תוצאה זו נראית בעייתית מאוד בעבור עמדת הפונקציונליזם העממי.

כיצד יכול הפונקציונליזם העממי להתמודד עם ביקורת זו? שימו לב שביקורת זו הסתמכה על ההנחה שלפני הפונקציונליזם העממי משמעותו של מונח נפשי אמורה להיקבע על ידי **צירופן של כל הטענות** של הפסיכולוגיה העממית אשר מכילות אותו. זוהי דרישה חמורה באשר למשמעות המונחים הנפשיים. ייתכן שהפונקציונליזם העממי יכול לחמוק מן התוצאה הבעייתית שמונחים נפשיים רבים הם על פיו ריקים, אם יאמץ עמדה פחות חמורה. ואמנם, רבים מחסידי השקפה זו דוגלים בעמדה פחות חמורה באשר למשמעות המונחים הנפשיים. לפי לואיס, למשל, מצב נפשי – המצב עליו חל מונח נפשי מסוים – אינו בהכרח המצב אשר מקיים את **כל** האפיונים הפונקציונליים שמייחסת לו הפסיכולוגיה העממית. אין זה בהכרח "**הממלא המושלם**" של הפונקציות הסיבתיות שמייחסת הפסיכולוגיה העממית, אלא זהו "**הממלא הבלתי-מושלם**" של הפונקציות הללו. כוונת הדברים היא שמצב מסוים יצוין על ידי מונח נפשי מסוים גם אם מצב זה ממלא חלק ניכר מן הפונקציות הרלוונטיות אך לא את כולן (וייתכן גם שחלק מן הפונקציות שהוא ממלא אינן זהות לחלוטין לפונקציות שמייחסת הפסיכולוגיה העממית, אלא רק **בקירוב**).

לפי גרסה מקובלת של הפונקציונליזם העממי, מונחים נפשיים מציינים את המצבים שהם הממלאים – ולו בקירוב לפחות – חלק ניכר מן הפונקציות שמייחסת להן הפסיכולוגיה העממית, ולא בהכרח ממלאים במדויק את כל הפונקציות הללו. אפשר להביע רעיון זה במילים הללו: מונחים נפשיים מציינים את הממלאים הבלתי מושלמים של הפונקציות שמייחסת להם הפסיכולוגיה העממית.

מאחר שהשקפת הפונקציונליזם העממי, כאמור, אינה רק השקפה בדבר משמעויותיהם של המונחים הנפשיים – היא השקפה המתחייבת לקיומן של זהויות בין מצבים נפשיים לבין מצבים פונקציונליים מסוימים, ובין תכונות נפשיות לבין תכונות פונקציונליות מסוימות, הרי שגרסה זו שלה מתייחסת גם לקריטריון הזהות של מצבים נפשיים ותכונות נפשיות: על פיה, מה שמשייך מצב נפשי מסוים לסוג נפשי מסוים (דהיינו, לכאב, לפחד, לרצייה בשוקולד, וכדומה) – מה שמקנה לו את תכונתו הנפשית – זוהי העובדה שהוא ממלא בקירוב **חלק ניכר** מן הפונקציות שהפסיכולוגיה העממית מייחסת למצב הנפשי המצוין (פונקציות הקושרות בין מצבים נפשיים לבין קלטים חושיים, פלטים התנהגותיים, ומצבים נפשיים אחרים). אין הוא חייב למלא את כולן.

דומה שגרסה זו של הפונקציונליזם העממי אמנם חומקת מן הקושי הנזכר בדבר שקריותן של חלק מטענות הפסיכולוגיה העממית. גם אם חלק מטענות הפסיכולוגיה העממית הכרוכות במונח "פחד", למשל, הן שקריות – כלומר, גם אם אין מצב נפשי שכל הטענות הללו חלות עליו – עדיין יכול מונח זה לציין מצב נפשי: הוא מציין את אותו מצב נפשי אשר חלק ניכר מטענות אלו חלות עליו.

בהקשר זה מתעוררות שתי שאלות. האחת היא, איזה חלק מטענות הפסיכולוגיה העממית הכרוכות במונח נפשי מסוים חייבות להיות אמיתיות כדי שמונח זה יציין מצב נפשי ולא יהיה ריק? מה פירוש "חלק ניכר": כמעט כולן? רובן? במילים אחרות, האם אין הקריטריון המוצע עמום מדי? בנקודה זו מסתמכים אנשי הפונקציונליזם העממי על גישה מסוימת באשר לשפה בכלל, גישה באשר למשמעויותיהם של מונחים (לא רק מונחים נפשיים). לפי גישה זו, אין קריטריון חמור המכתיב בכל מקרה ומקרה תשובה לשאלה האם מונח מסוים חל או שאינו חל על דבר מסוים. אנו קושרים עם מונחינו תיאורים שונים (ובמילים אחרות, תכונות שונות), והמונחים חלים על אותם דברים המספקים חלק ניכר מן התיאורים הללו (ובלשון אחרת, שיש להם חלק ניכר מן התכונות הללו). אם יתברר, למשל, שאחד התיאורים שהמדענים קושרים עם המונח "אלקטרון" הוא שגוי, עדיין יציין מונח זה בפי המדענים את החלקיק הרלוונטי, וזאת משום שחלקיק זה מספק את רוב התיאורים שהמדענים קושרים עם המונח. כך, הטענה המרוככת של אנשי הפונקציונליזם העממי, על פיה די למצב נפשי להתאפיין על ידי חלק ניכר של הפונקציות הקשורות במונח נפשי מסוים כדי שמונח זה יציין מצב זה,

נראית טענה סבירה (ולא טענה שכל הגיונה הוא הצורך להתגבר על הקושי הכרוך בבעיית הריקות). והתשובה להשגה כי גישה זו כרוכה בקריטריון עמום לחלוקתם של מונחים נפשיים על מצבים נפשיים (מה פירוש "חלק ניכר"?): היא: באשר למונחים מכל הסוגים, דוברי השפה מסוגלים לקבוע במרבית המקרים האם אמנם מונח מסוים חל על דבר מסוים או לא גם כאשר מתברר שהדבר אינו מספק חלק מן התיאורים הקשורים במונח. עם זאת, ייתכנו מקרי גבול (מקרים שבהם דוברי השפה אינם מסוגלים להשיב לשאלה זו, או שדוברים שונים חלוקים בדעותיהם). כך, העובדה שאין בידינו לומר כמה פונקציות מן הפונקציות שקושרת הפסיכולוגיה העממית במונח נפשי חייב מצב למלא כדי שיצוין על ידי מונח זה אינה תופעה חריגה, אלא תופעה טבעית ורגילה שאינה בעייתית כלל. הקריטריון המוצע על ידי הפונקציונליזם העממי לחלוקתו של מונח נפשי על מצב נפשי הוא אכן עמום במידה מסוימת, אולם אין בכך פגם משום שעמימות זו היא מעצם טבעה של השפה.

השאלה השנייה עניינה **בדרגת השקירות** שסביר לייחס לפסיכולוגיה העממית. הגרסה הנדונה של הפונקציונליזם העממי חומקת מבעיית הריקות של המונחים הנפשיים גם אם טענות שונות של הפסיכולוגיה העממית הינן שקריות. אולם אם היקף הטענות השקירות הוא רב, השקפת הפונקציונליזם העממי לא תוכל לחמוק מבעיית הריקות. שהרי גם הגרסה הנדונה של השקפה זו מחייבת את אמיתותן של חלק ניכר מטענות הפסיכולוגיה העממית הקשורות במונח נפשי כלשהו כדי שמונח זה יציין מצב נפשי ולא יהיה מונח ריק. האם עומדת הפסיכולוגיה העממית בתנאי זה? לפחות על-פני הדברים, דוגלי הפונקציונליזם העממי יוכלו לטעון שאין להשקפתם כל בעיה בעניין זה. אמנם סביר מאוד שהפסיכולוגיה העממית, בהיותה תיאוריה עממית ולא תיאוריה מדעית שיטתית, כרוכה בטעויות שונות. עם זאת, יש יסוד סביר להניח שמרבית הטענות המרכזיות של הפסיכולוגיה העממית הן אמיתיות. זאת משום שבמידה רבה זוהי תיאוריה מוצלחת מבחינה מעשית. הצלחתה המעשית מתבטאת בכך שאנו מצליחים לחזות ולהסביר את התנהגות הזולת במקרים רבים מאוד, באופן יומיומי, על ידי כך שאנו מייחסים מצבים נפשיים ומפעילים את הכללות הפסיכולוגיה העממית. במקרים רבים מאוד אנו חוזים בהצלחה מה יעשו אנשים כשאנו מייחסים להם כאב ראש, קנאה, רצון להצליח בעסקים, ועוד ועוד. ההסבר הטבעי והמתבקש להצלחה זו הוא שטענות הפסיכולוגיה העממית הן בעיקרן אמיתיות; לולא היו אמיתיות בעיקרן, החיזויים הללו היו נכשלים. דומה אפוא שקשה להתנגד לטענה שטענות הפסיכולוגיה העממית אמיתיות בעיקרן. ואם כך, דומה שהפונקציונליזם העממי מתגבר על ההתנגדות בעניין ריקות המונחים הנפשיים.

נזכיר כי קיימת עמדה מנוגדת – עמדה שלפיה הפסיכולוגיה העממית שגויה באורח רדיקלי. לעמדה זו השלכות מרחיקות-לכת לא רק באשר לפונקציונליזם העממי, אלא באשר לפילוסופיה של הנפש בכלל, אולם הדיון בה חורג ממסגרתו של קורס זה.

עתה נעבור לדון בגרסה חלופית של ההשקפה הפונקציונליסטית. בהמשך נדון בהבטים אחרים של הפונקציונליזם העממי.

## 4.2 פונקציונליזם אמפירי

כאמור, לפי ההשקפה הפונקציונליסטית, מצבים נפשיים מאופיינים על פי תפקידיהם בחיי הנפשיים של הייצור הנפשי כפי שתפקיד זה מתואר על ידי התיאוריה הפסיכולוגית החלה עליו. לפיכך עולה השאלה: מהי, לפי ההשקפה הפונקציונליסטית, אותה תיאוריה פסיכולוגית האמורה להוות את הבסיס לאפיוניהם של המצבים הנפשיים? תשובת הפונקציונליזם העממי לשאלה זו, כפי שראינו, היא שזוהי התיאוריה של הפסיכולוגיה העממית. עתה נציג את התשובה החלופית, המוצעת על-יד השקפת **הפונקציונליזם האמפירי**, הקרויה לעתים "**פסיכופונקציונליזם**" ("psychofunctionalism") או "**פונקציונליזם מדעי**". תשובה זו היא, בפשטות, שהתיאוריה הפסיכולוגית המדוברת היא **הפסיכולוגיה המדעית**. לעתים מדובר על הפסיכולוגיה המדעית "**המיטבית**" או "**האידיאלית**".

כפי שראינו, התשובה של הפונקציונליזם העממי לשאלה הנדונה הייתה נעוצה בעובדה שהשקפה זו הייתה השקפה אנליטית, השקפה בדבר משמעויות המונחים הנפשיים. עובדה זו חייבה פנייה לפסיכולוגיה העממית, שטענותיה ידועות בעיקרן לכל מי שמסוגלים לייחס מצבים נפשיים ליצורים נפשיים. לעומת זאת, הפונקציונליסטים האמפיריים רואים את השקפתם **כהשערה אמפירית**. במובן זה, השקפה זו דומה בטבעה ל – למרות שהיא שונה בתוכנה מ – התורה של זהות-סוג. כזכור, התורה של זהות-סוג קדם-הניחה כי אפשרית **זהות סינתטית בין תכונות**. הזהות בין תכונות נפשיות לבין תכונות פיזיקליות על פיה היא אומנם זהות סינתטית, כלומר זהות הניתנת לגילוי אמפירי. הפונקציונליסטים האמפיריים מקבלים גם הם שתיתכן זהות סינתטית בין תכונות. אך הם חולקים על טענת הזהות הספציפית שבה דוגלת התורה של זהות-סוג. טענתם היא שההשערה שתכונות נפשיות זהות עם **תכונות פונקציונליות** של מצבי-מוח היא השערה אמפירית סבירה ביותר, ובין היתר היא סבירה יותר מן ההשערה שתכונות נפשיות זהות עם **תכונות פיזיקליות** של מצבי-מוח.<sup>53</sup>

### 4.2.1 טיעונו של פטנאם בזכות ההשקפה הפונקציונליסטית

נחזור לרגע לגישת הניטרליות הנושאת, שבה דנו ביחידה 4-5. לפי גישה זו, מצבים נפשיים שונים מאופיינים על-ידי התנאים הרגילים שבהם הם מתרחשים. התנסות אדום, למשל, מאופיינת על ידי סמארט כמצב שמתרחש בי כאשר מולי נמצאת, למשל, עגבנייה בשלה,

<sup>53</sup> הילארי פטנאם, שרבים רואים בו את אבי הפונקציונליזם בכלל, הוא המפתח החשוב ביותר של הפונקציונליזם האמפירי. ראו בעיקר H. Putnam, "The Nature of Mental States", in his *Mind, Language and Reality: Philosophical Papers*, Volume 2, Cambridge: Cambridge University Press, 1975. פילוסופים בולטים אחרים שהגנו על השקפה זו הם ג'רי פודור וגילברט הרמן. ראו: J. Fodor, *Psychological Explanations*. Cambridge (Mass.): MIT Press, 1968; Harman, G. *Thought*. Princeton: Princeton University Press, 1973. הרמן מפתח, בין היתר, רעיונות של ווילפריד סלארס, הנחשב למבשר חשוב של הפונקציונליזם. ראו: W. Sellars, "Empiricism and the Philosophy of Mind", in his *Science, Perception, and Reality*, London: Routledge and Kegan Paul, 1963. W. Sellars, "Some Reflections on Language Games", *Philosophy of Science* 21, 1954, pp. 204-228. וויליאם לייקן פיתח גרסה מיוחדת של השקפה זו שבה נדון בהמשך. ראו, בין היתר: W. Lycan, *Consciousness*, Cambridge (Mass.): MIT Press, 1987.

כשעיני פקוחות, התיאוריה טובה, וכדומה. אמרנו שגישה זו כשלעצמה אינה מחייבת פיזיקליזם-סוג, אלא מראה את אפשרותה של עמדה זו. מתי נוכל לומר, לפי גישה זו, כי פיזיקליזם-סוג לא רק אפשרי אלא אמתי? לפי גישת הניטרליות הנושאית, פיזיקליזם-סוג אמתי אם בתנאים המאפיינים התנסות אדם מתרחשת אצל המתנסים, בכל המקרים, אותה תכונה פיזיקלית; אם בתנאים המאפיינים התנסות ירוק מתרחשת אצל המתנסים, בכל המקרים, אותה תכונה פיזיקלית, וכך הלאה. אולם אם אין מתממשת בתנאים כאלו אצל המתנסים, בכל מקרה ומקרה, אותה תכונה פיזיקלית, הרי פיזיקליזם-סוג אינו אמתי. אם יש יסוד סביר להניח שבכל מקרה כזה אין מתממשת בתנאים הנדונים אותה תכונה פיזיקלית אולם מתממשת אותה תכונה פונקציונלית – אותה פונקציה הקושרת בין מצבים פנימיים, גירויים חושיים ותגובות התנהגותיות – הרי יש יסוד סביר להניח שההשקפה הפונקציונליסטית אמיתית, דהיינו שתכונות נפשיות הן תכונות פונקציונליות כאלו. השאלה מהן תכונות נפשיות, לפי גישה זו, היא שאלה אמפירית. טיעונו של פטנאם בזכות ההשקפה הפונקציונליסטית הוא ברוח זו. הרעיון הבסיסי של טיעון זה הוא שזוהי השערה אמפירית סבירה שהתכונה המשותפת למקרים בהם יצור נפשי חש כאב, למשל, היא תכונה פונקציונלית. אם טיעון זה מוצלח, אם כן, הוא מבסס פונקציונליזם אמפירי.<sup>54</sup>

דרך אחת שתאפשר לנו לדעת מה סוגה של התכונה (דהיינו, תכונה פיזיקלית, תכונה פונקציונלית, וכדומה) משותפת לכל המצבים הנפשיים מאותו סוג, היא לשאול, ראשית, כיצד אנו מגלים שיצור מסוים נמצא במצב נפשי מסוים. ובכן, אומר פטנאם, אנו מגלים שיצור מסוים נמצא במצב של כאב, או רעב, או כעס, וכדומה, על בסיס התנהגותו. התנהגויות מסוימות של יצורים נפשיים מהוות ראיות לכך שהיצורים הללו נמצאים במצבים נפשיים מסוימים. השאלה שלפנינו היא, אם כן, מהי התכונה המשותפת ליצורים המפגינים התנהגויות כהתנהגות כאב, התנהגות כעס, וכדומה.

אפשר לחשוב, שאם אנו מגלים שיצורים נפשיים נמצאים במצבים נפשיים מסוימים על-סמך התנהגויותיהם, הרי שהתכונה המשותפת ליצורים החולקים מצבים נפשיים זה-סוג אינה אלא תכונה התנהגותית או נטייה התנהגותית. כלומר, לטענה הביהביוריסטית שלהימצא במצב של כאב משמעו להיות בעל נטייה התנהגותית מסוימת יש לכאורה יתרון: היא מקבילה לאופן שבו אנו מגלים שליצורים נפשיים יש מצבים נפשיים מסוימים. יותר מכך, אנו מייחסים ליצורים שונים מצבים נפשיים בהיעדר ידע הן על המבנה החומרי של מוחותיהם, העומד ביסוד התנהגויותיהם, והן על הארגון הפונקציונלי העומד ביסוד התנהגויותיהם. הידע שלנו עניינו אך ורק בהתנהגויותיהם. אולם שיקולים אלו אינם יכולים לספק הצדקה לטענה הביהביוריסטית. חשבו על הדוגמה הבאה: העובדה שאנו מזהים נמר על פי חברבורות עורו אין משמעה שלהיות נמר משמעו להיות יצור בעל חברבורות כאלו. מאפיינים גנטיים מסוימים, ולא המאפיינים החיצוניים המשמשים אותנו לזיהויים של נמרים, הם אלו האחראים להיותו של יצור מסוים נמר (איננו יכולים לשלול את האפשרות של קיומם של

<sup>54</sup> טיעונו של פטנאם מופיע במאמרו "The Nature of Mental States" (הנזכר בהערה הקודמת).

נמרים לבקנים, למשל). פטנאם נעזר בדוגמה אחרת כדי להראות שמן השיקול האמור אי-אפשר לפסול לא את האופציה של פיזיקליזם-סוג ולא את האופציה הפונקציונליסטית: לפסול אופציות אלו משיקול זה כמוהו כלטעון שחומו של גוף איננו אנרגיה קינטית מולקולרית ממוצעת מאחר שאנשים רבים אינם יודעים דבר על האנרגיה הקינטית המולקולרית הממוצעת של גוף (כך הם חושבים) כשהם יודעים שהוא קר או חם. מן הסתם, אין בשיקול זה כדי לבסס את אמיתותה של האופציה הביהביוריסטית.

אולם לא רק שאין בשיקול האמור כדי לבסס את האופציה הביהביוריסטית, אלא שיש שיקולים רציניים כנגדה. ביחידה 3 הצגנו מספר טיעונים כנגד הביהביוריזם הפילוסופי, וביניהם גם את טיעון הסופר-ספרטאנים של פטנאם עצמו. בטיעון זה הראה פטנאם שיצור החש כאב עשוי שלא להפגין התנהגות כאב, ולפיכך שהתנהגות כאב אינה תנאי מספיק לכאב. פטנאם פוסל אפוא את האופציה הביהביוריסטית. מה באשר לאופציה על פיה בתנאים המאפיינים התנסות אדום, למשל, מתרחשת אצל המתנסים, בכל המקרים, אותה תכונה פיזיקלית? במקרה זה, כאמור, היה עלינו לומר שהתכונה של התנסות אדום היא תכונה פיזיקלית, ואם כך הדבר באשר לכל התכונות הנפשיות, היה עלינו להצדיק פיזיקליזם-סוג. אם נניח שיצורים שאנו מייחסים להם מצבים נפשיים זה-סוג חולקים מצבים פנימיים **בעלי אותן תכונות פיזיקליות**, נוכל להסביר את התנהגויותיהם הדומות (ההתנהגויות ששימשו לנו כראיות בזכות ייחוס המצבים הנפשיים) – למצבים הזהים מבחינה פיזיקלית יש, באותם תנאים, אותן תולדות. אולם למעשה כבר ראינו את נימוקו של פטנאם כנגד אופציה זו. הצגנו ביחידה הקודמת את טיעון ריבוי המימושים של פטנאם. כזכור, טענתו של פטנאם הייתה שאין זה סביר (למרות שאין לשלול את עצם האפשרות) שכל המצבים הנפשיים מאותו סוג – לדוגמה, כאבים – חולקים באופן ייחודי אותה תכונה פיזיקלית. לפיכך אין זה סביר שהעמדה של פיזיקליזם-סוג אמתית.

פטנאם פוסל אפוא הן את האופציה הביהביוריסטית והן את האופציה הפיזיקליסטית כתשובות לשאלה מהי התכונה המשותפת לכל היצורים הנמצאים במצבים נפשיים מאותו סוג. האם זה סביר שכל המצבים הנפשיים מאותו סוג חולקים באופן ייחודי **אותה תכונה פונקציונלית**? טענתו של פטנאם היא שאכן זוהי אפשרות סבירה מאוד. ברור, אומר פטנאם, שזהות התנהגותית של שתי מערכות מהווה טעם לשער שהמערכות הללו זהות מבחינת ארגון הפונקציונלי – שהן זהות מבחינת הקשרים הפונקציונליים בין מצביהן. אם שתי מערכות מתנהגות באופנים דומים – דגמי הגירוי-תגובה שלהן דומים אלו לאלו, וזאת לא במקרים ספורים אלא באופן שיטתי, סביר שהארגון של מצביהם הפנימיים – הקשרים ביניהם, וביניהם לבין הגירויים והתגובות – אחראי לכך. במילים אחרות, זוהי השערה אמפירית סבירה ביותר שארגון הפונקציונלי של מערכות זהות מבחינה התנהגותית גם הוא זהה; סביר שהמצבים האחראים להתנהגויות הללו זהים מבחינת קשריהם עם מצבים נפשיים אחרים, גירויים ותגובות – יש זהות במאפייני "התוכנה". אם כן, זוהי השערה אמפירית סבירה שזהות נפשית היא זהות פונקציונלית – שהחקירה הפסיכולוגית תחשוף כי מצבים נפשיים המסווגים כמצבים השייכים לאותו סוגים – כאב, ההאמנה שיורד גשם,



חרדה, רצון לטוס לאלסקה, וכדומה – שייכים לאותם סוגים פונקציונליים. התכונות הפונקציונליות של המצבים הנפשיים הן אלו האחראיות לסיווגם הנפשי; התכונות הפונקציונליות הן הן התכונות הנפשיות.

לפי טיעון זה, אם כן (שבהמשך נתייחס לשאלת כוחו), ההשערה הפונקציונליסטית – ההשערה בדבר זהותן של תכונות נפשיות עם תכונות פונקציונליות – היא השערה אמפירית סבירה. זוהי השערה סבירה, כאמור, שהפסיכולוגיה המדעית תגלה זהויות בין סוגים נפשיים או תכונות נפשיות לבין סוגים פונקציונליים או תכונות פונקציונליות. כך, היא תחשוף את טבען האמתי של התכונות הנפשיות, שהוא טבע פונקציונלי (הוא אינו טבע פיזיקלי ואינו טבע התנהגותני). כאמור, לפי ההשקפה הפונקציונליסטית מצבים נפשיים מאופיינים על פי תפקידיהם בחייו הנפשיים של היצור הנפשי כפי שתפקיד זה מתואר על ידי התיאוריה הפסיכולוגית החלה עליו. עתה אנו יכולים לומר כי לפי הפונקציונליזם האמפירי התיאוריה הפסיכולוגית המדוברת היא התיאוריה הפסיכולוגית המדעית, התיאוריה העולה מן החקירה הפסיכולוגית האמפירית.

לפי הפונקציונליזם האמפירי, מצבים נפשיים מאופיינים על ידי מיקומם ברשת הקשרים המתוארת על ידי הפסיכולוגיה המדעית, קשרים אשר שוררים בין המצבים הנפשיים עצמם, ביניהם לבין קלטים חושיים, וביניהם לבין פלטים התנהגותיים.

#### 4.2.2 מהותו של הפונקציונליזם האמפירי

מאחר שלפי הפונקציונליזם האמפירי הזהויות בין תכונות נפשיות לבין תכונות פונקציונליות ניתנות לגילוי אמפירי בלבד, ברור שאלה הן זהויות סינתטיות, כדוגמת הזהויות בין תכונות נפשיות לבין תכונות פיזיקליות שלהן טוענים אנשי פיזיקליזם-סוג. השקפת הפונקציונליזם האמפירי אינה טוענת שטענות התיאוריה הפסיכולוגית המאפיינות את המצבים הנפשיים קובעות את משמעויות המונחים הנפשיים; היא טוענת שמונחים נפשיים מציינים (מורים על) תכונות פונקציונליות (מונחים אלו מציינים, כמובן, תכונות נפשיות, ואלו זהות לפי השקפה זו לתכונות פונקציונליות), אבל היא איננה השקפה על אודות משמעויות המונחים הנפשיים – אין היא השקפה אנליטית. בניגוד להשקפת הפונקציונליזם העממי, השקפת הפונקציונליזם האמפירי היא השקפה על אודות התכונות הנפשיות בלבד.

השקפת הפונקציונליזם האמפירי איננה השקפה על אודות משמעויות המונחים הנפשיים. היא השקפה על אודות התכונות הנפשיות בלבד. מונחים נפשיים מורים על פיה על תכונות פונקציונליות (משום שתכונות נפשיות הן בפועל תכונות פונקציונליות), אבל טענות הפסיכולוגיה המדעית על אודות התכונות הללו אינן קובעות את משמעויות המונחים הנפשיים.

אנו מזהים מצב ככאב לפי ההתנהגות האופיינית הכרוכה בו, ואולי לפי התפקיד הסיבתי "העממי" המיוחס לו – הפונקציונליסטים האמפיריים אינם מכחישים זאת; אבל הידע העממי ההתנהגותי אינו ממצה והתיאור העממי הפונקציונלי אינו בהכרח מדויק באשר לטבעם האמתי של כאבים. לפי גישה זו, הצלחתנו בייחוס כאבים לבני אדם (הצלחה המתאפשרת על-ידי שליטתנו בפסיכולוגיה העממית והבנתנו את משמעויותיהם של המונחים הנפשיים) מתיישבת עם העובדה שאנו עשויים לטעות, ואפילו במידה ניכרת, באשר לטבעם האמתי של כאבים. גילוי טבעם האמתי של המצבים הנפשיים הוא עניין מדעי ולא עממי – הוא עניין לפסיכולוגיה המדעית. זוהי כאמור השערה אמפירית סבירה, טוענים הפונקציונליסטים האמפיריים, שהטבע האמתי של המצבים הנפשיים הוא טבע פונקציונלי (שכל הכאבים, למשל, חולקים באופן ייחודי תכונה פונקציונלית מסוימת), אולם זהו עניין למדע לקבוע את הטבע הפונקציונלי הספציפי של כאבים, פחדים, וכדומה. המדע הרלוונטי הוא, כמובן, החקירה הפסיכולוגית המדעית. ניתן לראות כי שאלת קיומו של פער בין התיאור הפונקציונלי העממי לבין התיאור הפונקציונלי האמתי של מצבינו הנפשיים כלל אינה מאיימת על השקפת הפונקציונליזם האמפירי, מאחר שאין היא מחויבת כלל לאמתות התיאור הפונקציונלי העממי.

אמרנו, שלפי הפונקציונליזם האמפירי, הטבע הפונקציונלי של המצבים הנפשיים אשר ניתן לגילוי על ידי החקירה הפסיכולוגית המדעית הוא הטבע האמתי של המצבים הנפשיים. האם פירוש הדבר שטבעם של המצבים הנפשיים תלוי בשאלה מהי התיאוריה הפסיכולוגית המדעית הקיימת? פירוש זה של השקפת הפונקציונליזם האמפירי אינו סביר.

הפירוש הסביר של השקפת הפונקציונליזם האמפירי (למרות שלרוב אין הדבר נזכר במפורש) הוא זה המזהה את התכונות הנפשיות עם התכונות הפונקציונליות אשר באמת מאפיינות את המצבים הנפשיים, ולא עם התכונות הפונקציונליות המיוחסות להם על ידי תיאוריה פסיכולוגית כלשהי. הסימטריה הצורנית בין השקפת הפונקציונליזם האמפירי לבין השקפת הפונקציונליזם האנליטי (האמורה להתבטא בכך שכל אחת מן ההשקפות פונות לתיאוריה פסיכולוגית לשם אפיון המצבים הנפשיים) אינה מתקיימת אפוא בעניין זה; באופן שבו מאופיינת ההשקפה של פונקציונליזם אמפירי – כהשקפה הפונה אל ממצאי החקירה הפסיכולוגית – יש כדי להטעות, שכן למעשה החקירה הפסיכולוגית אינה רלוונטית במישרין לאפיונה הנכון של השקפה זו.

לעתים מאופיינת השקפת הפונקציונליזם האמפירי כהשקפה שעל פיה התפקיד הפונקציונלי הקובע את סיווגם של המצבים הנפשיים ומכונן את התכונות הנפשיות הוא זה אשר יתואר על ידי התיאוריה הפסיכולוגית **הטובה ביותר שבאפשר**, או התיאוריה הפסיכולוגית **האידיאלית**. לפי השקפה זו, מטרת התיאוריה הפסיכולוגית היא לתאר נכונה את התכונות הנפשיות (את הפונקציות הנפשיות), אולם מה שקובע את זהות התכונות הנפשיות אינו ממצאה של חקירה פסיכולוגית כלשהי אלא המציאות הפסיכולוגית – התכונות הנפשיות הן התכונות הפונקציונליות שיש **בפועל** למצבים הנפשיים. הטבע האמיתי של כאב הוא מילוי תפקיד סיבתי מסוים בחיים הנפשיים של הפרט, בדיוק כשם שהטבע האמיתי של מים הוא היותם מורכבים ממולקולות של  $H_2O$ , הטבע האמיתי של זהב הוא היותו בעל מספר אטומי 79, וכדומה. וכשם שבמקרים של המים, הזהב, ודומיהם, המדע אמנם שואף לגלות את הטבע האמיתי אבל הטבע האמיתי אינו תלוי בתגלית המדעית (מים הם  $H_2O$  אף אם איש לא היה יודע זאת), כך, לפי הפונקציונליזם האמפירי, הפסיכולוגיה אמנם שואפת לגלות את הטבע האמיתי של כאב ומצבים נפשיים אחרים (שהוא, לפי השקפה זו, טבע פונקציונלי מסוים), אבל הטבע האמיתי אינו תלוי בהצלחתה של הפסיכולוגיה לחשוף טבע זה.

גם תחת פשר זה של השקפת הפונקציונליזם האמפירי, משמרת השקפה זו היבט אמפירי חשוב (וכך היא ראויה לשמה), משום שעל פיה, בניגוד לפונקציונליזם העממי, הטענה שתכונות נפשיות הן תכונות פונקציונליות היא בגדר השערה אמפירית, השערה שממצאי החקירה הפסיכולוגית עשויים לאששה. התיאוריה הפסיכולוגית המדעית, ניתן לומר, היא זו שאמורה לפרוע את השטר הפונקציונליסטי. ואם ברצוננו לדעת מהן התכונות הנפשיות, על פי פשר זה, כמו על פי כל פשר של פונקציונליזם אמפירי, הדרך הטובה ביותר שתקרב אותנו למטרה זו היא ללמוד את ממצאי הפסיכולוגיה המדעית.

הפונקציונליזם האמפירי גם קשוב באופן מיוחד אל החקירה המדעית, האמפירית, של הנפש. ענפיה השונים של חקירה זו, ובמיוחד הפסיכולוגיה הקוגניטיבית וענפים אחרים של המדעים הקוגניטיביים, נבדלים מן הפסיכולוגיה העממית לא רק בכך שהם מגלים ומניחים קשרים סיבתיים השונים מאלו בהם מדובר בפסיכולוגיה העממית. גם טווח המצבים הנפשיים עצמם שאליהם הם מתייחסים שונה מטווח המצבים הנפשיים אליהם מתייחסת הפסיכולוגיה העממית. המדעים הקוגניטיביים מסבירים תהליכים פסיכולוגיים במונחים של **עיבוד מידע**: הם מתארים את סוגי המידע שאנו קולטים, כיצד אנו קולטים אותם, כיצד אנו שומרים את המידע הנקלט, כיצד אנו מאחדים מידע המגיע מערוצים שונים, כיצד אנו שולפים מידע מן הזיכרון, וכדומה. כאשר מתארים המדעים הקוגניטיביים תהליכי עיבוד מידע, הם נזקקים להניח את קיומם של מצבים נפשיים שונים שחלקם אינו זהה למצבים הנפשיים המוכרים לכולנו ואשר אליהם מתייחסת הפסיכולוגיה העממית, כמו האמנות ורציות. הם מתייחסים גם למצבים בלתי מודעים (מצבים שליזור הנפשי אין גישה אפיסטמית ישירה אליהם), אשר מאופיינים על ידי סוג המידע שהם נושאים, האופן שבו המידע הזה מקודד בהם, תהליכי עיבוד המידע שבהם הם נוטלים חלק, וכדומה. המדעים הקוגניטיביים מניחים את קיומם של מצבים כאלו, ובעזרת ההנחות בדבר קיומם של המצבים הללו הם מסבירים את תפקודינו

הפסיכולוגיים. אפשר לומר שאלה הם מצבים "תיאורטיים", מצבים שהתיאוריה מניחה כחלק מן המנגנון ההסברי שלה. ובכן, השקפת הפונקציונליזם האמפירי מתייחסת גם למצבים "תיאורטיים" כאלו. קדם-ההנחה של השקפה זו היא שמצבינו הנפשיים כרוכים ברשת קשרים סיבתיים מורכבת ועשירה, הכוללת גם מצבים כאלו, מצבים השונים ממצבי הפסיכולוגיה העממית המוכרים לכולנו. ההשערה האמפירית שלה היא שמצב נפשי מאופיין על ידי מיקומו ברשת כזו, כך שאפיונו כולל התייחסות גם למצבים נפשיים "תיאורטיים". במובן זה, רשת הקשרים שבמונחיה מאפיינת השקפה זו את המצבים הנפשיים היא רשת מעין זו המונחת על ידי המדעים הקוגניטיביים. למעשה, הפסיכולוגיה הקוגניטיבית וענפיה אחרים של המדעים הקוגניטיביים היוו מודל חשוב בהתפתחות ההשקפה הפונקציונליסטית.

רשת הקשרים שבמונחיה מאפיינת השקפת הפונקציונליזם האמפירי את המצבים הנפשיים היא רשת מעין זו המונחת על ידי הפסיכולוגיה הקוגניטיבית וענפיה אחרים של המדעים הקוגניטיביים. רשת זו כוללת לא רק מצבים נפשיים המוכרים לנו (ואשר אליהם יש לנו גישה אפיסטמית ישירה), אלא גם מצבים נפשיים "תיאורטיים".

לעתים מתארים את רשת המצבים והקשרים הזו, אותה שואף המחקר הפסיכולוגי לגלות, כ**פסיכולוגיית העומק** (depth psychology). עומקה של פסיכולוגיה זו מתבטא בכך שהיא עוסקת במצבים החורגים מעבר למצבים הנגישים לכולנו והידועים לכולנו. השקפת הפונקציונליזם האמפירי מהווה אפוא השערה באשר לפסיכולוגיית העומק של יצורים נפשיים. היא מזהה את מצביהם הנפשיים עם אפיונים פונקציונליים של פסיכולוגיית העומק שלהם. לפיה, מצבים נפשיים המשתייכים לאותו סוג פונקציונלי במונחי הפסיכולוגיה העממית (אותה ניתן לתאר כ**פסיכולוגיית פני-השטח**) אך נבדלים במאפיינים הפונקציונליים ברמת פסיכולוגיית העומק אינם שייכים לאותו סוג נפשי. לפי השקפה זו, התנאים לחלותו של מאפיין נפשי מסוים (כלומר, התנאים לכך שיצור נפשי מסוים נמצא במצב של כאב, שהוא מאמין שכדור האץ מתחמם, וכדומה) – תנאים ברמה של פסיכולוגיית העומק – הם תנאים מפורטים ומורכבים מאוד. לפי השקפה זו, הקריטריון לחלותו של מאפיין נפשי כלשהו – הקריטריון לזהות סוג של מצבים נפשיים – הוא קריטריון תובעני. בהמשך נראה שיש הטוענים שקריטריון זה של הפונקציונליזם האמפירי הוא תובעני מידי.

כאשר עומדים על האופי האמפירי של השקפת הפונקציונליזם האמפירי, ובמיוחד כאשר מנגידים אותו עם האופי האנליטי-לוגי של הפונקציונליזם העממי, יש לשים לב שלא להיגרר לטעות (שאליה נגררו פילוסופים שונים). על פי השקפת הפונקציונליזם האמפירי אמנם אין ה**כרח לוגי** שהמונחים הנפשיים יורו על התכונות הפונקציונליות שעליהן הם מורים בפועל, או על תכונות פונקציונליות בכלל; שהרי אפשרי היה מבחינה לוגית, לפי השקפה זו, שלכל הכאבים יש דווקא טבע פיזיקלי משותף – הפונקציונליסטים האמפיריים אינם טוענים שאין

זה אפשרי, אלא שבעניין שבעובדה אין זה נכון. תכונות נפשיות הן בפועל תכונות פונקציונליות, אולם הן אינן **חייבות** להיות תכונות פונקציונליות. לעומת זאת, בכיוון ההפוך מתקיים לפי ההשקפה הנדונה קשר לוגי: הפונקציונליסטים מחויבים לכך שהתממשותן של תכונות פונקציונליות מסוימות קובעת, בהכרח לוגי, את התממשותן של תכונות נפשיות. התממשותה של הפונקציה בה כרוך כאב, למשל, מחייבת מבחינה לוגית את התממשותו של כאב. כל יצור – למעשה, כל יצור **אפשרי** – המממש פונקציה זו, חש כאב. למצב המממש פונקציה זו תהיינה בהכרח התכונות שיש לכאבים.

מחויבות זו של הפונקציונליסטים – מחויבות המשותפת לכל הזרמים של הפונקציונליזם הפיזיקליסטי, ולא רק לגרסה האמפירית של השקפה זאת – היא מחויבות חזקה, ובהמשך דרכנו ניתקל בטענות על פיהן היא מחויבות חזקה מידי, שכן הפונקציונליסטים אינם מסוגלים להסביר כיצד מאפיינים שונים של מצבינו הנפשיים נקבעים על ידי תכונות פונקציונליות גרידא. אולם בשלב זה נציין את האופן בו מחויבות זו באה לידי ביטוי בהקשר של פונקציונליזם פיזיקליסטי. כזכור לנו מיחידה 4-5, כל התיאוריות הפיזיקליסטיות (ולמעשה כל התיאוריות המטריאליסטיות) מחויבות לכך שהתממשותן של תכונות פיזיקליות מסוימות קובעת, בהכרח לוגי, את התממשותן של תכונות נפשיות מסוימות. מן הסתם, פונקציונליזם פיזיקליסטי – בין אם עממי ובין אם אמפירי – מחויב לכך.<sup>55</sup> אמרנו גם שהתממשותן של תכונות פיזיקליות מסוימות קובעת, מבחינה לוגית, את התממשותן של תכונות פונקציונליות מסוימות (היזכרו, למשל, בתכונה של היות משאבה או בתכונה של היות מדחום). אין פירוש הדבר שכל התכונות הפונקציונליות הקיימות בעולם נקבעות על-ידי תכונות פיזיקליות – אם קיימות תכונות רוחניות, סביר שהן קובעות פונקציות מסוימות (הדואליסטים מכירים, כמובן, בכך שמצבים נפשיים ממלאים פונקציות שונות). אולם מי שדוגלים ב**אונטולוגיה פיזיקליסטית**, שלפיה אין בעולם התממשויות של תכונות רוחניות – מחויבים לכך שהתממשותן של **כל** התכונות הפונקציונליות נקבעת, מבחינה לוגית, על ידי התממשותן של תכונות פיזיקליות. על פי פונקציונליזם פיזיקליסטי אפוא התממשות התכונות הנפשיות נקבעת, מבחינה לוגית, על ידי התממשות של תכונות פיזיקליות. התוצאה המתקבלת היא שהתכונות הנפשיות נקבעות מבחינה לוגית על ידי תכונות פיזיקליות (אם  $X$  נקבע על ידי  $Y$  ו- $Y$  נקבע על ידי  $Z$ , הרי ש- $X$  נקבע על ידי  $Z$ ). הנפשי בתמונה זו נקבע על ידי הפיזיקלי, אולם הוא נקבע על ידי הפיזיקלי משום שהפיזיקלי מממש את הפונקציונלי. אמנם הנפשי איננו נקבע על ידי הפיזיקלי בשל טבעו הפיזיקלי (זה הלקח של טיעון ריבוי המימושים), אולם הפונקציונליזם מסביר כיצד בכל-זאת יכול הנפשי להיקבע על ידי הפיזיקלי; הוא נקבע על ידיו משום שלפיזיקלי יש גם יסוד מופשט, ובאופן ספציפי, יסוד פונקציונלי (כמובן, תקפותו של הסבר זה תלויה ביכולתו של הפונקציונליזם להתמודד עם

<sup>55</sup> אנו מדברים על פונקציונליזם פיזיקליסטי עממי ועל פונקציונליזם פיזיקליסטי אמפירי. ההבחנה בין פונקציונליזם פיזיקליסטי לבין פונקציונליזם דואליסטי וההבחנה בין פונקציונליזם עממי לבין פונקציונליזם אמפירי אינן תלויות זו בזו, כך שעשויות להיות (ואמנם קיימות) גרסה עממית של פונקציונליזם פיזיקליסטי וגרסה אמפירית של פונקציונליזם פיזיקליסטי, כשם שעשויות להיות גרסה עממית של פונקציונליזם דואליסטי וגרסה אמפירית של פונקציונליזם דואליסטי.

הביקורות שעל פיהן לא ניתן להסביר כיצד מאפיינים שונים של מצבינו הנפשיים נקבעים על ידי תכונות פונקציונליות גרידא. נדון בביקורות אלו בבוא הזמן).

על פי פונקציונליזם פיזיקליסטי – בין אם עממי ובין אם אמפירי – התממשות התכונות הנפשיות נקבעת, מבחינה לוגית, על ידי התממשות של תכונות פונקציונליות, הנקבעת בתורה, מבחינה לוגית, על ידי התממשות של תכונות פיזיקליות.

### 4.3 השוואות נוספות בין פונקציונליזם עממי לפונקציונליזם אמפירי

פילוסופים הדוגלים בפונקציונליזם האמפירי נוטים לדבוק במודל המחשב של הנפש יותר מאשר עמיתיהם הדוגלים בפונקציונליזם העממי (פטנאם, כזכור, הוא מן המפתחים החשובים הן של מודל המחשב והן של הפונקציונליזם האמפירי). אולם חשוב לציין שמודל המחשב ניתן להחלה לגבי שתי הגרסאות של הפונקציונליזם. על פי שתיהן ניתן לומר שהנפש היא תוכנת הגוף.

#### שאלה

האם הגרסאות האלה תיתנה אותה תשובה לשאלה באיזו תוכנה מדובר?

הגרסאות האלה נבדלות באשר לשאלה באיזו תוכנה בדיוק מדובר: האם התוכנה הרלוונטית היא זו המשתקפת בקשרים שקושרת הפסיכולוגיה העממית בין מצבינו הנפשיים לבין עצמם ובינם לבין הקלט החושי והפלט ההתנהגותי, או זו המשתקפת בקשרים הקיימים בפועל – ואשר משתקפים בפסיכולוגיה המדעית האידיאלית – בין מצבינו הנפשיים לבין עצמם (כשקבוצת המצבים הנפשיים במקרה זה היא, כאמור, רחבה יותר) ובינם לבין הקלט החושי והפלט ההתנהגותי? תוכנות אלו צפויות להיות שונות זו מזו. אולם כמובן, הבדל זה מעיד על-כך ששתי הגרסאות יכולות לקבל את רעיון הנפש כתוכנת הגוף.

באופן דומה, פילוסופים הדוגלים בפונקציונליזם העממי נוטים להתייחס לטכניקת הרמזיפיקציה כדי להסביר את השקפתם (לואיס, מן הבולטים בתומכי הפונקציונליזם העממי, הוא כזכור מי שפיתח טכניקה זו והחילה על אפיון המצבים הנפשיים), אולם גם פילוסופים הדוגלים בפונקציונליזם האמפירי יכולים להיעזר (ולעתים נעזרים) בטכניקה זו. כל עוד מתקיימת "עלילה" הקושרת בין המצבים הנפשיים לבין הקלטים החושיים והפלטם ההתנהגותיים ואשר מייחדת את זהותם של המצבים הנפשיים, ניתן לאפיין את המצבים הנפשיים באמצעות קשריהם למצבים נפשיים אחרים, לקלטים ולפלטם, מבלי להזכיר אף

מצב נפשי כמצב נפשי. אין זה משנה אם העלילה הנדונה היא העלילה של הפסיכולוגיה העממית, או העלילה של הפסיכולוגיה המדעית.

עמדנו על שתי נקודות משותפות לשתי הגרסאות של ההשקפה הפונקציונליסטית שבהן דנו עד כה. עתה נחدد את ההבדלים החשובים בין גרסאות אלו, וכך נבהיר טוב יותר את טיבן. הגרסאות הללו שונות זו מזו באשר למחויבויות שלהן. מחויבותה של האחת חזקה יותר בממד אחד, ומחויבותה של השנייה חזקה יותר בממד אחר. לגרסה האמפירית יש מחויבות אמפירית משמעותית: היא מחויבת להשערה שלכל הכאבים, למשל, משותפת אותה פסיכולוגית-עומק; משותף להם אותו אפיון פונקציונלי ברמת פסיכולוגית-העומק (הרמה אותה תחשוף הפסיכולוגיה המדעית המוצלחת).

## שאלה

מהי מחויבותה המקבילה של הגרסה העממית?

ובכן, לגרסה העממית אין כל מחויבות דומה. כשלעצמה, גרסה זו אינה מתחייבת לכך שחקירה מדעית – אפילו חקירה מדעית אידיאלית – תגלה שכל הכאבים הם בעלי טבע פיזיקלי משותף, או שהם בעלי טבע פונקציונלי משותף. טענתה אינה רלוונטית כלל לרמות העומק שאותן חוקר המדע (לא לרמה הפיזיקלית-נוירולוגית ולא לרמה של פסיכולוגית העומק); טענתה של גרסת הפונקציונליזם העממי היא ברמת פני-השטח – הרמה שבה אנו מזהים מצבים נפשיים בחיי היומיום שלנו, במסגרת הפרקטיקה של הפסיכולוגיה העממית, ואין היא חורגת כלל מרמה זאת. היא מחויבת לכך שלכל הכאבים משותף האפיון הפונקציונלי העולה מטענות הפסיכולוגיה העממית, אולם באשר לשאלה מה משותף להם (אם בכלל) ברמת עומק מדעית כלשהי, היא אינה מחויבת לכל תשובה שהיא. היא מתיישבת עם האפשרות שלכל הכאבים אותו טבע פונקציונלי ברמת העומק (טבע פונקציונלי הניתן לגילוי על ידי הפסיכולוגיה המדעית), עם האפשרות שלכל הכאבים של בני-האדם אותו טבע פיזיקלי, ואף עם האפשרות שלכל הכאבים בכלל (לאו דווקא האנושיים) אותו טבע פיזיקלי. אולם על פיה ההכרעה בשאלה אם אפשרות כלשהי מאלו מתקיימת, אינה רלוונטית לשאלה המושגית של מה עושה כאבים לכאבים.<sup>56</sup> ההכרעה בשאלה של מה עושה כאבים לכאבים ניתנת על ידי גרסה זו ברמת פני השטח: מה שעושה כאבים לכאבים זוהי העובדה שיש להם תפקיד פונקציונלי ("שטחי") מסוים, התפקיד הפונקציונלי העולה מטענות הפסיכולוגיה העממית. תשובה זו אינה עולה מחקירה אמפירית ואינה השערה אמפירית: זוהי תשובה אנליטית, תשובה העולה מניתוח משמעות המונח "כאב".

לעומת השקפת הפונקציונליזם העממי, השקפת הפונקציונליזם האמפירי כשלעצמה אינה כרוכה בעמדה מסוימת באשר למשמעות המונחים הנפשיים. יחד עם זאת, ניתן לומר יש לה

<sup>56</sup> לפילוסופים שונים הדוגלים בפונקציונליזם עממי עמדות שונות באשר לשאלה בדבר המשותף ברמת העומק המדעית למצבים נפשיים זה-סוג.

מחויבות מסוימת ומוגבלת בעניין זה, מחויבות המתבטאת בכך שהיא אינה מתיישבת עם כל עמדה אפשרית באשר לממד המשמעות של המונחים הנפשיים. למשל, היא אינה מתיישבת עם העמדה של הפונקציונליזם העממי באשר לממד זה, משום שלו הייתה מאמצת עמדה זו – שלפיה משמעותו של מונח נפשי היא התיאור הפונקציונלי העממי הקשור בו – הייתה למעשה מחויבת לכך שמה שעושה, למשל, כאב לכאב, זו העובדה שכאב כרוך באפיון פונקציונלי **עממי** מסוים. אולם לא ייתכן שהשקפת הפונקציונליזם האמפירי תהיה מחויבת לכך, משום שהיא מחויבת לתשובה אחרת לשאלות כמו מה עושה כאב לכאב: על פיה מה שעושה כאב לכאב זהו אפיונו הפונקציונלי **ברמת פסיכולוגיית העומק**. מטעמים דומים, השקפת הפונקציונליזם האמפירי אינה מתיישבת עם גישות נוספות לשאלת משמעותם של המונחים הנפשיים, אך נראה שהיא מסתדרת היטב עם גישות אחרות לשאלה זו (שאמנם על פי רוב מאומצות על ידי הדוגלים בה). אולם הדיון בנושא זה (הכרוך בהתייחסות לנושאים סבוכים בפילוסופיה של הלשון) חורג ממסגרת דיונונו הנוכחי.<sup>57</sup>

כאמור, לפי השקפת הפונקציונליזם האמפירי הקריטריונים לחלותם של מאפיינים נפשיים על יצורים – אפיונים פונקציונליים מורכבים ברמת פסיכולוגיית העומק – הם קריטריונים תובעניים. לעתים טוענים שבתובענות זו גלומה בעיה. נראה שהפונקציונליזם האמפירי מחויב להשערה שהמצב המזוהה בפרקטיקה העממית ככאב, למשל, כרוך באפיון פסיכולוגיית עומק זהים לחלוטין בכל מקרה ומקרה – הן כאבים שונים של אותו יצור נפשי, והן כאביהם של יצורים נפשיים שונים, קשורים באותם קשרים באותה רשת סבוכה של מצבים נפשיים. האם זוהי השערה סבירה? האם אין זה סביר שקיימים הבדלים – ולו קטנים – בין אפיוני פסיכולוגיית העומק של כאביהם או האמנותיהם של יצורים שונים? בעיה זו כרוכה בבעיה כללית יותר שהועלתה כנגד הפונקציונליזם ושנדון בה בהמשך (בעיית זהות התוכנה). כאן נסתפק ונאמר שבפני הפונקציונליסטים האמפיריים פתוחה הדרך להתפשר ולהחליש במידה מסוימת את הקריטריונים לחלותם של מאפיינים נפשיים. פשרה זו עשויה להיות דומה לפשרת "הממלא הבלתי-מושלם" של הפונקציונליזם העממי שתוארה קודם. באופן דומה, גם הפונקציונליסטים האמפיריים יכולים לומר שכדי ששני מצבים נפשיים ישתייכו לאותו סוג נפשי, הזהות הפונקציונלית ביניהם ברמת פסיכולוגיית העומק אינה חייבת להיות מושלמת – די בזהות מקורבת. מה שעושה כאבים לכאבים, לפי גישה מתוקנת זו, הוא שהם ממלאים **בקירוב** תפקיד פונקציונלי מסוים ברמת פסיכולוגיית העומק.

<sup>57</sup> נעיר גם, שהפונקציונליסטים האמפיריים יכולים (ויש הטוענים אף שהם חייבים) לקבל את הטענה שהאפיונים העממיים של המצבים הנפשיים הם המשמשים אותנו **לזיהוי** של מצבים אלו. ברוח זו, יש ביניהם כאלו המקבלים שלאפיונים אלו תפקיד הקשור במשמעות המונחים הנפשיים (הם קובעים על אילו מצבי מורים מונחים אלו), למרות שאינם מכוננים את משמעותם. גם הנושא של קביעת הוראה להבדיל מכינון משמעות הוא נושא בפילוסופיה של הלשון החורג ממסגרת דיונונו הנוכחי.



## 5. אפיון הקלט והפלט על פי הפונקציונליזם; ליברליזם ושוביניזם

כפי שנאמר ביחידה 4-5, תורה בפילוסופיה של הנפש, שממנה נובע שרק יצורים הדומים לנו, בני האדם, מאיזושהי בחינה – רק יצורים השייכים ל"לאום" שלנו, "הלאום האנושי", או הדומים לנו במידה מספקת מבחינה כלשהי – עשויים לאחוז במצבים נפשיים הדומים לאלו שלנו, מכונה "תורה שוביניסטית". הבעיה שחשף פטנאם באשר לפיזיקליזם-סוג היא שעמדה זו לוקה ב"שוביניזם חומרי" – רק מי שדומים לנו מבחינה חומרית עשויים, על פיה, להידמות לנו מבחינה נפשית, והרי סביר שגם ליצורים השונים מאתנו מבחינה חומרית עשויים להיות מצבים נפשיים הדומים לאלו שלנו. הליקוי של "שוביניזם יתר" (שוביניזם בלתי-מוצדק) באשר לתורות בפילוסופיה של הנפש הוא מקרה פרטי של ליקוי כללי יותר שתורות אחדות כאלו עשויות לעתים ללקות בו, והוא מחויבותן לשלול ייחוס מצבים נפשיים ליצורים כלשהם להם סביר לייחס מצבים נפשיים. קיים גם ליקוי נגדי, הליקוי של "ליברליזם" או "ליברליזם יתר". תורה בפילוסופיה של הנפש נחשבת ליברלית אם היא מחויבת לייחוס מצבים נפשיים לטווח גדול של יצורים או מערכות, והיא לוקה בליברליזם יתר אם היא מייחסת מצבים נפשיים (או מצבים נפשיים מסוגים מסוימים) ליצורים או למערכות שאין זה סביר לייחס להם מצבים נפשיים (או את המצבים הנפשיים המסוימים המיוחסים להם על ידי התורה). כפי שאמרנו ביחידה 1, אחד ממבחניה של תורה בפילוסופיה של הנפש הוא האם ייחוס המצבים הנפשיים הנובעים ממנה הם סבירים, דהיינו אם היא מייחסת מצבים נפשיים, לפחות בקירוב, לכל מי שסביר לייחס להם מצבים נפשיים ורק למי שסביר לייחס להם מצבים נפשיים. תורה שתנבע ממנה שלילת ייחוסם של מצבים נפשיים לכלבים או לקופים היא בלתי סבירה, כשם שתורה שינבע ממנה ייחוס מצבים נפשיים לווסת-חום היא בלתי סבירה. כפי שאמרנו, אחד ממניעיה של ההשקפה הפונקציונליסטית הוא להתגבר על השוביניזם החומרי של התורה הפיזיקליסטית של זהות-סוג, והאופן שבו היא מתמודדת עמו נחשב ליתרון חשוב שלה. אולם כפי שנראה עתה, הנושא אינו כה פשוט. לעתים נדמה שההשקפה הפונקציונליסטית, שעל פיה החומר אינו מהותי לנפשות, נקלעת לליברליזם יתר, ולעתים, כאשר ההשקפה מאופיינת באופן המתגבר על ליקוי זה, נדמה שהיא נקלעת לשוביניזם יתר. נושא זה בא לידי ביטוי בולט בהקשר של ההבחנה בין גרסאות שונות של פונקציונליזם, ובמיוחד בהקשר של ההבחנה בין פונקציונליזם עממי לבין פונקציונליזם אמפירי.

הרושם שעשויים ליצור מודל המחשב של הנפש ורעיון המופשטות שייחסנו להשקפה הפונקציונליסטית הוא שכל המצבים המרכיבים את רשתות הקשרים הפסיכולוגיות שאליהם מתייחסת השקפה זו מאופיינים באופן מופשט לחלוטין; שהם מאופיינים אך ורק על ידי מיקומם ברשת. המופשטות משמעה שכל תכונה של מצב ממצבי הרשת וכל זיהוי שלו מעבר ליחסים המתקיימים בינו לבין מצבים אחרים הנוטלים חלק ברשת אינם רלוונטיים לאפיון

הפונקציונלי של המצב (וכך לאפיונו הנפשי). וכאמור, גם הקשרים "הסיבתיים", ולא רק המצבים שביניהם חלים הקשרים הללו, מאופיינים באופן מופשט. אחד מביטוייה של מופשטות זו של האפיונים הפונקציונליים הוא שגם הקלט והפלט של הרשתות מאופיינים באופן מופשט. כלומר, הזהות של רשתות הקשרים ושל מצבים שונים בהן איננה תלויה כלל בטיבם של הקלט והפלט; אין זה משנה בכלל מהם הקלט והפלט של המערכת. מערכות קשרים תהיינה זהות מבחינה פונקציונלית אם הן בעלות אותו מבנה, אף אם הקלט והפלט של האחת הוא זרמים חשמליים, הקלט והפלט של השנייה הוא מטבעות המוכנסים אליה ונפלטים ממנה (נניח שמדובר במכונה לפריטת כסף), הקלט של השלישית הוא הקלדת אותיות על מקלדת והפלט שלה הוא הדפסת אותיות על נייר, וכדומה. ההבחנה בין קלטים שונים (שאפשר לציין כ- $I_1, I_2$ , וכדומה) וההבחנה בין פלטים שונים (שאפשר לציין כ- $O_1, O_2$ , וכדומה) הן הבחנות מופשטות לחלוטין. פלטים וקלטים שונים מאופיינים אך ורק באמצעות רשתות הקשרים שבהן הם קשורים והאופנים שבהם הם קשורים בהן. התיאור במונחי תוכנה (כדוגמת מכונת טיורינג) הוא תיאור מופשט, תיאור האדיש הן לזהות הספציפית של המצבים הפנימיים והן לזהות הספציפית של הקלט והפלט.

כזכור, מטרת הפונקציונליזם היא המטרה הרדוקציוניסטית של אפיון המצבים הנפשיים מבלי לציין כנפשיים, כלומר מבלי להשתמש במונחים נפשיים. הדבר מתבטא בכך שבעלילה הפסיכולוגית המקבילה לעלילה הבלשית של לואיס, המצבים הנפשיים מצוינים על ידי משתנים ( $M_1, M_2$ , וכדומה). אולם את העלילה הפסיכולוגית (כמו את העלילה הבלשית) ניתן לספר באופן מופשט יותר או מופשט פחות. היא תסופר באופן המופשט ביותר אם יצונו באמצעות משתנים **כל המצבים** הנזכרים בה, דהיינו לא רק המצבים הנפשיים אלא גם מצבי הקלט החושי והפלט ההתנהגותי (בעלילה כזאת הן המונחים הנפשיים והן מונחי הקלט והפלט הם "מונחים תיאורטיים" המאופיינים באמצעות תפקידם בעלילה). היא תסופר באופן מופשט פחות אם מצבי הקלט והפלט יצונו בה באופן קונקרטי ולא על ידי משתנים, למשל פגיעת קרן אור מסוימת ברשתית (במקרה של קלט) או בעיטה בדלת (במקרה של פלט). חשוב להדגיש שהמטרה הרדוקציוניסטית עשויה להיות מושגת גם במקרה זה, שכן הסייג הרדוקציוניסטי הוא הימנעות משימוש במונחים **נפשיים**, שאותם אנו רוצים לאפיין ולא את מונחי הקלט והפלט (אם כי נראה להלן שאפיון קונקרטי **מסוים** של הפלט עלול לפגוע במטרה הרדוקציוניסטית).

האפשרות של אפיון הקלט והפלט במונחים קונקרטיים מסוימים פתוחה אפוא, באופן עקרוני, בפני הפונקציונליסטים. אולם מדוע עליהם לפנות אליה ולא לבחור באפשרות המופשטת לחלוטין? נחזור לרגע לדוגמה של שלושת האחים מבאר-שבע. דובר בכתבה המתארת את  $X$  כאחיו של  $Y$ , המבוגר ממנו ב-30 שנה, ואת  $Y$  כאחיו של  $Z$ , המבוגר ממנו ב-25 שנה. ל- $X$  דירה ברחוב הרצל בבאר-שבע, ול- $Z$  קיוסק ברחוב מצדה בבאר-שבע. אם יש רק שלישייה אחת של אחים שביניהם מתקיימים פערי גילים כאלו ואשר לאחד דירה ברחוב הרצל בבאר-שבע ולאחר קיוסק ברחוב מצדה בבאר-שבע, הרי מידע זה מספיק כדי לזהות את השלישייה. אולם נניח שלא רק זהותם של האחים הייתה ניתנת בכתבה באמצעות משתנים,

אלא גם הנכסים שיש להם (ל-X יש נכס T ול-Z יש נכס U). במקרה כזה, ייתכן שהמידע לא יספיק כדי לזהות את השלישייה, שכן ייתכן שהוא יאפיין שלישיות אחים נוספות; והסיכוי שהמידע לא יספיק לזיהוי השלישייה יגבר אם הכתבה תאפיין באמצעות משתנים (במקום באמצעות מספרים) גם את פערי הגילים שבין האחים. באופן כללי, ככל שהתיאור מופשט יותר, הוא עשוי לחול על מספר גדול יותר של קבוצות של ישויות (בין אם הישויות הללו הן אנשים ובין אם הן אובייקטים פיזיקליים למיניהם, מצבים נפשיים, וכדומה).

טענה מקובלת היא שזהו המצב באשר לאפיון הפונקציונלי המופשט ביותר של המצבים הנפשיים, אפיון המתאר גם את מצבי הקלט והפלט של חיינו הנפשיים באופן מופשט לחלוטין, באמצעות תיאור יחסיהם למצבים אחרים ברשת ותו לא. אפיון זה, לפי הטענה, איננו אפיון הולם של מצבינו הנפשיים, משום שהוא חל לא רק עליהם, לא רק על המערכת הנפשית שלנו. נניח שניקח את מערכת כללי הפסיכולוגיה החלים על אדם מסוים – את "התכנית הפסיכולוגית" (המופשטת לחלוטין) של אדם זה, ונציב כמצבי הקלט שלה מצבים של קבלת כסף (באופנים שונים ובסכומים שונים), כמצבי הפלט שלה מצבים של נתינת כסף (באופנים שונים ובסכומים שונים), וכמצביה הפנימיים מצבים כלכליים כמו היות שער רבית מסוים, היות שער חליפין מסוים, וכדומה. נקבל מערכת כלכלית לכל דבר ועניין (ניתן לדמיון שאדם עשיר משתלט על כלכלתה של מדינה קטנה ומפעיל אותה בהתאם לתכנית של הפסיכולוגיה של אדם כלשהו).<sup>58</sup> כלומר, ייתכן שאותו אפיון פונקציונלי מופשט החל על הפסיכולוגיה של אדם יחול גם על מערכת כלכלית כלשהי.

## שאלה

איזו מסקנה אפשר להסיק מניסוי חשיבה זה?

מאחר שאיש לא יקבל שמערכת כלכלית כזו היא מערכת נפשית, מערכת בעלת מצבים נפשיים, ומאחר שיש לה אותו אפיון פונקציונלי מופשט שיש למערכת בעלת מצבים נפשיים, ניתן להסיק שאפיון פונקציונלי מופשט איננו מספיק לאפיונה של נפשיות. השקפה על פיה אפיון כזה הוא אפיון מספיק לנפשיות היא השקפה ליברלית מידי, השקפה המייחסת נפשיות למי שאין להם מצבים נפשיים, ויש אפוא לדחותה. שימו לב שאין זה משנה כלל אם אומנם מתקיימת או סביר שתתקיים מערכת כלכלית בעלת אפיון פונקציונלי מופשט הזהה לחלוטין לאפיון הפונקציונלי המופשט של חיי הנפש של אדם כלשהו. כל מה שצריך להניח הוא שמערכת כלכלית כזו היא אפשרית, וקשה לדחות את הטענה שהיא אפשרית: נראה שהנחת קיומה אינה מתנגשת לא בחוקי הלוגיקה ואף לא בחוקים מדעיים כלשהם (ובכללם חוקי הפסיכולוגיה).

<sup>58</sup> ניסוי חשיבה זה לקוח, בשינויים קלים, מהפילוסוף נד בלוק. ראו:

N. Block, "Troubles with Functionalism", in Ned Block (ed.), *Reading in Philosophy of Psychology*: Volume I, Cambridge (Mass.): Harvard University Press, 1980.

בקצרה, הטענה היא, שאפיון פונקציונלי מופשט לחלוטין של חיינו הנפשיים הוא אפיון הלוקה בליברליזם-יתר. כזכור, מניע מרכזי של ההשקפה הפונקציונליסטית היה ניסיונה להתגבר על שוביניזם-היתר החומרי של התורה הפיזיקליסטית של זהות-סוג. הרעיון היה לפנות לאפיון מופשט יותר וליברלי יותר, כזה שיחול גם על חיים נפשיים הממומשים באופנים פיזיקליים שונים מהאופן שבו ממומשים חיינו הנפשיים. אולם עתה מתברר שהמטוטלת נעה רחוק מידי בכיוון הנגדי: האפיון הפונקציונלי המופשט לחלוטין – זה המאפיין באופן מופשט גם את הקלט והפלט של המערכת הפסיכולוגית שלנו – הוא **ליברלי מידי**. אין הוא יכול לשמש כאפיון הולם של היסוד הנפשי, משום שהוא חל גם על מערכות שאינן נפשיות, וכך אינו יכול להבחין בין הנפשי לבין מה שאינו נפשי.

השקפה פונקציונליסטית מופשטת לחלוטין, השקפה המאפיינת באופן מופשט גם את הקלט והפלט של מערכות נפשיות, לוקה בליברליזם-יתר, ולפיכך אינה יכולה להיות ההשקפה הנכונה בדבר מהותם של המצבים הנפשיים.

דומה, אם כן, שאפיון תכניתי מופשט, כדוגמת אפיון של מכונת טיורינג מסוימת, לא יכול להיות אפיון הולם של חיינו הנפשיים. על הפונקציונליסטים הדוגלים ברעיון של הנפש כתוכנת הגוף להתפשר במידה מסוימת, ולהגביל באופן כלשהו את המופשטות המיוחסת ליסוד הנפשי. כאשר מאפיינים מצבים נפשיים באמצעות מיקומם ברשתות הקשרים הפסיכולוגיות, יש לאפיין באופן קונקרטי את מצבי הקלט והפלט הכלולים ברשתות אלו. מרבית הדוגלים בהשקפה הפונקציונליסטית אמנם עושים זאת. אולם הם עושים זאת באופנים שונים. לדוגלים בפונקציונליזם אמפירי דרך אופיינית לעשות זאת, ולדוגלים בפונקציונליזם עממי דרכים אופייניות לעשות זאת.

בהקשר זה עולות שתי שאלות: האחת היא, **מהם** הקלט והפלט שבאמצעותם מאפיינת כל השקפה את המצבים הנפשיים? השנייה היא, **כיצד** מאפיינת כל השקפה את הקלט והפלט? הקלט והפלט באמצעותם מאפיינים חלק מהדוגלים בהשקפת הפונקציונליזם האמפירי את המצבים הנפשיים הם הקלט והפלט של מערכת העצבים המרכזית. אם נתייחס לשרשרת סיבתית של תפיסה חושית, שרשרת המתחילה באירוע שבו מעורב האובייקט הנתפס (האובייקט בעולם החוץ-נפשי) ונגמרת באירוע מסוים במוח, הקלט הרלוונטי הוא החוליה האחרונה בשרשרת זו לפני החוליות המתרחשות במערכת העצבים המרכזית (למשל, האות הנשלח מהעין למוח). חוליה זו היא הקלט של מערכת העצבים המרכזית, הגירוי שהיא מקבלת. ואם נתייחס לשרשרת סיבתית המוליכה לפעולה, שרשרת המתחילה באירוע מסוים במוח ומסתיימת בפעולה מסוימת, הפלט הרלוונטי הוא החוליה הראשונה בשרשרת זו לאחר החוליות המתרחשות במערכת העצבים המרכזית (למשל, האות הנשלח מהמוח לשרירים). חוליה זו היא הפלט של מערכת העצבים המרכזית, התגובה שהיא מפיקה.

אירועים אלו מאופיינים במונחים פיזיקליים – במונחי התכונות הנורו-פיזיולוגיות שלהם (לו היו מאופיינים אפיון פונקציונלי, הגישה המתקבלת הייתה פונקציונליזם מופשט לחלוטין). אנו יכולים לראות שגישה זו מתגברת על בעיית ליברליזם-היתר שבה כרוכה הגישה הפונקציונליסטית המאפיינת גם את הקלט והפלט באופן מופשט לחלוטין. על פיה, בניגוד לפונקציונליזם המופשט, לא יתקבל מצב שבו מערכת כלכלית כלשהי אמורה להיות מערכת נפשית, שהרי הפלט והקלט של מערכות כלכליות אינם אירועים נורו-פיזיולוגיים. אולם בחמקה מליברליזם-יתר, גישה זו צועדת הרחק מידי בכיוון המנוגד: היא כרוכה במה שנראה כשוביניזם-יתר. אמנם גישה זו אינה דורשת ממצבים נפשיים זהים להיות זהים מבחינה פיזיקלית, אולם היא דורשת מהם להיות קשורים ברשתות קשרים שמצבי הקלט והפלט שלהם זהים מבחינה פיזיקלית. היא מחויבת אפוא לשלול מצבים נפשיים מיצורים שהפיזיקה שלהם שונה מהפיזיקה. בכך היא חותרת תחת אחת המוטיבציות העיקריות של הרעיון הפונקציונליסטי. היא תשלול נפשיות גם מיצורים שסביר לייחס להם נפשיות, כמו תמנון (בהתאם לדוגמה שבה השתמש פטנאם בטיעון ריבוי המימושים). המטוטלת של שוביניזם-ליברליזם נעה שוב לכיוון השוביניסטי.

השקפה פונקציונליסטית המאפיינת את הקלט והפלט של מערכות נפשיות במונחים פיזיקליים לוקה בשוביניזם-יתר, ולפיכך אינה יכולה להיות ההשקפה הנכונה בדבר מהותם של המצבים הנפשיים.

הפונקציונליסטים עשויים להרחיב את נקודת-המבט ולהתייחס באפיון הקלט והפלט לגורמים שמחוץ למערכת העצבים המרכזית. אפשר, למשל, להחשיב את האותות המשוגרים ממערכת העצבים ההיקפית לשרירים בתור הפלט, ואת האירועים העצביים הראשונים של השרשרות הסיבתיות הכרוכות בתפיסה חושית בתור הקלט. אולם גם במקרה זה, מאחר שהדרך היחידה הבאה בחשבון לאפיון אירועים אלו היא במונחים פיזיקליים, גם חלופה זו תלקה בשוביניזם-יתר.

חלופה דומה, שאליה נטו חלק מהדוגלים בפונקציונליזם עממי, היא להרחיב את נקודת המבט עוד יותר, ולהחשיב בתור פלט את התנועות גופניות הכרוכות בהתנהגות, ובתור קלט את הגירויים של איברי החושים (למשל, פגיעת קרני אור ברשתית). אולם כל עוד אירועים אלו מאופיינים במונחים פיזיקליים, לא נימלט משוביניזם-יתר. בצד של הקלט, חלופה זו תשלול נפשיות מיצורים שאיברי החושים שלהם שונים משלנו, ושמסוים כך הם רגישים לגירויים פיזיקליים שונים, כמו עטלפים שיש להם חוש תהודה (זכרו את הדוגמה מיחידה 1). שימו לב, שחלופה זו תשלול מהם לא רק התנסויות חושיות הכרוכות באיברי חושים שיש לנו ואין להם – בכך אין בעיה, כי סביר להניח שאומנם אין להם התנסויות כאלו; חלופה זו תשלול מהם נפשיות בכלל, או על-כל פנים מצבים נפשיים אחרים, כמו האמנות שונות, משום שהקלט הוא חלק מהותי של האפיון הפונקציונלי הדרוש לנפשיות. אפשר לחשוב שנוכל לכלול

באפיון הקלט מצבים של איברי חושים שיש לכל היצורים הנפשיים וכך לחמוק מן הקושי. אולם עלינו לזכור שהשקפה הולמת של היסוד הנפשי צריכה לייחס נפשיות לכל יצור אפשרי בעל נפשיות, ולפיכך אפיון הקלט צריך לכלול איברי חושים בעלי מבנים חומריים שונים ומשונים, כדי כך שאין זה ברור מה אפיון כזה לא יכלול. במילים אחרות, אפיון זה יהיה אפיון ריק מתוכן. בצד של הפלט, חלופה זו תשלול נפשיות מיצורים אפשריים שהפיזיקה שלהם וכך גם תנועותיהם הגופניות שונות לחלוטין משלנו – למשל, מיצורים אפשריים העשויים מסיליקון. התוצאה שעל פיה יצורים העשויים מסיליקון שיש להם מצבים נפשיים אינם אפשריים כלל היא תוצאה בעייתית, כפי שכבר ראינו. ייתכן שניתן היה להתגבר על קושי זה על ידי אפיון גאומטרי של התנועות הגופניות (להבדיל מאפיון במונחי חומר), אולם גם אפיון כזה ישלול נפשיות מיצורים מסוימים שסביר לייחס להם נפשיות, למשל יצורים שאף אם הפיזיקה שלהם דומה ביסודה לשלנו, איבריהם החיצוניים שונים משלנו (כמו נחשים, החסרים ידיים ורגליים<sup>59</sup>). הרעיון שנוכל לכלול באפיון הפלט תנועות של כל היצורים הנפשיים הוא בעייתי, בדיוק כמו הרעיון המקביל באשר לאפיון הפלט, כפי שהראינו בפסקה הקודמת.

בשלב זה עשויה להתעורר המחשבה שאולי המהות הספציפית של הקלט והפלט אינה משנה כלל לאפיון היסוד הנפשי, ושמערכת קשרים פנימית מתאימה תהייה מערכת נפשית גם אם יודבקו לה קלטים ופלטטים מסוגים שונים, כל עוד קלטים ופלטטים אלו יקיימו את האפיונים הפונקציונליים הנדרשים.

### שאלה

מה הבעיה הכרוכה ברעיון זה?

רעיון זה אינו אלא הרעיון של פונקציונליזם מופשט לחלוטין. כך, רעיון זה חוזר ומניע את המטוטלת באופן מוגזם לכיוון המנוגד: הוא מוליך אותנו לליברליזם-יתר.

כיצד יוכל הפונקציונליזם להתגבר על בעיה זו? כפי שאומר נד בלוק, השאלה היא:

**האם יש תיאור של קלטים ופלטטים שהינו מספיק ספציפי כך שיימנע ליברליזם, אך מספיק כללי כדי שיימנע שוביניזם?**<sup>60</sup>

נראה שהמטוטלת נעה פעם לכיוון הליברלי מידי – לייחוס מצבים נפשיים למי שאין להם מצבים נפשיים, ופעם לכיוון השוביניסטי מידי – לשלילת מצבים נפשיים ממי שיש להם מצבים נפשיים. זוהי דילמת הקלט-פלט של הפונקציונליזם.

<sup>59</sup> זו דוגמה של נד בלוק ממאמרו הנזכר בהערה הקודמת.

<sup>60</sup> N. Block, "Troubles with Functionalism", in N. Block (ed.), *Reading in Philosophy of Psychology: Volume I*, Cambridge (Mass.): Harvard University Press, 1980.

דרך המקובלת על הפונקציונליסטים העממיים לאפיון הקלט והפלט נראית ממבט ראשון ככזו המתגברת על הקושי. הרעיון הוא לאפיין את הקלט והפלט של חיינו הנפשיים במונחים יומיומיים, כלומר במונחים שבאמצעותם אנו מאפיינים את הקלט והפלט הללו במהלך החיים הרגיל, כחלק מהפרקטיקה של הפסיכולוגיה העממית. כך, אנו אומרים שאנו רואים שולחנות, שומעים חריקת בלמים של מכונית, מסמנים למונית לעצור, ועוד. חשוב לשים לב, כי כשאנו מתארים פעולה כסימון למונית לעצור, תיאור זה איננו תיאור פיזיקלי. למעשה, אין זה תיאור של תנועה גופנית מסוימת (או תיאור הכרוך בתנועה גופנית מסוימת), שהרי קיימות דרכים פיזיקליות רבות לסמן למונית לעצור. שום אפיון פיזיקלי ספציפי איננו מהותי לתיאור כזה. אולם בהקשר זה מתעוררת שאלה: מה מהותי לתיאור התנהגות זה? מה עומד ביסוד התנועות הפיזיקליות השונות המהוות כולן סימון למונית לעצור? בעטיו של מה הן מהוות סימון למונית? ניתן לחשוב שמה שעומד ביסודן הוא (בין היתר) שהן נגרמות על ידי הרצון לגרום לנהג המונית שיעצור. אם כך, האפיון הנוכחי של הפלט נתקל בבעיה הנעוצה בעובדה שרצון הוא מצב נפשי. אם אפיון הפלט מניח התייחסות למצב נפשי, אין הוא יכול להוות חלק מתיאור פונקציונלי של מצבים נפשיים שאמור להינתן כולו במונחים שאינם נפשיים (זוהי, כזכור, היומרה הרדוקציוניסטית של הפונקציונליזם).

ברדון-מיצ'ל וג'קסון הצביעו על דרך להיחלצות מקושי זה.<sup>61</sup> הם טוענים שניתן לאפיין פעולות, כדוגמת הסימון למונית, מבלי להניח כל אפיון נפשי. הרעיון הוא שניתן לאפיין פעולות במונחי ההשפעה שיש להן באופן טיפוסי (או בהתקיים תנאים מסוימים) על מצבו של הסובייקט. כך, את כל התנועות המהוות סימון למונית לעצור ניתן לאפיין כפעולות הנוטות להביא למצב שבו הסובייקט ימצא בתוך מונית (או כפעולות שיביאו למצב כזה אם יתקיימו תנאים נוספים מסוימים). מאחר שלחלק מפעולותינו אין השפעה מעין זו על מצבינו אנו, הרי שיש להרחיב אפיון זה: ניתן לאפיין פעולות על פי ההשפעה שיש להן על גורמים כלשהם, בין אם על הסובייקט עצמו, על סובייקטים אחרים, או על חפצים בסביבה.

דרך אפיון זו אינה מניחה התייחסות למצבים נפשיים. נראה שהיא אינה נגועה בשוביניזם-יתר (היא אינה מתייחסת לפיזיקה של גופו של הסובייקט), וגם לא בליברליזם-יתר (מערכות כלכליות, למשל, אינן מוצאות את עצמן בתוך מונית). דרך אפיון זו נראית קרובה ברוחה לפונקציונליזם העממי, משום שתיאורי הפעולות הרלוונטיים הם תיאורים יומיומיים של פעולותינו (דרך זו אמנם הוצעה על ידי מי שדוגלים בהשקפה זו). לעתים מייחסים יתרון לפונקציונליזם העממי על-פני הפונקציונליזם האמפירי בשל אופן טיפולו בסוגיית הקלט והפלט. אולם אין מניעה לכך שדרך אפיון זו של הקלט והפלט תשתלב גם בפונקציונליזם האמפירי. אמנם הסברי הפסיכולוגיה המדעית שונים מהסברי הפסיכולוגיה העממית בכך שהם מתייחסים גם לתהליכים ומצבים פנימיים אחרים, אולם עתה מדובר לא באפיון של מצבים ותהליכים נפשיים אלא באפיון של פעולות, והפונקציונליזם האמפירי בהחלט יכול

D. Braddon-Mitchell and F. Jackson, *Philosophy of Mind and Cognition*, Oxford: Blackwell, 1996.

הדוגמה של סימון למונית לקוחה גם היא מספר זה.

לאפיין פעולות כפי שהן מאופיינות על ידינו באופן יומיומי: מטרה מרכזית של הפסיכולוגיה המדעית היא להסביר את פעולותינו כפי שהן מאופיינות על ידינו באופן יומיומי, ואין כל מניעה לכך שהסברים של פעולותינו, כפי שהן מאופיינות על ידינו באופן יומיומי, יתייחסו למצבים ולתהליכים פנימיים שאינם מוזכרים בפרקטיקת הפסיכולוגיה העממית. אם כך, אין לפונקציונליזם העממי עדיפות בעניין הנדון על-פני הפונקציונליזם האמפירי. ההצעה הנדונה לאפיון הפלט של רשתות הקשרים הנפשיות פתוחה גם בפני הפונקציונליזם האמפירי.



## 6. בעיית זהות התוכנה, ופונקציונליזם הומונקולארי

### 6.1. בעיית זהות התוכנה

דיון בקושי נוסף אשר מועלה כנגד ההשקפה הפונקציונליסטית יוביל אותנו להצגתה של גרסה נוספת של השקפה זו. קושי זה עניינו בקריטריון הזהות של מצב נפשי. על פי ההשקפה הפונקציונליסטית, שני מצבים נפשיים של שני יצורים נפשיים שונים שייכים לאותו סוג נפשי אם הם מממשים אותה פונקציה בחיים הפסיכולוגיים של היצורים הללו. מודל המחשב של הנפש נתן פשר לרעיון זה. על פיו, מימוש אותה פונקציה משמע מימוש אותו מצב חישובי, וראינו את פשרו המדויק של הרעיון בדבר מימוש אותו מצב חישובי. כאמור,

**ב<sub>2</sub>\*. שני מצבים מממשים את אותו מצב חישובי, אם, ורק אם, (1) רשתות הקשרים הסיבתיים של המערכות שלהן שייך כל אחד מהם זהות במבניהן; ו-(2) הם תופסים מקומות זהים ברשתות הללו.**

ולפי הפונקציונליזם,

**ב<sub>2</sub>\*. שני מצבים מממשים את אותו מצב נפשי (הם שייכים לאותו סוג נפשי), אם, ורק אם, (1) רשתות הקשרים הסיבתיים של המערכות שלהן שייך כל אחד מהם זהות במבניהן; ו-(2) הם תופסים מקומות זהים ברשתות הללו.**

אם כן, בהתאם לתנאי (1), כדי ששני מצבים נפשיים של שני יצורים נפשיים ישתייכו לאותו סוג נפשי (למשל, ההאמנה שיורד גשם), רשתות הקשרים של חיהם הנפשיים של יצורים אלו חייבות להיות זהות. באופן אחר, יצורים אלו חייבים לממש את אותה תוכנה – מצבים זהים מבחינה פונקציונלית הם מצבים התופסים אותו מקום באותה תוכנה. ובאופן אחר נוסף, על אותה פסיכולוגיה לחול עליהם. הפסיכולוגיה היא שמתארת את רשתות הקשרים שבהן כרוכים מצבינו הנפשיים – את כללי המעבר ממצב אחד לאחר, את התוכנה הנפשית שלנו. אנו רואים אפוא שתנאי הכרחי לכך שלשני יצורים נפשיים יהיו מצבים נפשיים מאותו סוג הוא שתחול עליהם אותה פסיכולוגיה (זהו תנאי הכרחי, אך לא תנאי מספיק, שהרי נדרש גם תנאי (2)).

מההשקפה הפונקציונליסטית נובע שתנאי הכרחי לכך שלשני יצורים נפשיים יהיו מצבים נפשיים מאותו סוג הוא שתחול עליהם אותה פסיכולוגיה.

טענתם של מבקרים אחדים של הפונקציונליזם היא שהשלכה זו של הפונקציונליזם היא בלתי סבירה. ברור, לטענתם, שיצורים שונים עשויים להימצא במצבים נפשיים מאותו סוג גם אם הם נבדלים זה מזה במידה משמעותית בפסיכולוגיה שלהם. בני אדם שמתקיימים ביניהם הבדלי אישיות ניכרים – אולי אף כל בני האדם – חולקים כאבים, פחדים, רצונות למיניהם, האמנות שונות, או לפחות חלק מאלו. האם ניתן לומר שעל כל בני האדם, או מרביתם, חלה אותה חוקיות פסיכולוגית, חרף הבדלי האישיות הניכרים שביניהם? אם נטייתכם היא להשיב על שאלה זו בחיוב, חשבו על מבוגרים לעומת ילדים, על גאונים לעומת בני אדם מוגבלים בשכלם, על בני אדם שפויים לעומת מטורפים למיניהם. האומנם סביר לומר באשר לכל אחד מצמדי הקבוצות הללו שאותה חוקיות פסיכולוגית חלה על חבריהם? זאת ועוד, האומנם אותה חוקיות פסיכולוגית מאפיינת את בני האדם ואת אותן חיות שלהן אנו מייחסים מצבים נפשיים כמו כאבים, פחדים ושמחות? לפחות על-פני הדברים, נדמה שהקריטריון הפונקציונליסטי לזהות נפשית נתקל בקושי רציני. נראה שקריטריון זה מחויב לכך שכל היצורים החשים כאב, למשל, יהיו כפופים לאותה חוקיות פסיכולוגית, ושתוצאה זו היא בלתי סבירה. תוצאה זו נראית בעייתית במיוחד כשמדובר בגרסת הפונקציונליזם האמפירי, משום שגרסה זו מחויבת לזהות פונקציונלית ברמה פרטנית מאוד.

ייתכן שקושי זה ייראה פחות חמור אם נזכור שמצבים נפשיים רבים (אם לא כולם) מאופיינים על פי הפונקציונליזם באמצעות קשרים **מותניים**, ולא קשרים בפועל – בינם לבין מצבים אחרים. בדוגמה שהבאנו, האמנתו של אדם שיורד גשם, מדובר, לפי הפונקציונליזם, במצב אשר יובילו על פי רוב לשימוש במטרייה **אם**, למשל, אדם זה מאמין גם שהציאה לרחוב כשיורד גשם תיכרך בהירטבות אלא אם כן יינקט אמצעי כלשהו, הוא רוצה שלא להירטב, ומאמין ששימוש במטרייה הוא אמצעי יעיל למניעת הירטבות. את החוקיות הפסיכולוגית שלפי הפונקציונליזם על שני יצורים נפשיים לחלוק כדי שיוכלו להימצא באותו מצב נפשי יש להבין אפוא כ**חוקיות מותנית**: אין זו חוקיות הקובעת שכל היצורים הנפשיים הנמצאים במצב נפשי X יעברו ממנו לאותו מצב נפשי אחר. זוהי חוקיות הקובעת שהם יעברו ממצב נפשי X לאותו מצב נפשי אחר רק אם גם מצבים נפשיים נוספים שלהם זהים. כך, הבדלים במה שיצורים נפשיים שונים מסיקים מאותה האמנה, הבדלים בפעולות שאליהן מוליכה אותם אותה שאיפה, וכדומה, אין משמעותם היעדר חוקיות פסיכולוגית משותפת ביניהם. הבדלים כאלו מתיישבים עם חוקיות (מותנית) משותפת. הפונקציונליסטים יכולים לקבל את קיומם של הבדלים כאלו במעברים הפסיכולוגיים ביו יצורים נפשיים שונים. יצורים נפשיים שונים ואף בני אדם שונים נבדלים זה מזה בזיכרונותיהם, בהאמנותיהם, בהעדפותיהם, ועוד. בהבדלים כאלו, יכולים הפונקציונליסטים לטעון, מתבטאים הבדלי האישיות בין היצורים הנפשיים השונים, ובהבדלים אלו נעוצים ההבדלים במעברים הפסיכולוגיים שלהם. מיקי לא תעבור מההאמנה שיורד גשם לרצון להצטייד במטרייה בצאתה מביתה ויובל כן יעבור מההאמנה זו לרצון זה, משום שיוכל אינו רוצה להירטב ואילו למיקי לא אכפת. לו העדפתה של מיקי הייתה זהה להעדפתו של יובל, גם היא הייתה רוצה להצטייד במטרייה. כך, הם כפופים לאותה חוקיות פסיכולוגית.

אולם, האם סביר שכל הבדל במעברים הפסיכולוגיים של יצורים נפשיים שונים מקורו בהבדל במצבים נפשיים נוספים שלהם? סביר שגאון ואדם מוגבל בשכלו יסיקו מסקנות שונות לחלוטין מאותן הנחות בנושאים רבים מאוד, ובכללם גם בנושאים שבהם הם נחשפו למידע זהה. קשה לייחס הבדל זה בהיסקיהם להאמנות שונות. האין זה סביר שגם אם יחלקו את כל המצבים הנפשיים בזמן נתון, מסקנותיהם תהיינה שונות, ולפיכך שחוקיות פסיכולוגית שונה (גם חוקיות פסיכולוגית מותנית שונה) מאפיינת אותם. במילים אחרות, האין זה סביר שהם כפופים לכללי מעבר שונים בין מצבים? ייתכן שיהיו פונקציונליסטים שיטענו שניתן לייחס הבדלים כאלו בהיסקים להבדלים במצבים ברמת פסיכולוגיית העומק – הבדלים במצבי מוח הנושאים מידע, שאינם המצבים המוכרים מן הפרקטיקה של הפסיכולוגיה העממית (ובאופן ספציפי יותר, אף על פי שמצבים אלו, בהיותם מצבים הנושאים מידע, דומים להאמנות, הם אינם האמנות; בין היתר, הם נבדלים מהאמנות בכך שאינם מודעים לבעליהם – בעליהם אינם יודעים עליהם כשם שהם יודעים על האמנותיהם). האם יכולים הבדלי המידע שבין גאון לבין אדם מוגבל בשכלו, הנעוצים במצבי מוח כאלו, להסביר את ההבדלים בהסקת מסקנות בין שני בני אדם כאלו? ספק. ההבדלים בהסקת מסקנות בין בני אדם כאלו הם כלליים ומשתרעים כמעט על כל תחום חשיבה, והבדלי מידע ספציפיים אינם יכולים להסביר הבדל כללי בכשרים ההיסקיים (הבדל המתבטא בנושאים רבים). לפיכך, מצבי המוח הללו אמורים לשאת מידע כללי ומופשט, כמו חוקים לוגיים מורכבים. אולם ההשערה שקיימים מבנים מוחיים הנושאים מידע לוגי כזה ואשר מתפקדים כהאמנות היא השערה מרחיקת לכת למדי, והדרך לבססה אינה נראית באופק.

פונקציונליסטים שונים מקבלים את הטענה שאין פסיכולוגיה המשותפת לכל היצורים הנפשיים המסוגלים להימצא באותם מצבים נפשיים, ומציגים את השקפתם באופן שאינו מתנגש בטענה זו. אחד הרעיונות, בקיצור נמרץ, הוא כדלקמן: אם אנו מחלקים את מערכות הקשרים של חיינו הנפשיים לתת-מערכות – תוחמים חלקים מסוימים של רשתות הקשרים – ניתן לאפיין מצבים נפשיים במסגרת תפקידם או מיקומם בתת-מערכות אלו. סביר שגם אצל יצורים שונים מבחינה פסיכולוגית ניתן לבדוד תת-מערכות פסיכולוגיות דומות. יצורים שונים מבחינה פסיכולוגית ייבדלו זה מזה בחלק מתת-מערכות הפסיכולוגיות שלהם, ובקשרי הגומלין, באינטראקציות, שבין תת-מערכות פסיכולוגיות שונות (למשל, הפלט של אותה תת-מערכת עשוי להיות מוזן לתת-מערכת אחת אצל יצור אחד ולתת-מערכת אחרת אצל יצור אחר). אולם יש יסוד להניח שיצורים כאלו גם יחלקו תת-מערכות פסיכולוגיות (אלו וכמה? הדבר תלוי במידת הדומות שבין הפסיכולוגיות שלהם ובמידת הדומות של מורכבותן). אפשר אפוא לאפיין מצבים נפשיים באמצעות תפקידיהם בתת-מערכות המשותפות. בסעיף הבא נדגים תת-מערכות כאלו, ואז נוכל להבין טוב יותר את הרעיון של אפיון מצב נפשי במסגרת תפקידו בתת-מערכת. רעיון זה כרוך בגרסה ספציפית של ההשקפה הפונקציונליסטית שעתה נעבור להציגה, הגרסה של פונקציונליזם הומונקולארי.

## 6.2 פונקציונליזם הומונקולארי

מערכות רבות בטבע בנויות כמו בובה רוסית ("בבושקה"). הבובה הרוסית היא מערכת של בובות – בובה גדולה, ובתוכה בובה קטנה יותר, ובתוכה של זו בובה קטנה עוד יותר. כך, לדוגמה, איברי הגוף של אורגניזמים מורכבים מרקמות המאורגנות באופן מסוים, ואלו מורכבות מתאים המאורגנים באופן מסוים, ואלו מורכבים מגרעינים ומממברנות... וכך הלאה, עד שמגיעים לחלקיקים התת-אטומיים היסודיים ביותר. העובדה שהתחלנו באזכור איברי גוף היא שרירותית: הם אינם המערכות המורכבות ביותר. איברי הגוף של אורגניזמים מרכיבים את הגופים של האורגניזמים עצמם, והאורגניזמים, בתורם, מרכיבים קבוצות שונות של אורגניזמים (עמים, למשל). התמונה היא תמונה של מערכות ותת-מערכות המרכיבות אותן. המושג של **תת-מערכת** (subsystem) הוא מושג יחסותי: התאים הם תת-מערכות של הרקמות שאותן הם מרכיבים, אולם הרקמות עצמן הן תת-מערכות ביחס לאיברים שאותם הן מרכיבות. פונקציונליזם הומונקולארי (Homuncular Functionalism) הוא החלה של תמונה זו על הנפש.

נחשוב, כדוגמה, על המשימה הפסיכולוגית של גיבוש עמדות על סמך תפיסה חושית. זוהי משימה שכולנו עוסקים בה חלק ניכר מן הזמן. איך מסבירה הפסיכולוגיה את ביצועה של משימה זאת? הפסיכולוגיה מסבירה שזוהי משימה מורכבת למדי. ביצועה כרוך, בין היתר, בביצוע משימה של גילוי אובייקטים בסביבה. ביצועה של המשימה של גילוי אובייקטים בסביבה כרוך, בין היתר, בביצוע של משימה של גילוי צורות. ביצועה של המשימה של גילוי צורות כרוך בביצוע של משימה של זיהוי קווים וגבולות של משטחים, וכך הלאה. כל משימה מבוצעת על ידי תת-מערכת המתמחה במשימה זו: היא מתמחה בעיבוד מידע מסוג מסוים – מקבלת קלט מסוים ומפיקה פלט מסוים בהתאם; והפלט שלה מוזן לתת-המערכת שמעליה בהיררכיה ומהווה את הקלט של תת-מערכת זו.<sup>62</sup> לכל תת-מערכת, ניתן לומר, פונקציה משלה.

כיוונו של ההסבר הפסיכולוגי הוא מלמעלה למטה – מתת-המערכות המורכבות יותר לכיוון תת-המערכות הפשוטות יותר. ההסבר הפסיכולוגי נמשך עד אשר מגיעים למשימה שביצועה ניתן להסבר פיזיקלי פשוט. כך, למעשה, מוסברת תבונה ומאפיינים פסיכולוגיים אחרים. פעילות תבונית, פעילות אינטליגנטית, מוסברת על ידי פירוקה לגורמים, כפי שהדגמנו לעיל לגבי הפעילות של גיבוש עמדות על סמך תפיסה חושית. הפונקציות של מערכת מוסברות על ידי הפונקציות של תת-המערכות שלה והאינטראקציות שביניהן, וגם הפונקציות של תת-המערכות מוסברות באותו אופן, וכך הלאה, עד שמגיעים למערכות אותן רואים כמבנים פיזיקליים פשוטים. המשימה שמבצעת כל תת-מערכת היא פשוטה יותר מן המשימה שמבצעת המערכת (או תת-המערכת) שמעליה בהיררכיה. הפעילות המורכבת מוסברת על ידי פעילות פשוטה יותר, עד אשר מגיעים לפעילות פיזיקלית פשוטה. הפעילות האינטליגנטית מוסברת על ידי פעילות פחות ופחות אינטליגנטית. ניתן לומר, שכשאדם מבצע פעילות

<sup>62</sup> תת-מערכת עשויה לקבל קלט מתת-מערכות אחדות (ולאו דווקא מאחת), ולהעביר את הפלט שלה לתת-מערכות אחדות (ולאו דווקא אחת).

אינטליגנטית, היא מורכבת מפעילויות שונות שכל-אחת מהן מבוצעת על ידי "אדם קטן" או "הומונקולוס" (homunculus) – תת מערכת של האדם עצמו – המתמחה במשימה ספציפית מאוד. "בני-האדם הקטנים" או ה"הומונקולי" (homunculi – צורת הריבוי של homunculus) הסמוכים בהיררכיה לאדם עצמו הם פחות אינטליגנטיים מן האדם עצמו, שהרי הפעילות שהם מסוגלים לבצע היא מוגבלת ופשוטה יותר. אולם ההומונקולי שתחתיהם בהיררכיה, שפעילותם מוגבלת ופשוטה עוד יותר, פחות אינטליגנטיים מהם. התמונה המצטיירת היא תמונה מדורגת של הומונקולי שהם פחות ופחות אינטליגנטיים – יותר ויותר "טיפשים" – עד אשר מגיעים לתת-מערכת המבצעת פעילות פשוטה ביותר, פעילות שאינה אינטליגנטית כלל. כך מוסברת פעילות אינטליגנטית על ידי פעילות פיזיקלית פשוטה שאינה אינטליגנטית כלל. כך, למעשה, נעשית רדוקציה לאינטליגנציה, ובאופן כללי, ליסוד הנפשי, המוסבר על ידי יסודות שאינם נפשיים.

ההשקפה של פונקציונליזם הומונקולארי רואה את הנפש כמערכת היררכית של יחידות כאלו.<sup>63</sup> הנפש אינה כמחשב אחד, אלא כמערכת היררכית של מחשבים או מעבדי מידע הקשורים ביניהם. מצב נפשי מסוג מסוים, לפי השקפה זו, הוא **מצב בעל פונקציה מסוימת בתת-מערכת מסוימת**. אפיון זה של סוגים של מצבים נפשיים הוא למעשה אפיון פונקציונליסטי כפול: לא רק המצבים מאופיינים במונחי הפונקציות שלהם במסגרת תת-מערכות, אלא גם תת-המערכות מאופיינות במונחי הפונקציה שלהן, המתבטאת בקשריהן עם תת-מערכות אחרות, תת-מערכות הסמוכות אליהן בהיררכיה.<sup>64</sup> האם מצבים נפשיים מזוהים עם מצבים של תת-מערכות פשוטות ("טיפשות") או עם מצבים של תת-מערכות מורכבות ו"אינטליגנטיות" יותר (תת-מערכות אשר להן עצמן יש תת-מערכות, ושגם לאלו יש תת-מערכות...)? ובכן, ראשית, התשובה לשאלה בדבר האפיון ההומונקולרי של מצב נפשי מסוג מסוים היא אמפירית. אין לשאלה כזו תשובה אפריורי. החקירה הפסיכולוגית האמפירית היא שאמורה לגלות את האפיונים הפונקציונליים ההומונקולריים של סוגיהם של המצבים הנפשיים, ובכלל זה היא אמורה לגלות לאיזו תת-מערכת שייכים מצבים נפשיים מסוגים שונים. במובן זה, ניתן לראות את הפונקציונליזם ההומונקולרי כגרסה של הפונקציונליזם האמפירי. שנית, סביר להניח שסוגים שונים של מצבים נפשיים ייבדלו זה מזה בדרגת המורכבות של תת-המערכות שהם מצביהן. סביר להניח שהאמנה מתוחכמת במיוחד תהיה מצב של תת-מערכת מורכבת, ורציה לאכול – מצב נפשי המאפיין גם יצורים שהפסיכולוגיה שלהם אינה מורכבת במיוחד – תהיה מצב של תת-מערכת פשוטה יותר.

<sup>63</sup> מפתחיה הבולטים של השקפה זו הם הפילוסופים דניאל דנט וויליאם לייקאן. ראו:

D.C. Dennett, *Brainstorms*, Cambridge (Mass.): MIT Press, 1978.

W.G. Lycan, "Toward a Homuncular Theory of Believing", *Cognition and Brain Theory*, 1981, pp. 139-159.

W.G. Lycan, *Consciousness*, Cambridge (Mass.): MIT Press, 1987.

<sup>64</sup> הפילוסוף ויליאם לייקאן מזהה את הפונקציות המאפיינות מצבים נפשיים ומערכות נפשיות על פי הפונקציונליזם ההומונקולרי **כפונקציות ביולוגיות-אבולוציוניות** (או, כפי שהן מכונות לעתים, **פונקציות טלאולוגיות**). הכוונה היא שהתפקיד הסיבתי הרלוונטי לאפיונם של מצבים נפשיים ומערכות נפשיות הוא התפקיד שהם "נבחרו" למלא בתהליך האבולוציוני. גם פונקציונליסטים אחדים שאינם דוגלים בפונקציונליזם הומונקולארי מזהים את הפונקציות הנפשיות כפונקציות טלאולוגיות. גרסה כזו של "פונקציונליזם טלאולוגי" ("Teleological Functionalism") ראו במאמר:

E. Sober, "Panglossian Functionalism and the Philosophy of Mind", *Synthese* 64, 1985, pp. 165-193.

## 7. מבחן טיורינג

ההשקפה הפונקציונליסטית, ובמיוחד כאשר היא פונה אל מודל המחשב של הנפש, מעלה אל פני השטח, באופן בולט ביותר, את השאלה האם מחשבים מסוגלים לחשוב? השאלה איננה האם המחשבים של היום חושבים, אלא האם, באופן עקרוני, מחשבים מסוגלים לחשוב. יהיו שיאמרו כי המוח הוא מחשב, ואם כך, כמעט כל מי שמקבלים את האונטולוגיה המטריאליסטית יסכימו שמחשבים מסוגלים לחשוב. אולם כאשר נשאלת השאלה בדבר מסוגלותם של מחשבים לחשוב, על פי רוב הכוונה ספציפית יותר: האם מחשבים מסוגלים לחשוב בשל-כך שהם מממשים תוכנה מסוימת? האם, במילים אחרות, התוכנה היא זו שרלוונטית לחשיבה? לפחות על-פני הדברים, מן ההשקפה הפונקציונליסטית מתחייבת תשובה חיובית לשאלה זו. לפי השקפה זו, מחשב המסוגל להריץ את התוכנה הנכונה הוא מחשב המסוגל לחשוב. גם אם המחשבים של היום אינם מורכבים מספיק, אין כל מניעה עקרונית שיתקיימו מחשבים בעלי רמת מורכבות מספקת – מחשבים שיהיו מסוגלים להריץ את התוכנה המתאימה, וכך לחשוב.

בהקשר זה כדאי להזכיר את תחום המחקר המכונה "בינה מלאכותית" (artificial intelligence, ובקיצור "AI"). עיסוקו של תחום מחקר זה הוא בפיתוחן של תוכנות מחשב המחקות חשיבה אנושית – ביצועים אינטלקטואליים של בני אדם – בתחומים שונים. השימוש במונח "בינה מלאכותית" אינו אחיד. לעתים משתמשים בו כדי לציין את פיתוחן של תוכנות המיועדות להתמודד עם משימות שבאופן רגיל מחייבות תבונה (למשל, קבלת החלטות בתנאי אי ודאות, אבחון רפואי, ניתוחים כלכליים) אף אם תוכנות אלו מתמודדות עם המשימות הללו בדרכים שונות מן הדרכים האנושיות (במילים אחרות, תוכנות של בינה מלאכותית במובן זה מחקות את יחסי הקלט-פלט האופייניים לפעולה תבונית); ולעתים משתמשים במונח זה כדי לציין תוכנות המתמודדות עם משימות כאלו באופן שבו בני אדם היו מתמודדים אתן. מטרתו של חלק חשוב מתחום מחקר זה אינה לקבל כנתון את אופן פעולת התבונה האנושית ואז לנסות לחקותה באמצעות תוכנה מתאימה, אלא לבחון, באמצעות פיתוח והרצת תוכנות כאלו, איך עשויה התבונה האנושית לפעול. למשל, לשאול אילו תוכנות עשויות להתגבר על הבעיות שעמן מתמודדת התבונה האנושית, כדי להעלות ולבחון השערות בדבר אופן פעולת התבונה האנושית. ענף זה של הבינה המלאכותית, המנסה להבין באמצעות יישומים של בינה מלאכותית את התבונה הטבעית, מהווה חלק חשוב של המדעים הקוגניטיביים. חוקרים בענף זה אינם מעלים ובוחנים השערות בדבר התוכנה המורצת במוחנו רק על סמך המשימות הקוגניטיביות שעמן מתמודדת תבונתנו (סוגי החישובים שאנו עורכים, תהליכי קבלת ההחלטות שעמם עלינו להתמודד בתנאים שונים, וכדומה); אלא הם מתחשבים בנתונים שונים כמו מהירויותיהם של חישובים שונים שלנו, אילוצי החומרה שלנו (כלומר, מגבלותיו הפיזיקליות של המוח), ועוד. המאפיין המהותי

לתחום של בינה מלאכותית, המשותף לענפיו השונים, הוא, כאמור, פיתוח תוכנות מחשב המחקות את החשיבה האנושית.

חיקויים או הדמיות (simulations) של ביצועים והתרחשויות שונים רווחים בתחומי מחקר רבים (בפיזיקה, בהנדסה, ועוד). בתחומים אלו ברורה לכל ההבחנה בין ההדמיה לבין הדבר האמתי, הדבר שאותו מחקים. איש אינו חושב שהדמיית מחשב של מכונית הנוסעת בתנאי שטח קשים, למשל, היא מכונית הנוסעת בתנאי שטח קשים. אולם פילוסופים וחוקרים שונים חושבים שפעילותו של מחשב המריץ תוכנה הולמת של בינה מלאכותית, **אינה הדמיה בלבד** של חשיבה אלא היא בעצמה חשיבה, היא הדבר האמתי. לפי גישה זו – שכאמור נראה כי היא מתחייבת מן ההשקפה הפונקציונליסטית – ההבחנה בין הדמיה לבין הדבר האמתי אינה קיימת כאשר מדובר בחשיבה. יש להבחין את המקרה של הבינה המלאכותית מן המקרה הרגיל של הדמיה – יש טעם טוב בזכות ההתייחסות אל התוכנה ההולמת של הבינה המלאכותית כאל חשיבה ממש ולא רק כאל הדמיה של חשיבה. הטעם נעוץ בכך שחשיבה, לפי גישתם של הפונקציונליסטים, כמו פעילויות נפשיות אחרות, היא מאפיין מופשט. אמנם כל תהליך חשיבה ממומש באופן חומרי כלשהו, אולם אין לזהות את תהליך החשיבה עצמו עם מימוש חומרי זה או אחר. כל מימוש שלו הוא חשיבה, ולכן גם הדמיית מחשב של תהליך היא חשיבה. כבר דנו רבות ברעיון הזה בדבר מופשטות היסוד הנפשי.

יש הסוברים כי אין פשר לייחוס חשיבה למחשבים, משום שהחשיבה המעורבת בפעולתם אינה חשיבתם שלהם אלא של **מתכנתיהם**: המתכנתים הם שחושבים, ולא המחשב. אולם אמירה זו אינה מדויקת. המתכנתים אינם חייבים לבצע בפועל בראשם את כל החישובים שהתכנית מאפשרת למחשב לבצע (ברור שמי שמתכנת מכונת טיורינג לבצע פעולות חיבור מספרים לא ביצע בראשו את כל פעולות החיבור שמכונה זו עשויה לבצע). אם אמנם מעורבת חשיבה בפעולתם של מחשבים, אי-אפשר אפוא להתייחס אליה כאל חשיבתם של המתכנתים. אפשר לטעון, בתגובה, שעדיין אין לייחס את החשיבה הכרוכה בפעולתם של מחשבים למחשבים עצמם, משום שפעולת המחשבים נעשית על פי **כללים** שקובעים עבורם המתכנתים. אולם מדוע מתנגשת עובדה זו בהיותם של מחשבים חושבים? גם הכללים שעל פיהם בני אדם חושבים אינם פרי חשיבתם שלהם: ייתכן שחלקם טבועים מלידה (הם נקבעים על ידי הגנים שלנו) ורבים מהם בוודאי נלמדו מאחרים. טענה אחרת היא שמחשבים אינם חושבים משום שהם יצירים מלאכותיים. אולם מדוע מתנגשת עובדת המלאכותיות של מחשבים באפשרות היותם של מחשבים חושבים? לפחות על פני

הדברים, סביר להניח שתכונותיהם העכשוויות של ישויות הן הרלוונטיות לשאלה האם הן חושבות, ולא עובדות היסטוריות באשר לאופן היווצרותן. כדי לטעון שעובדות באשר להיווצרותן של ישויות רלוונטיות למאפיינים נפשיים כלשהם שלהן, נדרש נימוק מורכב.<sup>65</sup>

למעשה, כפי שאמרנו, על פי הגרסאות הרווחות של ההשקפה הפונקציונליסטית, חשיבה, ונפשיות בכלל, דורשות יותר מאשר הרצת התוכנה המתאימה. הן דורשות גם קלט ופלט מסוגים מסוימים. אולם עתה נתרכז בשאלת התוכנה. לשם דיון בשאלה שאותה אנו רוצים לבחון עכשיו, ניתן להתעלם משאלת הקלט והפלט. אנו נניח כעת שאכן פעולה על פי תכנית בעלת רמת מורכבות הולמת – תכנית בעלת ארגון פונקציונלי הולם – מכוננת חשיבה. האם אנו יכולים ללמוד מעמדה זו **מהו המבחן הנכון לחשיבה**, המבחן לקביעה האם מערכת מסוימת חושבת או לא? ובכן, הטענה שחשיבה (כמו נפשיות בכלל) מכוננת על ידי ארגון פונקציונלי ברמת מורכבות מסוימת עדיין אינה מספקת כדי ללמדנו מהי רמת המורכבות הפונקציונלית הדרושה לקיומה של חשיבה. נניח שחשיבתו של האדם מאופיינת על ידי רמת ארגון פונקציונלי גבוהה במיוחד. אין זה סביר לומר שחשיבה בכלל מכוננת על ידי ארגון פונקציונלי ברמת מורכבות זו לפחות, שהרי יש יצורים חושבים שחשיבתם מורכבת במידה פחותה מחשיבתו של האדם. עד כמה צריך מחשב, למשל, להיות "מתוחכם" – עד כמה מורכבת צריכה תכניתו להיות, כדי שנוכל לומר שהוא חושב? ומהו סוג המורכבות הנדרש? מהו קריטריון המורכבות ההולמת?

תשובה מעניינת (אם כי חלקית, כפי שנראה) לשאלה זו, הציע אלן טיורינג.<sup>66</sup> הרעיון הגלום בתשובתו של טיורינג הוא להשיב לשאלה זו בדרך עקיפין. טיורינג הציע **מבחן מעשי** או "אופרציונלי" (operational) שיספק את התשובה לשאלה האם מחשב מסוים (אם נרצה, מערכת מסוימת מסוג כלשהו ולא דווקא מחשב) חושב או לא. מחשב העומד במבחן זה הוא מחשב שתכניתו היא בעלת רמת המורכבות הדרושה, כלומר הוא מחשב חושב. מבחן זה מכונה היום על פי רוב "מבחן טיורינג" ("Turing Test"); טיורינג עצמו כינה אותו "משחק החיקוי" ("The Imitation Game"). במה כרוך מבחן זה? ובכן, בחדר מסוים נמצא המחשב העומד למבחן. בחדר אחר נמצא אדם. אנו ("השופטים") איננו יודעים באיזה חדר נמצא המחשב ובאיזה חדר נמצא האדם – מטרתנו לגלות זאת. דרכנו לגלות זאת באמצעות ראיונות זהים שאנו עורכים למחשב ולאדם. אנו מקלידים את שאלותינו על מקלדת, והן מוזנות הן לאדם והן למחשב. תשובות שניהם מופיעות על-גבי מסכים הניצבים מולנו. האדם מתבקש לענות באופן שיסייע לנו לגלות שהוא האדם; המחשב מתוכנת לענות באופן שיסווה את היותו מחשב – הוא מנסה לחקות אדם. כמובן, ככל שתוכנת המחשב מתוחכמת יותר,

<sup>65</sup> ביחידה 8 נפגוש בהשקפה, שלפיה כדי שיהיה לישות מאפיין נפשי ספציפי – מכוונות – עליה להיות בעלת היסטוריה מסוימת – היסטוריה אבולוציונית.

<sup>66</sup> תשובתו זו של טיורינג פותחה במאמרו: "Computing Machinery and Intelligence", *Mind* 59 (1950), pp. 433-460.



יהא עלינו, השופטים, להציג שאלות מתוחכמות יותר כדי להפיל את המחשב בפח, ולגרום לו שישב באופן שיאפשר לנו לעמוד על זהותו. אם למשל, נשאל את המחשב "האם אתה מחשב?", יהא עליו לשקר; אם נשאל אותו מה גילו, יהא עליו להשיב באופן שיתיישב עם תשובותיו לשאלות באשר לעבודתו, וכן הלאה. עתה, לטענת טיורינג, מחשב שיעמוד בהצלחה באתגר המתוחכם ביותר מסוג זה שיוצב בפניו, כך שלא נוכל להבחין על פי תשובותיו בינו לבין האדם, הוא מחשב חושב – הוא מחשב בעל אינטליגנציה.

טיורינג חשב שבימיו לא יהיה מחשב שמסוגל לעמוד בהצלחה באתגר זה (כאשר השאלות הנשאלות מתוחכמות מספיק, כמובן). הוא חזה שבתוך כ-50 שנה מן הזמן שבו פיתח את רעיון משחק החיקוי (בשנת 1950) יהיו מחשבים שיעמדו באתגר מתוחכם ביותר מסוג זה שנציב בפניהם, בהצלחה של 70%. תחזיתו זו לא התממשה: אנו עדיין מסוגלים להציב בפני המחשבים הקיימים היום (למעלה מ-50 שנה מאז) שאלות שיכשילו אותם.

שימו לב, שטיורינג לא טען שמחשב שנכשל במבחן זה אינו חושב. הוא טען רק את הטענה ההופכית, שלפיה מחשב שעובר מבחן זה אכן חושב. ואמנם יש טעמים טובים להתנגד לרעיון שמחשב שנכשל במבחן טיורינג אינו חושב. אנשים שונים עשויים להיכשל במבחן טיורינג, ואין פירוש הדבר שהם אינם חושבים. לכל היותר, כישלונם במבחן מהווה ראייה לכך שרמת האינטליגנציה שלהם אינה גבוהה. כישלון במבחן טיורינג עשוי גם להיות נעוץ בתפקוד לקוי בתחום אינטלקטואלי אחד (או בתחומים אינטלקטואליים אחדים), ואנו יודעים היטב שמי שתפקודו האינטלקטואלי לקוי בתחומים אחדים עשוי לתפקד היטב בתחומים אינטלקטואליים אחרים, כלומר תפקודו הלקוי בתחומים מסוימים רחוק מלהעיד על כך שאינו בעל כושר חשיבה בכלל. לפיכך, כשלון במבחן טיורינג אינו מהווה ראייה להיעדר כושר חשיבה.

לעומת זאת, מחשב שעובר בהצלחה את המבחן – מחשב שמשיב לשאלות המתוחכמות ביותר באופן שאינו נבדל מן האופן שבו בני אדם מסוימים, כשירים מבחינה שכלית, עשויים להשיב לשאלות אלו (והראייה לכך היא שאיננו מסוגלים להבחין בין המחשב לבין האדם על פי התשובות) – הוא מחשב חושב. כך טען טיורינג, ורבים השתכנעו.

שימו לב, שמבחן טיורינג **אינו מבחן פיזיקלי** לחשיבה. עניינו כלל אינו במבנהו הפיזיקלי של המחשב. יהא מבנהו הפיזיקלי אשר יהא, מחשב המספק תשובות הולמות לשאלות מתאפין בחשיבה.

## שאלה

אפשר לומר שמבחן טיורינג הוא מבחן ביהביוריסטי לחשיבה. מדוע?

עניינו של מבחן זה, כפי שאנו רואים, **בדגמי הקלט-פלט** של המחשב: השאלות הן הקלט, התשובות הן הפלט, ומתן תשובות הולמות לשאלות משמעו הפגנת דגמי קלט-פלט הולמים. במובן חשוב אפוא מבחן טיורינג הוא **מבחן ביהביוריסטי** לחשיבה. מחשב חושב, על פי מבחן זה, הוא כזה שהתנהגותו החיצונית היא כשל יצור חושב, כשל אדם.

ראוי להדגיש כי דגמי התנהגותו של מחשב העובר את מבחן טיורינג אינם זהים לחלוטין לדגמי התנהגותו של אדם. דגמי התנהגותו של מחשב כזה זהים לדגמי התנהגות אנושיות ממקטע חלקי בלבד – התנהגויות לשוניות. אולם אין זה סביר שניתן להתנגד על בסיס זה לטענה שמחשב העובר את מבחן טיורינג הוא ישות חושבת, משום שבתוך טווח ההתנהגויות האנושיות, סביר כי ההתנהגות הלשונית היא הביטוי החיצוני המובהק ביותר של חשיבה. בהנחה שהתנהגות חיצונית עשויה להוות ראיה טובה לקיומה של נפשיות, דומה שקיומה של התנהגות לשונית מתאימה – למשל, סיפוק תשובות לשאלות באופן זהה לזה של יצור נפשי כשיר מבחינה שכלית – מהווה ראיה טובה לקיומה של חשיבה. ייתכן שיש טעמים טובים לומר שמחשב המפגין התנהגות לשונית גרידא חסר מאפיינים נפשיים כלשהם (למשל, מאפיינים רגשיים), אך אם התנהגותו הלשונית הולמת במובן שבו אנו דנים, הרי, עד כמה שהתנהגות עשויה להוות ראיה לקיומה של נפשיות, יש טעם טוב להניח שמחשב זה מתאפיין בהיבט הרלוונטי של נפשיות – דהיינו שהוא חושב.

אולם האם ההנחה שהתנהגות חיצונית עשויה להוות ראיה טובה לקיומה של נפשיות היא הנחה סבירה? האם, במילים אחרות, מבחן ביהביוריסטי עשוי להיות מבחן הולם לנפשיות, וכך, באופן ספציפי, לחשיבה? מבחן טיורינג קדם-מניח שמכונה המפגינה את יחסי הקלט-פלט האופייניים ליצור נפשי היא יצור נפשי (או לפחות, שיש לה מאפיינים נפשיים מסוימים). אין זאת שפעילותה של מכונה כזו היא רק **חיקוי** של תפקוד נפשי, כמו חשיבה; לפי מבחן טיורינג, פעילותה של מכונה כזו היא **בעצמה** תפקוד נפשי, כמו חשיבה.

אין פירוש הדבר שמבחן טיורינג קדם-מניח את נכונותה של ההשקפה הביהביוריסטית. עלינו להבחין בהקשר זה בין שתי עמדות שונות. העמדה האחת היא כי התנהגות מסוימת מהווה **קריטריון** לחשיבה (או לנפשיות בכלל); העמדה השנייה היא כי התנהגות מסוימת מהווה **אבן בוחן** לחשיבה (או לנפשיות בכלל). על פי העמדה הראשונה, ההתנהגות העומדת על הפרק **מכוננת** חשיבה – זוהי העמדה הביהביוריסטית. על פי העמדה השנייה, ההתנהגות העומדת על הפרק מהווה **ראיה**, או **סימן** (אינדיקציה) לחשיבה – זוהי העמדה שאותה קדם-מניח מבחן טיורינג. אין הוא קדם-מניח כי התנהגות מסוימת היא חשיבה;<sup>67</sup> הוא קדם-מניח אך ורק שהתנהגות מסוימת היא ראיה לחשיבה. הרעיון שביסוד מבחן טיורינג מתיישב עם העובדה שחשיבה היא משהו הנבדל מהתנהגות (או מנטייה התנהגותית) מסוגים מסוימים – למשל, שחשיבה מכוננת על ידי ארגון פונקציונלי מסוים (או תוכנה מסוימת). במקרה זה, המבחן

<sup>67</sup> זאת לא רק משום שהרעיון שביסודו הוא, כאמור, שמי שעוברים את המבחן הם חושבים, אך לא שמי שנכשלים בו אינם חושבים.

ההתנהגותי הוא מבחן לקיומו של הארגון הפונקציונלי המורכב ברמה הדרושה, דהיינו, יש טעם טוב לחשוב שמי שמתנהגים באופנים הנדרשים מתאפיינים בארגון הפונקציונלי המכונן חשיבה. מבחן טיורינג הוא מבחן לחשיבה, משום שהוא מהווה מבחן לקיומו של ארגון פונקציונלי בעל מורכבות מסוימת, וארגון פונקציונלי זה מכונן חשיבה.

לפי מבחן טיורינג התנהגות מסוימת היא אבן בוחן לחשיבה, ואינה קריטריון לחשיבה: ההתנהגות הנדרשת למעבר המבחן היא **ראיה** לקיומה של חשיבה, אך **אינה מכוננת** חשיבה. מבחן טיורינג הוא אבן בוחן לחשיבה במובן מוגבל, "חד-כיווני": מעבר המבחן הוא ראיה לחשיבה, אך כישלון במבחן אינו ראיה להיעדר חשיבה.<sup>68</sup>

נקודה חשובה נוספת באשר למהותו של מבחן טיורינג היא שהוא הוצג כמבחן **לעצם קיומה של חשיבה**, ולא כמבחן לחשיבה **באופן מסוים**: הרעיון אינו שכל מי שמשיבים לאותן שאלות באותם אופנים חושבים באותם אופנים. טיורינג ידע היטב, כפי שמעידה עבודתו על מכונת טיורינג, שהתנהגויות זהות אינן מהוות ראיה לארגונים פונקציונליים זהים.

#### שאלה

איזו טענה שעלתה בדיון במכונת טיורינג משמעה הוא שהתנהגויות זהות אינן מהוות ראיה לארגונים פונקציונליים זהים?

הטענה הרלוונטית היא שמכונות טיורינג שונות יכולות לבצע את אותם חישובים – להגיע מאותם קלטים לאותם פלטים – **בדרכים שונות**, דהיינו הן עשויות להתאפיין על ידי ארגונים פונקציונליים שונים. כפי שראינו, תכניות שונות של מכונת טיורינג – תכניות הפועלות על פי כללים שונים ומבצעות צעדים שונים בהתאם לכללים השונים הללו – עשויות להיות תכניות שקולות קלט-פלט, כלומר להפגין את אותם דגמי התנהגויות. משמע שהפגנת אותם דגמי התנהגויות אינה מהווה ראיה לקיום של ארגונים פונקציונליים זהים.

אם אחת משתי תכניות מחשב שהן שקולות פלט-קלט תעבור את מבחן טיורינג, כך גם רעותה, שהרי מבחן זה אין עניינו אלא בהתנהגות, דהיינו ביחסי קלט-פלט. ותכניות כאלו, כאמור, עשויות להיות שונות מאוד זו מזו, להיות בעלות ארגונים פונקציונליים שונים. ובכן, על פי הרעיון שביסוד מבחן טיורינג, אם מחשב הפועל בהתאם לאחת משתי תכניות כאלו חושב, גם מחשב הפועל בהתאם לרעותה חושב; אולם רעיון זה אינו מחויב לכך שהמחשבים הללו חושבים **באותו אופן**. רעיון זה מתיישב היטב עם ההשקפה הפונקציונליסטית, שעל

<sup>68</sup> יש לציין שפילוסופים אחדים רואים במבחן טיורינג קריטריון לחשיבה, ולא רק אבן בוחן לחשיבה. דומה כי טיורינג עצמו סבר שבעתיד, כאשר מחשבים יעברו את המבחן, הוא יהפוך לקריטריון לחשיבה. אנו נתייחס למבחן **טיורינג** כאבן בוחן לחשיבה – כפי שנראה, גם כשהוא מפורש באופן זה, הוא נתקל בקשיים רציניים.

פיה שקילות קלט-פלט בין התכניות שעל פיהן פועלות שתי ישויות אינה מחייבת שהן מתאפיינות על ידי אותם תהליכי חשיבה. אם אמנם התכניות שונות, הרי לפי הפונקציונליזם, הישויות הללו (ויהיו הן מחשבים או ישויות אחרות) **אינן מתאפיינות על ידי אותם תהליכי חשיבה**. במילים אחרות, מתקיימים ביניהן הבדלים נפשיים.

מבחן טיורינג הוא מבחן לעצם קיומה של חשיבה – לעצם היותה של מערכת מערכת חושבת, ואינו מבחן לזהותם של תהליכי חשיבה; תשובות זהות במבחן טיורינג אינן ראייה לתהליכי חשיבה זהים.

בעבור מי שמקבלים את ההשקפה הביהביוריסטית, מבחן טיורינג הוא מבחן מתבקש. אם אמנם חשיבה מכוננת על ידי דגמי התנהגות מסוימים, הרי מי שמפגינים דגמי-התנהגות אלו הם בגדר חושבים. אולם אם, כפי שאמרנו זה עתה, הרעיון שביסוד מבחן טיורינג אינו כרוך בהשקפה הביהביוריסטית, מדוע עלינו לקבלו כמבחן הולם לחשיבה? במילים אחרות, מדוע שקילות קלט-פלט בין מערכת מסוימת לבין יצור חושב (כאדם) היא ראייה לכך שמערכת זו היא מערכת חושבת? ובאופן ספציפי יותר, מדוע על פונקציונליסטים לקבל מבחן זה? מדוע על מי שסוברים כי חשיבה, ונפשיות בכלל, מכוננות על ידי ארגון פונקציונלי מסוים, לקבל **מבחן ביהביוריסטי** כזה לקיומה של חשיבה, כאשר הם דוחים **קריטריון ביהביוריסטי** לחשיבה?

מאחר שבעבור הפונקציונליסטים חשיבה (כמו נפשיות בכלל) מכוננת על ידי ארגון פונקציונלי ברמת מורכבות מסוימת, השאלה בעבורם היא מדוע שקילות קלט-פלט בין מערכת מסוימת לבין יצור חושב (כאדם) היא ראייה לכך שלמערכת זו ארגון פונקציונלי ברמת מורכבות כמו ארגונו הפונקציונלי של יצור חושב. פונקציונליסטים הדוגלים במבחן טיורינג אינם יכולים להצדיק מבחן זה בטענה ששקילות קלט-פלט היא ראייה **לשקילות** פונקציונלית, שהרי, כפי שחזרנו והדגשנו, תכניות שקולות קלט-פלט עשויות להיות שונות מאוד זו מזו, להיות בעלות ארגונים פונקציונליים שונים. ובכן, ייתכן שניתן להצדיק את הרעיון שביסוד מבחן טיורינג במסגרת פונקציונליסטית באופן הבא: הצלחה במבחן טיורינג אינה סתם הצלחה בחיקוי דגמי קלט-פלט. זוהי הצלחה בחיקוי דגמי קלט-פלט מורכבים למדי, כאלו שמאפיינים חשיבה אנושית תקינה. זוהי הצלחה שדורשת תחכום רב למדי. נכון שמחשב שמצליח לחקות דגמי קלט-פלט מורכבים, ובאופן שיטתי, אינו חייב להיות זהה בארגונו הפונקציונלי לאדם המתאפיין באותם דגמי קלט-פלט, אבל הוא חייב בהחלט להיות בעל ארגון פונקציונלי מורכב ומתוחכם ביותר. נכון שמחשב שמתמודד בהצלחה עם אותן שאלות ובעיות, שעמן מתמודד בהצלחה אדם בעל חשיבה תקינה, אינו חייב להיות אינטליגנטי באותה מידה כמו אדם כזה – ייתכן שהוא מתמודד עם הבעיות באופן מסורבל ובדרך שאינה חכמה במיוחד (כמובן, ייתכן גם שהוא מתמודד עם הבעיות בדרך חכמה ואלגנטית יותר). אבל אם הבעיות אמנם מורכבות והוא מתמודד עמן בהצלחה, הוא חייב להיות אינטליגנטי ומתוחכם במידה משמעותית; אם

אמנם מורכבותו של הארגון הפונקציונלי היא הקריטריון לחשיבה, ארגונו הפונקציונלי של מחשב כזה חייב להיות מורכב במידה משמעותית.

קו מחשבה זה עשוי לשכנע ממבט ראשון, אולם אפשר להתנגד לו באופן הבא: גם דגמי הקלט-פלט המתוחכמים ביותר ניתנים לחיקוי באמצעות תוכנות "טיפשות" למדי (ואולי אף "טיפשות" באופן מוחלט, כפי שנראה מייד), כאלו שארגון הפונקציונלי אינו בעל רמת מורכבות גבוהה במיוחד. נחשוב על מחשב המבצע חישובים מתמטיים סבוכים ביותר. מחשב עשוי אמנם לבצע חישובים כאלו באופנים דומים לאופנים בהם אנו מבצעים חישובים מתמטיים, אולם קיימת גם אפשרות אחרת. ייתכן שתוצאות החישובים פשוט הוזנו לזיכרון המחשב, וכשעל המחשב לבצע "חישוב" מסוים, הוא פשוט שולף את התשובה הנכונה מזיכרונו. דומה שלא נרצה לזכות מחשב זה בתואר "חושב", אולם לפי מבחן טיורינג, שעניינו ביחסי קלט-פלט בלבד, מחשב כזה חושב. למעשה, דוגמה זו אינה בדיונית לחלוטין. עד לפני שנים לא רבות פעלו רוב מחשבי השחמט באופן זה: הוזנו לזיכרוןם משחקי שחמט של מיטב רבי-האומנים, והם חיקו את המהלכים שעל פי רוב הוליכו לניצחונות.<sup>69</sup> מחשבים כאלו לא הפעילו הליכי חשיבה בניסיון למצוא את המהלך הנכון – הם שלפו מזיכרוןם את המהלך שהוכח כיעיל. דומה שלא נרצה להתייחס למהלכים כאלו של שליפה מן הזיכרון כאל חשיבה. האין דוגמאות כאלו חותרות תחת הרעיון שביסוד מבחן טיורינג, לפיו יחסי קלט-פלט הולמים הם אכן ראייה לקיומה חשיבה? האם אין דוגמאות כאלו מראות שמבחן טיורינג לוקה בליברליזם-יתר?<sup>70</sup>

דומה שלהתנגדות זו כוח רב כאשר מדובר בתוכנות מחשב המוגבלות למשימות ספציפיות שונות, כמו משחק שחמט או חישובים מתמטיים. אולם ספק אם התנגדות זו מצליחה לחתור תחת עצם הרעיון של מבחן טיורינג. זאת משום שייתכן שכאשר מדובר בתחומים אחרים הדורשים תבונה – למשל, הבנת בני אדם אחרים, הבנת יצירות ספרות, או הבנת תהליכים חברתיים והיסטוריים – "תוכנות-זיכרון" לא יצליחו להפיק את אותן תשובות כבני אדם. במילים אחרות, ייתכן ששאלון של מבחן טיורינג שיהיה מתוחכם מאוד ואולי מפורט מאוד יוכל להפיל בפח מחשבים הפועלים על-סמך זיכרון בלבד.

על הדיון באפשרות זו להיכרך בדוגמאות של תוכנות כאלו ושאלונים המנסים להכשילן. אנו נוותר על דיון זה. נסתפק בחזרה על העובדה, שאנו מסוגלים להציב בפני המחשבים הקיימים היום שאלונים שיכשילו אותם. עתה נציג ביקורת כנגד ההשקפה הפונקציונליסטית והשקפות הקרובות אליה, הכרוכה גם בהתנגדות עקרונית לעצם הרעיון של מבחן טיורינג. לפי ביקורת זו, יש מובן חשוב שבו גם המחשב "המתוחכם" ביותר איננו חושב.

<sup>69</sup> דוגמת השחמט הובאה בהקשר דומה על ידי לואו (E.J. Lowe) בספרו:

*An Introduction to Philosophy of Mind*, Cambridge: Cambridge University Press, 2000, p. 210.

<sup>70</sup> בנקודה זו אפשר לראות מדוע הדיון שלעיל בסוגיית הקלט והפלט אינו רלוונטי לדיונינו הנוכחי: נניח שמחשב השחמט הנדון בדוגמה זו יישתל בגופו של רובוט, אשר יתפקד כשחקן שחמט לכל דבר: יראה לוח שחמט ממש, ויזיז בעצמו את הכלים על הלוח בתגובה ל"חישוב" של המחשב שבקרבו. ברור שמי שחושב שהמחשב הנדון אינו חושב, לא ישנה את דעתו לאור השתלת המחשב בגופו של רובוט כזה.

## 8. טיעון החדר הסיני כנגד ההשקפה הפונקציונליסטית

להשקפה הפונקציונליסטית תומכים רבים, אולם הועלו כלפיה גם התנגדויות רבות ומגוונות. עתה נסקור אחדות מהתנגדויות אלו.

ההתנגדות לפונקציונליזם שבה נדון עתה היא התנגדותו של הפילוסוף ג'ון סרל. התנגדות זו פורסמה תחילה במאמר שהופיע ב-1980 בכתב-העת *Behavioral and Brain Sciences*, יחד עם תגובות של פילוסופים, מדענים העוסקים בבנייה מלאכותית, פסיכולוגים וחוקרים אחרים העוסקים במדעים הקוגניטיביים, ותשובותיו של סרל לתגובות אלו. הפולמוס סביב התנגדות זו נמשך בעוצמה שנים אחר-כך, והדיו עדיין נשמעים.<sup>71</sup>

נבהיר מונחים אחדים המופיעים במאמר. התזה שסרל מכנה **"התזה החזקה של בינה מלאכותית"** ("Strong AI") שקולה, פחות או יותר, למודל המחשב של הנפש שבו דוגלים הפונקציונליסטים, וכך אנו נבין אותה.<sup>72</sup> "SHRDLU" ו-"ELIZA" הם שמות של תכניות ספציפיות של בינה מלאכותית. את תופעת **המכוונות** (Intentionality) הצגנו ביחידה 1, ונדון בה בהרחבה רבה ביחידה 8. אולם עתה נחזור ונבהיר בקצרה את משמעות המונח "מכוונות". היותם של מצבים נפשיים בעלי מכוונות משמע שהם נסבים על אודות ישויות או מצבי עניינים שונים. למשל, האמנתי שיורד גשם מכוונת אל, או נסבה על אודות, ירידת הגשם. דרך אחרת להביע את הרעיון שלהאמנתי שיורד גשם יש מכוונות היא לומר שיש לה **תוכן** – התוכן שלה הוא (הטענה) **שיורד גשם**. גם משפטים בשפה ופעולות דיבור נסבים על אודות ישויות או מצבי עניינים: לא רק ההאמנה שיורד גשם נסבה על אודות ירידת הגשם, אלא גם האמירה שיורד גשם. דרך אחרת להביע את הרעיון שלאמירה שיורד גשם יש מכוונות היא לומר שיש לה **משמעות** – משמעותה היא (הטענה) שיורד גשם (כלומר, את קיומה של המכוונות בתחום הנפשי מאפיינים באמצעות המונח "תוכן" ואת קיומה של המכוונות בתחום הלשוני מאפיינים באמצעות המונח "משמעות"). אנו מזכירים זאת משום שהמונח "משמעות" מופיע בדיונו של סרל.<sup>73</sup>

<sup>71</sup> ראו, בין היתר, את הפולמוס בין סרל לבין פול מ. צ'רצילנד (Paul M. Churchland) ופטרישיה ס. צ'רצילנד (Patricia S. Churchland) בכתב העת *Scientific American*, ינואר 1990. לאחרונה יצא לאור ספר המוקדש כולו לטיעון החדר הסיני:

J. Preston, and M. Bishop (eds.), *Views into the Chinese Room: Essays on Searle and Artificial Intelligence*. New York: Oxford University Press, 2002.

סקירה של תגובות שונות לטיעונו של סרל ראו

D. Cole, "The Chinese Room Argument", in *Stanford Encyclopedia of Philosophy* (Fall 2004 Edition), Edward N. Zalta (ed.). URL = <http://plato.stanford.edu/archives/fall 2004/entries/chinese-room/>.

<sup>72</sup> למעשה, כפי שנבהיר בהמשך, תזה זו שקולה לגרסה מסוימת של ההשקפה הפונקציונליסטית.

<sup>73</sup> על הקשר בין תוכן (נפשי) למשמעות (לשונית), ועל הטעם להבחין ביניהם, נעמוד ביחידה 8.

נקודה נוספת שנבהיר עניינה במונח "סמל צורני" ("formal symbol"). נחזור לרגע למכונת טיורינג. המכונה מגיבה באופנים שונים, הנקבעים על פי כללי המכונה, לסימנים השונים שהיא קוראת. מה שמשנה בעבור המכונה זו הצורה של הסימן. ניתן להציג את כללי הפעולה של המכונה באופן כדוגמת זה: אם המכונה נמצאת במצב פנימי  $S_1$  אז, אם היא מזהה בתא הנסרק סימן שצורתו "+", עליה לפעול כך וכך (פעולות העשויות לכלול הדפסת סימן בעל צורה מסוימת); ואם היא מזהה בתא הנסרק סימן שצורתו "1", עליה לפעול כך וכך, וכך הלאה. הכללים של המכונה – למעשה, כללי התוכנה שעל פיהם מתפקד כל מחשב – רגישים לצורה של הסימנים בלבד (ולא, למשל, למה שהם מסמנים). כללים כאלו נקראים "כללים צורניים", או "פורמליים", או "סינטקטיים", והסימנים שעליהם פועלים כללים כאלו נקראים "סמלים צורניים". מתייחסים אל אובייקטים כאל סמלים צורניים בהקשר שבו מה שמשנה זו צורתם, ובהקשר של תפקודה של מכונת טיורינג – למעשה בהקשר של כל תוכנת מחשב – אכן מה שמשנה זו צורתם של הסימנים, שכן כללי התוכנה רגישים לצורתם בלבד.

כללי התוכנה שעל פיהם מתפקדים מחשבים הם כללים צורניים הפועלים על סמלים צורניים – הם רגישים רק לצורת הסימנים (ולא, למשל, למה שהם מסמנים).

קראו נא עתה בנספח המצורף ליחידה את מאמרו של סרל, "נפש, מוח ותוכנות מחשב" ("Minds Programs and Brains").<sup>74</sup>

## 8.1 החדר הסיני ומבחן טיורינג

במרכז מאמרו של סרל עומד ניסוי החשיבה של "החדר הסיני". גיבורו של ניסוי חשיבה זה, האדם המאייש את "החדר הסיני", אינו דובר סינית. עם זאת הוא מסוגל לתפקד בהקשר הנדון כדובר סינית – הוא מתפקד כמי שמבין את הסיפור הכתוב בסינית.

### שאלה

סרל אינו עוסק ישירות במבחן טיורינג,<sup>75</sup> אולם יש קשר בין מבחן זה לבין העובדות שהזכרנו זה עתה. מהו לדעתכם הקשר בין עובדות אלו לבין מבחן טיורינג?

<sup>74</sup> הטעון המרכזי של סרל הופיע גם במאמרו "Can Computers think?", מאמר המהווה את פרק 2 של ספרו של סרל: J.R. Searle, *Minds, Brains, Science*, London: Penguin Books, 1984. במאמר זה מופיעים מעט מונחים טכניים שיייתכן ואינכם מכירים.

<sup>75</sup> הוא מתייחס למבחן זה בתגובותיו לביקורות שהועלו כנגד המאמר.

דומה שהאיש בחדר הסיני היה עובר בהצלחה את מבחן טיורינג. אמנם תחום ה"שאלות" שהופנו אליו על פי ניסוי-החשיבה של סרל הוא מוגבל, אבל דומה שאין קושי עקרוני לדמיין (גם אם האפשרות המעשית הזו נראית קלושה) שאדם כזה משיב בהצלחה לשאלות רבות ומגוונות בתחומים רבים ומגוונים – שהוא מתפקד כמו דובר סינית רגיל, דובר המבין את הסיפור ואת השאלות המופנות אליו. יחסי הקלט-פלט של אדם זה – המתאמים בין השאלות המופנות אליו לבין תשובותיו – עשויים לאפיין דובר סינית רגיל, המבין את השאלות. ואמנם, כפי שמציין סרל, מי שצופים מבחוץ יחשבו שהאדם שבחדר הסיני דובר סינית ומבין את הסיפור, את השאלות, ואת והתשובות. אולם אדם זה (זוהי הנחת ניסוי החשיבה) אינו דובר סינית, כך שאין הוא מבין כלל את הסיפור, את השאלות, ואת התשובות שהוא מנפיק. כל מה שהוא יודע זה שמה שמסמן רצף הסימנים "@##" הוא תגובה נכונה למה שמסמן רצף הסימנים "%#%". אולם הוא עובר את מבחן טיורינג באשר לסיפור זה, שכן צופים מבחוץ לא יוכלו להבחין בין תשובותיו לבין התשובות שהיו ניתנות על ידי דובר סינית. כלומר, אפשרי מצב שבו אדם עובר את מבחן טיורינג בנושא שבו אינו מבין דבר. זהו אפוא אבסורד שאליו נקלע הרעיון של מבחן טיורינג (שימו לב לניסוחה של טענה זו: האבסורד הנדון אינו נעוץ בכך שהאדם עובר את מבחן טיורינג אינו בעל תבונה, שכן האדם בחדר הסיני הוא בעל תבונה, תבונה המאפשרת לו להבין את מדריך הכללים. ההתקפה על הרעיון שביסוד מבחן טיורינג היא עקיפה: האבסורד נעוץ בכך שהוא עובר את מבחן טיורינג בנושא שבו אינו מבין דבר).

האמנם ניתן לראות בניסוי-החשיבה של סרל את הבאתו של מבחן טיורינג לאבסורד? לדיונו בנושא הישיר של ניסוי-חשיבה זה – השלכתו על ההשקפה הפונקציונליסטית, ובאופן ספציפי יותר, על מודל המחשב של הנפש – רלוונטיות לשאלה זו. נפנה עתה לדון בנושא זה.

## 8.2 התוכנה הנכונה, הבנה ומכוונות

כאמור, האדם בחדר הסיני אינו מבין את הסיפור, את השאלות ואת והתשובות שהוא מנפיק. אבל מודל המחשב של הנפש מחויב לכך שהוא מבין את הסיפור, ולפיכך יש לדחות את מודל המחשב. זהו גרעין טיעונו של סרל.

### גרעין טיעון החדר הסיני:

- (1) האדם בחדר הסיני אינו מבין את הסיפור, את השאלות ואת התשובות.
  - (2) אם מודל המחשב של הנפש אמת, האדם בחדר הסיני מבין את הסיפור, את השאלות ואת התשובות.
- לפיכך,  
מודל המחשב של הנפש אינו אמת.



קשה לחלוק על ההנחה הראשונה של הטיעון – על ההנחה שהאדם אינו מבין את הסיפור. מה באשר להנחה השנייה? מדוע, לטענת סרל, מחויב מודל המחשב של הנפש לטענה שהאדם בחדר הסיני מבין את הסיפור? משום שלפי מודל זה הבנה (כמו כל מאפיין נפשי) מתמצה בהרצת התוכנה המתאימה, בתפקוד לפי תכנית ספציפית או, בלשון אחרת, לפי כללי מעבר ספציפיים ממצב למצב, והאדם בחדר הסיני פועל לפי הכללים הספציפיים הנדרשים: תפקודו מתאפיין על ידי אותם מעברים בין מצבים שהיו מאפיינים את תפקודו של דובר סינית אשר היה נשאל את אותן שאלות על אותו סיפור; הוא מגיב לצורות הסמלים הרלוונטיים באופן הדרוש. הוא אמור אפוא להבין את הסיפור.

אם אמנם האדם בחדר הסיני פועל על פי התוכנה הדרושה, אם הוא פועל על פי הכללים הצורניים הנכונים ומגיב באופנים הנכונים לסמלים צורניים, אולם הוא חסר הבנה, הבנה אינה יכולה להתמצות בתפקוד על פי התוכנה הנכונה, בתמרון הנכון של סמלים צורניים. לפיכך מודל המחשב איננו מודל הולם של הנפש. מודל זה מחמיץ גורם חשוב בחיינו הנפשיים.

טיעונו של סרל רלוונטי, כמובן, לשאלה האם מחשבים יכולים לחשוב, אם כי הקשר בינו לבין שאלה זו אינו מאוד פשוט. על פי מודל המחשב של הנפש (ליתר דיוק, לפי מודל זה בגרסתו המופשטת לחלוטין), מחשבים יכולים לחשוב. הם יחשבו, אם רק יריצו את התוכנה הדרושה, תוכנה הדומה במידה מספקת לתוכנה של הפסיכולוגית האנושית. ניתן לחשוב שטיעונו של סרל מיועד להראות שמחשבים אינם יכולים לחשוב. אולם אין זה מדויק. טיעונו של סרל, אף אם הוא מוצלח, איננו מראה שמחשבים אינם יכולים לחשוב. אם הוא מוצלח, הוא מראה שמחשבים אינם יכולים לחשוב **רק בשל כך שהם מריצים את התוכנה הנכונה**. התוכנה איננה הגורם הרלוונטי לקיומה של חשיבה בפרט ונפשיות בכלל. לפי השקפתו של סרל באשר לטבעם של המצבים הנפשיים, מחשבים שיתקיימו באשר אליהם תנאים מסוימים הקשורים בחומרה, יהיו מסוגלים לחשוב.<sup>76</sup> טיעון החדר הסיני עניינו בקשר בין מצבים נפשיים לבין הרצת תוכנה בלבד.

אם טיעון החדר הסיני מוצלח, הוא מראה שמחשבים אינם יכולים לחשוב **רק בשל כך שהם מריצים את התוכנה הנכונה**. אין די בהרצת התוכנה הנכונה כדי לכוון מחשבה.

המערכת של האדם בחדר הסיני היא שקולת קלט-פלט לדובר סינית בהקשר הנדון.

<sup>76</sup> על השקפתו של סרל באשר למצבים הנפשיים ראו בעיקר את ספרו

*The Rediscovery of the Mind*, Cambridge (Mass.): MIT Press, 1992.

## שאלה

לו הייתה המערכת של האדם בחדר הסיני שקולת קלט-פלט **בלבד** לדובר סינית, האם יכולה הייתה דוגמה זו לחתור תחת מודל המחשב של הנפש?

לו הייתה מערכת האדם בחדר הסיני שקולת קלט-פלט **בלבד** לדובר סינית, דוגמה זו לא הייתה יכולה להפריך את מודל המחשב. במקרה כזה הייתה הדוגמה מפריכה רק מודל ביהביוריסטי, ואילו מודל המחשב איננו מודל ביהביוריסטי. מצבים ותהליכים נפשיים מזוהים על ידינו גם על פי הקשרים הפנימיים בין המצבים שבהם הם כרוכים – על פי העיבוד הפנימי. ואמנם דוגמת החדר הסיני אינה אמורה להדגים שקילות קלט-פלט בלבד. מטרתה להתייחס לתהליכי עיבוד פנימיים. אמנם הדבר לא נזכר במפורש, אבל על פי אופן ההבנה הסביר של הדוגמה יש לראותה כמייחסת לאדם מצבים פנימיים ותהליכים פנימיים שונים. על האדם לקלוט בחושיו את הטקסט של הסיפור, לקלוט בחושיו את הטקסט של השאלות, לחבר (לערוך סינתזה) ביניהם, לזהות את הכללים הרלוונטיים. בתהליך זה כרוכים באורח בלתי-נמנע מצבים פנימיים שונים.

כדי שהדוגמה תשכנע ואמנם תהווה הפרכה של מודל המחשב של הנפש, עליה להיות כרוכה בארגון פונקציונלי פנימי **הזהה** לארגון הנפשי של דובר סינית. האומנם היא כרוכה בארגון פונקציונלי זהה? ניתן לטעון שלא, למשל משום שעיבוד המידע של האדם בחדר הסיני, כאשר הוא משיב על השאלות, כרוך בהתבוננות במדריך הכללים ובעיבוד המידע הרלוונטי הכלול בו. דובר סינית אינו נזקק לתהליכים כאלו. אולם לסרל תשובה מוכנה לטענה זו: על פיו, ביכולתנו לשנות מעט את הדוגמה, ולהניח שהאדם בחדר הסיני לומד את מדריך הכללים בעל-פה. הוא **מפנים** את הכללים, ואינו נזקק כלל להתבונן במדריך (ייתכן שלאחר אימון ותרגול ממושכים, הוא אף יפעל בהתאם לכללים באופן אוטומטי, מבלי להיות מודע אליהם בכלל). במצב זה הארגון הפונקציונלי של תפקודו הנפשי יתקרב במידה משמעותית לתפקודו הנפשי של דובר סינית, למרות שאינו מבין מילה בסינית.

האם לאחר הפנמת הכללים יהיה ארגונו הפונקציונלי של האדם בחדר הסיני **זהה** לזה של דובר סינית? ספק. כפי שמתואר ניסוי החשיבה, סביר מאוד שיתקיימו ביניהם הבדלים שונים. לדוגמה, דובר הסינית מפעיל את התכונה המכונה "יצירתיות": כשדובר סינית נשאל שאלה, הוא אינו שולף משפט שלם ממאגר המשפטים שלמד. הוא מרכיב בעצמו את המשפט ממאגר המילים שלמד, ולפי כללי ההרכבה שלמד, וכך הוא יכול להשיב באמצעות משפטים שמעולם לא נתקל בהם. למשל, הוא יוכל להשיב לשאלה מסוימת באמרו (בסינית) "ההמבורגר היה טעים" אף אם מעולם לא נתקל במשפט זה, אולם נתקל במשפטים כמו "הסטייק היה טעים" ו"ההמבורגר היה יבש". האדם בחדר הסיני (כפי שמתואר ניסוי החשיבה) אינו מפעיל יצירתיות כזו. הכללים מורים לו אילו **משפטים** עליו להשיב ("אם אתה נשאל שאלה '&^', ובסיפור כתוב '@\$#', ענה '^&#^'). עליו לשלוף משפט שלם בסינית (את

המשפט הנכון, כפי שמכתיבים הכללים והסיפור) ממאגר המשפטים הנקובים בכללים שהפנים.

דומה שאפשר לתקן את דוגמת החדר הסיני כך שתתגבר על הבדל זה והבדלים נוספים. ניתן להניח שהאדם בחדר הסיני מופעל על פי כללים מתוחכמים יותר, שיאפשרו לו, למשל, להפעיל את אותה יצירתיות לשונית שמפעיל דובר סינית. אף על פי שהתמונה במקרה זה תהיה סבוכה מאוד, היא אינה כרוכה בבעיה עקרונית. אולם, האם ביכולתנו לקבוע שכאשר האדם בחדר הסיני יפעל על פי כללים מתוחכמים ביותר – שאותם נתקשה מאוד לדמיין – הוא עדיין לא יבין סינית ולא יבין את הסיפור? האם פעולה מאוד מתוחכמת על פי כללים, פעולה הכוללת יצירתיות לשונית, עדיין אינה מקנה הבנת שפה? האם תוכנה עשירה יותר, שתכלול גם כללים שעל פיהם מעובד מידע הנקלט בחושים, כללים שעל פיהם אנו מגבשים האמנות, כללים שעל פיהם אנו מבצעים היסקים, עדיין איננה מקנה הבנה? ובכן, סרל משיב על-כך בשלילה. לטענתו, אין זה משנה כמה מתוחכמת תהיה פעולת האדם בחדר הסיני. העובדה שאנו משוכנעים שהאדם בחדר הסיני – בגרסתה הראשונה של הדוגמה – אינו מבין את הסיפור, אינה נעוצה בהיעדר תחכום בפעולתו, או בכך שפעולתו אינה זהה לחלוטין לפעולתו של דובר סינית. עובדה זו נעוצה בכך שהוא פועל על פי כללים צורניים בלבד; בכך שפעולתו מוכתבת רק על ידי הצורה של הסימנים שהוא סוקר. כל עוד הוא פועל באופן זה, ולו בדרגת התחכום הגבוהה ביותר, אין הוא מבין דבר מן הסיפור. מודל המחשב של הנפש מתאר תפקוד נפשי במונחים של פעולה על פי כללים צורניים בלבד, ומטעם זה הוא מחמיץ משהו חשוב בחיינו הנפשיים: הוא מחמיץ את ממד ההבנה.

הבנת שפה בודאי דורשת תחכום מסוים. סרל אינו חולק על-כך. טענתו היא שהבנת שפה כרוכה בדבר מה נוסף. אם אני מבין שפה מסוימת, כשאני משתמש במילותיה **יש להן משמעות, או מכוונות**. המילה "המבורגר" מכוונת להמבורגר – היא נסבה על אודות המבורגר. אולם המילה הסינית הנסבה על אודות המבורגר בשימוש של דובר סינית (נניח "\$^#@###"), איננה נסבה על אודות המבורגר בשימוש של האדם בחדר הסיני. היא איננה נסבה על אודות דבר; היא חסרת משמעות. שהרי האדם אינו מקשר מילה זו עם המבורגרים. למעשה, הוא אינה מקשר אותה עם שום דבר בעולם. תפקידה של מילה זו בחייו הנפשיים מתמצה בכך שפועלים עליה כללים צורניים. בהקשר זה היא סמל צורני בלבד. למשמעותה – למה שהיא מסמנת – אין כל תפקיד. משום כך, גם אם תפקודו של האדם הכרוך במילה זו יהיה מתוחכם ביותר, גם אם הכללים על פיהם הוא פועל כאשר הוא פוגש בה והכללים הכרוכים בהפקת משפטים שבהם היא נכללת יהיו מתוחכמים ביותר – אף אם הם יהיו זהים לכללים על פיהם פועל דובר סינית, המילה בשימוש תהיה חסרת משמעות, והוא לא יבין שום משפט שבו היא נכללת. פעולה על פי כללים צורניים פשוט אינה מספקת להקניית משמעות, ולפיכך להקניית הבנה.

המסקנה העולה מטיעונו של סרל היא שמודל המחשב של הנפש, וכך ההשקפה הפונקציונליסטית בכללה (שהרי כאמור הרעיון שביסודו של מודל זה מהותי לכל גרסאותיה

של ההשקפה הפונקציונליסטית), אינם מייצגים נאמנה את החיים הנפשיים על כל היבטיהם. הם מותירים מחוץ לתמונה היבט חשוב של חיים אלו – את מכוונותם כלפי היותם נסבים על אודות – אובייקטים בעולם החוץ-נפשי. אם טיעון זה מוצלח, עלינו לדחות אפוא את ההשקפה הפונקציונליסטית, או לפחות לא לקבלה כהשקפה ממצה על אודות הנפש.

### 8.3 ביקורת טיעון החדר הסיני – מכוונות, מחשב ורובוט

טיעונו של סרל עורר דיון נרחב וסוער, והותקף מכיוונים שונים. אנו נתייחס כאן בקצרה לאחד הכיוונים הללו. תחילה נעיר, שהטיעון משכנע יותר כטיעון נגד גרסה ספציפית של פונקציונליזם – גרסת הפונקציונליזם המופשט לחלוטין, המתייחסת גם אל הקלט והפלט של חיי הנפש אך ורק על פי "מיקומם בתוכנה". נראה ש"התזה החזקה של בינה מלאכותית" ("Strong AI") שאותה מבקר סרל שקולה לגרסה מופשטת זו של פונקציונליזם ומודל המחשב, וגרסה זו אמנם פגיעה יותר לביקורתו של סרל. במסגרתה של גישה כזו קשה לתת פשר להיותן של מילים נסבות על אודות דברים בעולם החוץ-נפשי. סביר להניח שמה שמקנה למצבים הנפשיים את מכוונותם לעולם – לכיסאות, לעצים, להמבורגרים – מה שעושה אותם נסבים על אודות אובייקטים חוץ-נפשיים – קשור באופן כלשהו לקלט ולפלט של המערכת הפסיכולוגית. הקלט והפלט הם נקודות החיבור בין המצבים הנפשיים לבין העולם החוץ-נפשי. אולם כאשר הקלט והפלט מאופיינים באופן צורני בלבד, המגע עם העולם החוץ-נפשי אובד. אם אף קלט אינו מאופיין כתפיסה חושית של המבורגר, למשל, או כפעולה המיועדת לאכול המבורגר, למשל – אם האפיון של הקלט והפלט ובעקבותיו האפיון של המערכת כולה הוא אפיון צורני שאדיש למאפיינים בעולם החיצון או לקשר ממשי איתם – אפיון זה אינו מסוגל להקנות מכוונות ומשמעות למצבים הנפשיים, לעשותם נסבים על אודות דברים כמו המבורגרים.

אולם הדברים שונים כשמדובר באפיון ממשי של הקלט והפלט. טענה סבירה היא שכדי שמונח בשימושי יהיה מכוון להמבורגר, אני צריך, בין היתר, להיות מסוגל לקשור אותו להמבורגר. למשל, אני צריך להיות מסוגל להחיל את המונח כשאני רואה המבורגר. אכן, האדם בחדר הסיני אינו מסוגל להחיל את המונח הסיני הרלוונטי על המבורגרים, אבל תיאור פעולתו של האדם בחדר הסיני הוא כתיאור פונקציונלי מופשט. אם אנו רוצים לבחון גרסאות של פונקציונליזם המאפיינות את הקלט והפלט באופן שקשור לעולם, עלינו לשנות את ניסוי החשיבה. למשל, עלינו להניח שלאדם יש מצלמת טלוויזיה ומכשיר טלוויזיה, ושכללי הפעולה שלו כוללים גם התייחסויות לתמונות שהוא רואה דרכם. כך, הוא יוכל לקשור את המונח הסיני הרלוונטי לתמונות של המבורגר, שבעצמן קשורות להמבורגר. במצב עניינים זה, דומה שקשה יותר לקבל את הטענה שכשהאדם בחדר הסיני (המשוכלל) משתמש במונח "###@#\$%^" (המבורגר בסינית, כזכור), מונח זה חסר משמעות, ואינו נסב על אודות המבורגר. קשה יהיה לקבל את הטענה שהאדם אינו מבין את המונח הזה. לפי גישה זו, כשנכלול

בתמונה הפונקציונליסטית את איברי החושים השונים שלנו (האנלוגיים, כמובן, לטלוויזיה בחדר הסיני המשוכלל, והמספקים לנו קשרים לאובייקטים חיצוניים) תמונה זו תוכל להסביר מכוונות לאובייקטים חיצוניים, משמעות והבנה.

תשובה זו לטיעון החדר הסיני אינה מייחסת חשיבה למחשבים. האדם בחדר הסיני פועל כמחשב: הוא מקבל קלטים שהם בעבורו סמלים צורניים, עובר בין מצבים שונים כשכללי המעברים הם כללים צורניים, ולבסוף מפיק פלטים. התשובה הנדונה איננה מתנגדת לטענה שפעולה כמחשב בהתאם לעקרונות אלו איננה כרוכה בחשיבה. אולם לפי תשובה זו, כשנשפץ מחשב הפועל באופן זה בגוף של **רובוט**, פעולתו עשויה להיות כרוכה בחשיבה. האדם בחדר הסיני הוא המחשב (או המוח) בגופו של הרובוט. המצלמה ומקלט הטלוויזיה שהוזכרו הם כעיניים, המאפשרות לקלוט גירויים חיצוניים ולהעבירם למחשב. הם מאפשרים קשר לאובייקטים חיצוניים, כמו המבורגרים. הרעיון הגלום בתגובה הוא שמה שמעניק מכוונות ומשמעות למצבינו הנפשיים הוא לא רק התפקוד הפנימי שלהם, אלא קשר סיבתי בינם לבין אובייקטים בעולם. אם המילים שבשימושי לא רק מתפקדות באופנים פנימיים שונים, אלא קשורות בקשרים סיבתיים מסוימים להמבורגרים, ואני יכול להחיל אותן על המבורגרים, נראה שהן עשויות להיות מכוונות כלפי המבורגרים, להיות נסבות על אודות המבורגרים.

תשובה בכיוון זה לטיעון החדר הסיני הוצעה על ידי פילוסופים אחדים. זוהי "תשובת הרובוט" ("the Robot Reply") שבה דן סרל במאמר.<sup>77</sup> כמובן, רובוטים קשורים לאובייקטים בעולם לא רק באמצעות מצלמות המתפקדות כאיברי חושים, אלא גם באמצעות פעולות על אובייקטים בעולם (הם יכולים להכין המבורגרים, להגיש אותם, וכדומה). חלק מן הפילוסופים הדוגלים בתשובת הרובוט סבורים שגם קשרים כאלו לאובייקטים בעולם – קשרים בצד של הפלט – משחקים תפקיד בהקניית מכוונות ומשמעות. לפיהם, התמונה הפונקציונליסטית תוכל להסביר מכוונות לאובייקטים חיצוניים, משמעות והבנה, אם נכלול בה הן את איברי החושים שלנו (בצד של הקלט) והן את פעולותינו על אובייקטים בעולם (בצד של הפלט).

אם קו מחשבה זה הוא בכיוון הנכון, אפשר לומר שטיעונו של סרל לא הצליח לחתור תחת גרסאות לא מופשטות של פונקציונליזם. נוכל להעריך בצורה טובה יותר את קו המחשבה הזה, לאחר שנטפל (ביחידה 8) באופן מקיף בנושא של מכוונות וכיצד היא מושגת, אולם כבר עתה אפשר לראות שטיעון החדר הסיני אינו חף מקשיים.

<sup>77</sup> בין הפילוסופים שהציעו תשובות ברוח זו לטיעון החדר הסיני, ראו:

M. Boden, *Computer Models of Mind*, Cambridge: Cambridge University Press, 1988.

J. Fodor, "Yin and Yang in the Chinese Room", in D. Rosenthal (ed.), *The Nature of Mind*, Oxford and New York: Oxford University Press, 1991.

סרל הגיב על תשובת הרובוט גם במסגרת תגובותיו לתשובות שונות לטיעונו בגיליון הנזכר של כתב-העת *Behavioral and Brain Sciences*.

מובן שלפי קו מחשבה זה, העובדה שרובוט (או האדם בחדר הסיני המשוכלל) עשוי לעבור את מבחן טיורינג אינה בעייתית – על כל פנים, טיעון החדר הסיני אינו מראה שהיא בעייתית – שכן מצביו של רובוט (או של האדם בחדר הסיני המשוכלל) עשויים להיות בעלי מכוונות. מה באשר לטענה שהאדם בחדר הסיני המקורי עובר את מבחן טיורינג? ייתכן שמי שמצדדים בתשובת הרובוט לטיעון של סרל יכולים לומר שהאדם בחדר הסיני המקורי **אינו עובר את המבחן**: הוא אינו עובר את המבחן ההולם, מפני שהמבחן ההולם צריך לכלול שאלות הכרוכות בהצגת דברים בעולם כמו המבורגר, או תמונות שלהם (שאלות כמו: האם מה שאתה רואה כאן זה "##@#\$^"?), ואלו נמצאים מחוץ לתחום עבורו.

כדאי גם להעיר, שייתכן שמטרתו של מבחן טיורינג הייתה להוות מבחן רק **לארגון הפונקציונלי** המכונן חשיבה, ולא לכל היבטיה של חשיבה. לפי פרשנות זו, השאלה שעליה בא המבחן להשיב היא, איך אפשר לזהות האם למערכת מסוימת יש **ארגון פונקציונלי** הנדרש לחשיבה? מבחן טיורינג המפורש באופן זה איננו מבוסס על ההנחה שארגון פונקציונלי כלשהו מספיק כדי לכונן חשיבה – אין הוא שולל את הרעיון שלחשיבה נדרשים תנאים נוספים. אבל לפיו, כדי שמערכת תהיה מערכת חושבת עליה להיות, בין השאר, בעלת ארגון פונקציונלי ברמת מורכבות גבוהה, ומבחן טיורינג הוא תשובה לשאלה: מהו המבחן לכך שארגונה הפונקציונלי של מערכת מסוימת מספיק כדי לכונן חשיבה, בהנחה שאם נדרשים תנאים נוספים לחשיבה, מערכת זו מקיימת אותם? בהנחה שמבחן טיורינג אמנם בא להשיב לשאלה זו, ברור שטיעון החדר הסיני אינו מפריך אותו, כי מבחן זה אינו מניח שארגון פונקציונלי כלשהו בעצמו מכונן חשיבה, וזוהי ההנחה שמנסה טיעון החדר הסיני להפריך. למבחן כזה – מבחן לארגון הפונקציונלי הדרוש לחשיבה – יש חשיבות תיאורטית בין אם הוא תואם את הרעיון של טיורינג ובין אם לאו.

## 9. פונקציונליזם ו-qualia

המאפיינים של חיינו הנפשיים שלפחות לכאורה מאיימים באופן המשמעותי ביותר על ההשקפה הפונקציונליסטית הם התכונות ההתנסותיות של המצבים הנפשיים – ה-qualia. התכונות הללו הן האופנים שבהם דברים נראים, נשמעים, או מורגשים, או הדרכים שבהם כאבים ותחושות גופניות שונות נחווים: האיכות התחושתית האופיינית של מתיקות המלווה טעימת דברים מתוקים, האיכות הבלתי נעימה הכרוכה בכאבים, תחושת ה"אודם" הכרוכה בראיית דברים אדומים, וכדומה. כפי שאמרנו ביחידות קודמות, תכונות אלו מציבות אתגר קשה בפני תורה פילוסופית בדבר היסוד הנפשי, ואתגר קשה במיוחד בפני תורת מטריאליסטיות למיניהן.

בהקשר של ההשקפה הפונקציונליסטית, אתגר זה לובש צורה מיוחדת. כאמור, ההשקפה הפונקציונליסטית מאפיינת ומסווגת מצבים נפשיים על פי קשריהם עם מצבים אחרים. כלומר, היא מאפיינת ומסווגת אותם על פי **תכונות יחסותיות** שלהם. תכונות נפשיות, על פיה, הן תכונות יחסותיות: מה שעושה מצב לבעל תכונות המכאיבות, למשל, זו העובדה שמצב זה קשור עם מצבים נפשיים אחרים, למשל, שהוא נגרם על ידי נזק גופני, גורם רצייה ליטול משכך כאבים, וכדומה. אולם השקפה טבעית באשר ל-qualia היא שהן **תכונות פנימיות**, או **אינטרינסיטיות** – **תכונות בלתי יחסותיות** – של מצבים נפשיים. על פי השקפה טבעית זו, המכאיבות, למשל, היא טבעו הפנימי של הכאב. אמנם כאבים כרוכים, בין היתר, ברציות שונות, אולם המכאיבות האופיינית שלהם, טבעם הבלתי נעים (לעתים נורא), חורגת מהקשר שלהם לרציות ולמצבים אחרים ואינה תלויה בו. טבע זה – "איך זה להרגיש כאב" – כך נראה, הוא פנימי, או אינטרינסי, לכאב, הוא מאפיין אותו בפני עצמו. אם אמנם qualia הן תכונות בלתי יחסותיות של מצבים נפשיים, ההשקפה הפונקציונליסטית מאוימת על ידי קושי משמעותי, משום שעל פיה התכונות הנפשיות הן תכונות יחסותיות, תכונות המכוננות על ידי קשריהן למצבים אחרים; נראה ש-qualia חומקות מן הרשת הפונקציונליסטית – שאין להן מקום במסגרת תמונת עולם פונקציונליסטית.

**הבעיה שמציבות תכונות התנסותיות (qualia) בפני ההשקפה הפונקציונליסטית:** לפי השקפה זו התכונות הנפשיות הן תכונות יחסותיות, אולם נראה שתכונות התנסותיות הן תכונות פנימיות (בלתי יחסותיות) של מצבים נפשיים.

באמור לעיל אין כדי להוות טיעון מספק כנגד הפונקציונליזם, משום שאין לנו ערובה שאמנם הדעה ה"טבעית" על פיה תכונות התנסותיות הן תכונות פנימיות ולא יחסותיות היא נכונה. ואמנם, פילוסופים שונים ניסו להמחיש את הבעיה שמציבות התכונות ההתנסותיות בפני הפונקציונליסטים באמצעות טיעונים ספציפיים. ההיגיון שביסוד מרבית הטיעונים הללו הוא

דומה: הפונקציונליסטים מחויבים לכך שמערכות זהות מבחינה פונקציונלית הן זהות מבחינה נפשית, ובכלל זה מבחינה התנסותית (כלומר, הן כרוכות באותן תכונות התנסותיות). אם, לפיכך, ניתן להצביע על האפשרות של קיומן של מערכות הזהות זו לזו מבחינה פונקציונלית אולם נבדלות זו מזו מבחינה התנסותית, פירוש הדבר שבמסגרת ההשקפה הפונקציונליסטית אין מקום לתכונות התנסותיות, וכל מי שמקבלים את קיומן של תכונות אלו ואת מהותיותן לחיי הנפש שלנו מחויבים לדחות השקפה זאת.

### שאלה

אילו צורות עשויה ללבוש האפשרות של מערכות הזהות זו לזו מבחינה פונקציונלית, אולם נבדלות זו מזו מבחינה התנסותית? במלים אחרות, באילו מובנים יכולות מערכות להיבדל זו מזו מבחינה התנסותית?

מערכות עשויות להיבדל זו מזו מבחינה התנסותית בשני אופנים. ראשית, ייתכן שתהייה להן התנסויות שונות, התנסויות בעלות qualia שונות. שנית, ייתכן שלאחת תהייה התנסויות ולשנייה לא תהייה התנסויות כלל. הטיעון המתבסס על האפשרות הראשונה מכונה "טיעון ה-qualia המהופכות" ("the inverted qualia argument"), או לעתים (מסיבה שתובהר להלן) טיעון "הספקטרום המהופך" ("the inverted spectrum argument"); הטיעון המתבסס על האפשרות השנייה מכונה "טיעון ה-qualia הנעדרות" ("the absent qualia argument").

## 9.1 טיעון הספקטרום המהופך

הפילוסוף בן המאה ה-17 ג'ון לוק (John Locke) הצביע על האפשרות שיצורים נפשיים שונים נבדלים זה מזה בספקטרום התנסויות הצבעים שלהם באופן כדוגמת הבא: כאשר אחד מהם מתבונן במלפפונים בשלים, הם נראים לו ירוקים (יש לו qualia של ירוק), וכשהוא מתבונן בעגבניות בשלות הן נראות לו אדומות (יש לו qualia של אדום); כאשר רעהו מתבונן באותם מלפפונים הם נראים לו אדומים (יש לו qualia של אדום), וכשהוא מתבונן באותן עגבניות הן נראות לו ירוקות (יש לו qualia של ירוק). פילוסופים בני זמננו השתמשו ברעיון זה של ספקטרום התנסויות הצבעים המהופך כדי לטעון כנגד ההשקפה הפונקציונליסטית. לטענתם, אפשרי מקרה של היפוך כזה בין שני יצורים נפשיים גם אם מתקיימת ביניהן זהות פונקציונלית מוחלטת.<sup>78</sup> בין יצורים בעלי אותו ארגון פונקציונלי שלהם עשוי להתקיים היפוך ספקטרום כזה אם הם נבדלים זה מזה במבנה הפיזיקלי. הזהות הפונקציונלית ביניהם תתבטא בכך ששניהם יחשבו "הו, הנה מלפפון ירוק", לשניהם תהיה נטייה לאכול את אותם מלפפונים (שכן שניהם חושבים שמלפפונים שנראים "כך" הם בשלים), וכדומה. עובדות כאלו

<sup>78</sup> ראו

N. Block, and J. Fodor, "What Psychological States are Not", *The Philosophical Review* 81, 1972, pp. 159-81.



מתיישבות, כך נראה, עם האפשרות שה-qualia הרלוונטיות שלהם שונות – שהאופנים בהם הדברים נראים להם הם שונים. כלומר, זהות פונקציונלית מתיישבת עם הבדל התנסותי. אם כך, ההשקפה הפונקציונליסטית "מותירה בחוץ" היבט חשוב של התנסויות, ההיבט ההתנסותי. אין היא אפוא השקפה הולמת באשר לטבען של התנסויות, ומכאן שאין היא השקפה ממצה על אודות היסוד הנפשי.

פילוסופים אחדים חושבים שדי להצביע על **אפשרות לוגית** של היפוך כזה (או של qualia נעדרות) כדי להפריך את ההשקפה הפונקציונליסטית על שתי גרסאותיה, העממית והאמפירית, ואילו אחרים חושבים שההצבעה על אפשרות לוגית כזו מספיקה כדי להפריך את הפונקציונליזם העממי, אך כדי להפריך את הפונקציונליזם האמפירי יש לשכנע שזוהי לא רק אפשרות לוגית אלא גם **אפשרות נומולוגית**, דהיינו אפשרות המתיישבת עם חוקי הטבע השוררים בעולמנו. לדעתנו, די בהצבעה על אפשרות לוגית בכדי להפריך את שתי הגרסאות. זאת משום שפי שאמרנו, כל הגרסאות של הפונקציונליזם מחויבות לכך שהתממשותן של תכונות פונקציונליות מסוימת מחייבת מבחינה לוגית את התממשותן של התכונות הנפשיות.

אנלוגיה שיש בה כדי להבהיר את האפשרות של ספקטרום צבעים מהופך היא זו: חשבו על רשתות זהות של קשרים פונקציונליים, אשר נבדלות זו מזו מבחינה פיזיקלית. כמו כן, התרחשותו של כל מצב פנימי מלווה בהארה על ידי נורה צבעונית. צבעי הנורות בשתי הרשתות מהופכים: נורה ירוקה אצל האחת מקבילה לאדומה אצל השנייה ולהפך, נורה כחולה מקבילה לנורה צהובה ולהפך, וכדומה.

**הרעיון שביסוד טיעון הספקטרום המהופך: אפשרות זהות פונקציונלית חרף שונות התנסותית.**

## שאלה

כשם שמועלית האפשרות של ספקטרום צבעים מהופך, אפשר להעלות את האפשרות של היפוך בין כאבים להנאות. חשבו מה פשרה של אפשרות כזו, ומדוע, לפחות לכאורה, היא פחות סבירה מן האפשרות של היפוך צבעים.

היפוך בין כאבים להנאות משמעו שלכאבים יהיה את התפקיד הפונקציונלי שיש במצב רגיל להנאות, ולהנאות יהיה את התפקיד הפונקציונלי שיש במצב רגיל לכאבים. אחד הביטויים של היפוך כזה, שלפחות על פני הדברים אינו סביר, הוא שאם מגע בחפץ מסוים יגרום ליצור מהופך כזה כאב, הוא ינסה לגעת בו עוד, ואם מגע בחפץ מסוים יגרום ליצור הנאה, הוא יבכר להתרחק ממנו.

פילוסופים אחדים, הנוטים להשקפה הפונקציונליסטית, הסכימו שטעון זה מראה שההשקפה הפונקציונליסטית איננה ההשקפה ממצה באשר ליסוד הנפשי, ושהיא זקוקה להשלמה כלשהי.<sup>79</sup> פילוסופים אחרים ניסו להצביע על פגמים בטעון. נזכיר בקצרה קו מחשבה אחד המנסה לחתור תחת טעון הספקטרום המהופך. לפי קו מחשבה זה, ספקטרום מהופך מן הסוג שתיארנו – ספקטרום מהופך בין שני יצורים חרף זהות פונקציונלית ביניהם – איננו אפשרי. בין המצבים הנפשיים האחרים הכרוכים באפיונה של תחושת צבע נמנים **שיפוטי שונות ודומות** בין צבעים, שיפוטים כמו: הצבע הזה דומה לצבע הזה; צהוב שונה מאדום יותר מאשר מכתום, וכדומה. אם כך, הרי אם התפקיד הפונקציונלי של התנסות האדום של יצור אחד זהה לתפקיד הפונקציונלי של התנסות הירוק של יצור אחר, על שיפוטי הדומות והשונות שהם מחילים באשר לצבעים להיות זהים. בין היתר, אם האחד שופט בהתבוננו בעגבנייה שהצבע שלה דומה מאוד לכתום, גם רעהו המתבונן בה אמור לשפוט שהצבע שלה דומה מאוד לכתום. אבל רעהו, כאמור, רואה את העגבנייה כירוקה, וירוק אינו נראה לו דומה לכתום, ולכן, בניגוד אליו, הוא לא ישפוט שהצבע של העגבנייה דומה לכתום. במילים אחרות, בין שני היצורים הללו, שהנחנו שה-qualia שלהם מהופכות, חייב להתקיים הבדל פונקציונלי: כלומר, **ההבדל הנדון ב-qualia, ההבדל ההתנסותי, יחייב הבדל פונקציונלי**, שיתבטא בהיות המצבים ההתנסותיים השונים קשורים עם האמנות שונות. אם כך, דומה שטעון הספקטרום המהופך אינו מצליח לחתור תחת ההשקפה הפונקציונלית, כי בניגוד ליומרתו הוא לא מצליח להציג מקרה של qualia שונות שיש להן **אותו תפקיד פונקציונלי**.

הדרך של בעלי טעון הספקטרום המהופך להתמודד עם התנגדות זו לטיעונם היא להתייחס לאפשרות של היפוך ספקטרום מקיף יותר, ולא היפוך מקומי בלבד (היפוך בין שתי התנסויות צבע). כלומר, יש להניח גם שהמצב הפונקציונלי שמימש **התנסות כתום** אצל האחד מממש התנסות צבע אחרת אצל האחר, וכך גם לגבי צבעים אחרים. למעשה, ניתן לראות שכדי לחמוק מן ההתנגדות הנדונה על בעלי טעון הספקטרום המהופך להניח שכל מצביהם הפונקציונליים הזהים המממשים התנסויות צבע מממשים התנסויות צבע שונות; ועליהם להניח שחילוף גלובלי כזה יהיה שיטתי באופן שיבטיח זהויות בשיפוטי הדומות והשונות של צבעים. כשהיצור האחד ישפוט – בהתאם להתנסויותיו – שצבע מסוים דומה מאוד לצבע אחר, ישפוט גם רעהו שהצבע הראשון דומה מאוד לשני, חרף העובדה שההתנסויות שונות. האם חילוף qualia כזה הוא אמנם אפשרי? נסתפק כאן בהעלאת שאלה קשה זו, ונשאירה כחומר למחשבה.

<sup>79</sup> ראו, למשל:

S. Shoemaker, "Functionalism and Qualia", *Philosophical Studies* 27, 1974, pp. 291-315.

## 9.2 טיעון ה-qualia הנעדרות והמוח הגדול של סין

גם טיעון זה מצביע על מערכות שנטען שהן זהות זו לזו מבחינה פונקציונלית אולם נבדלות זו מזו מבחינה התנסותית. אולם הפעם ההבדל ההתנסותי אינו בין מערכות עם qualia שונות, אלא בין מערכת עם qualia לבין מערכת הזוהה לה פונקציונלית אולם חסרה qualia לחלוטין. אם תיתכנה שתי מערכות כאלו, הרי שהאפיון הפונקציונלי שאמור לכונן qualia אינו מכונן אותן, שכן ייתכן שהוא יתממש ללא qualia. כך qualia אינן תכונות פונקציונליות, ומכאן שההשקפה הפונקציונליסטית אינה השקפה ממצה על אודות היסוד הנפשי. נקרא את מילותיו הבאות של הפילוסוף נד בלוק:

עד כמה שאנו יודעים, זוהי אפשרות נומולוגית ששני מצבים פסיכולוגיים יהיו זהים מבחינה פונקציונלית (כלומר, קשורים באופנים זהים עם קלטים, פלטים, ומצבים עוקבים), אף אם רק לאחד מהם יש תוכן התנסותי. במקרה זה, תיאוריות פונקציונליסטיות תדרושנה מאתנו לומר שאורגניזם עשוי להיות במצב של כאב אפילו אם הוא אינו חש דבר, ותוצאה זו נראית בלתי קבילה לחלוטין.<sup>80</sup>

האמנם מצב כזה אפשרי? על פי רוב בעלי הטיעון אינם מסתפקים בטענה סתמית שהוא אפשרי. הם מתארים מערכות ספציפיות שלטענתם הן זהות מבחינה פונקציונלית למערכת הפסיכולוגית האנושית, אולם בניגוד למערכת הפסיכולוגית האנושית, למערכות אלו אין התנסויות – אין להן מצבים עם תכונות התנסותיות. הדוגמה הידועה ביותר של מערכת כזאת היא דוגמת "המוח הגדול של סין" של נד בלוק. נדמיין אדם מסוים, שתאי העצב של הקלטים (דהיינו, התפיסות החושיות) ושל הפלטים (דהיינו, התנועות) שלו אינם מקושרים למוח. איך אדם זה פועל? ובכן, כל אזרחי סין מפעילים אותו, והם למעשה מהווים, ביחד, את מוחו, כאשר כל אזרח מתפקד כנירון ב"מוח" זה (בלוק בחר באזרחי סין בשל מספרם העצום). אזרחי סין מצוידים במכשירי רדיו דו-כיווניים, המקשרים אותם גם זה לזה, וגם לתאי העצב הללו של הקלטים והפלטים, באמצעות גלי רדיו. כשהאדם תופס משהו בחושיו (נניח, כשהוא רואה משהו) תאי העצב הרלוונטיים מעבירים אותות לא אל תאי מוח כלשהו (לאדם אין מוח), אלא אל חלק מאזרחי סין. אזרחי סין אלו מקבלים אותות נוספים מאזרחי סין אחרים, ומעבירים אותות לאזרחי סין נוספים ולתאי הפלט של האדם, כלומר לתאי העצב האחראים לתנועותיו. למעשה, כך ממומש ארגון פונקציונלי מסוים, שהקלט והפלט שלו הם הקלט והפלט של מערכת פסיכולוגית אנושית, אולם המצבים "הפנימיים" הם כמובן שונים. האותות שאנשי סין משדרים ונמעניהם (כלומר, למי הם משדרים אותם) נקבעים על פי הוראות שהם מקבלים. ניתן להניח שההוראות שהם מקבלים הן כאלו, שדגמי הגירוי-תגובה של מערכת זו – של "האדם בעל המוח של סין" זהים לדגמי הגירוי והתגובה של אדם נורמלי כלשהו; כלומר, כשהאדם הנורמלי פועל באופן מסוים בתגובה לתפיסה חושית מסוימת, גם בעל המוח של סין יפעל באותו אופן כשהוא קולט את אותה תפיסה חושית. אולם יותר מכך,

<sup>80</sup> ראו:

N. Block, "Troubles with Functionalism", in N. Block (ed.), *Readings in Philosophy of Psychology: Volume 1*. Cambridge (Mass.): Harvard University Press, 1980.

אפשר לדמיין שההוראות שהם מקבלים הן כאלו, שהארגון הפונקציונלי של בעל המוח של סין זהה לחלוטין לארגון הפונקציונלי של האדם הנורמלי, כלומר שמערכת הקשרים הסיבתיים של בעל המוח של סין זהה לגמרי במבנה שלה למערכת הקשרים הפסיכולוגית של האדם הנורמלי. אולם לאדם הנורמלי יש התנסויות ו-qualia: יש לו כאבים, תחושות ויזואליות, וכדומה. כלומר, **המערכת של בעל המוח של סין זהה מבחינה פונקציונלית למערכת פסיכולוגית אנושית בעלת qualia**. מן ההשקפה הפונקציונליסטית מתחייב אפוא שלמערכת זו יש qualia. אולם, טוען בלוק, זהו אבסורד. לכל אחד מאזרחי סין יש התנסויות ו-qualia (סביר להניח), אולם לכל המערכת המתוארת אין התנסויות נוספות מעבר להתנסויותיהם של האזרחים. כך, שוב אנו מגיעים למסקנה שבתמונת העולם הפונקציונליסטית אין מקום ל-qualia, ומכאן שאין היא השקפה סבירה באשר ליסוד הנפשי.<sup>81</sup>

### שאלה

התוכלו להציג את המבנה של טיעון המוח הגדול של סין?

### טיעון המוח הגדול של סין:

- (1) למוח הגדול של סין ארגון פונקציונלי זהה לזה של אדם בעל התנסויות.
  - (2) מההשקפה הפונקציונליסטית מתחייב שלמערכת בעלת ארגון פונקציונלי הזהה לזה של אדם בעל התנסויות תהיינה התנסויות.
  - (3) (נובע מ-1 ו-2) מההשקפה הפונקציונליסטית מתחייב שלמוח הגדול של סין תהיינה התנסויות.
  - (4) למוח הגדול של סין אין התנסויות.
- לפיכך (נובע מ-3 ו-4),  
ההשקפה הפונקציונליסטית שקרית.

השאלה העיקרית באשר לטיעון זה נוגעת לאמיתותה של טענה 4. מה הבסיס לטענה זו? לרבים נדמה שהיא ברורה מאליה. האם מישהו ירצה לטעון ברצינות שלמערכת המשונה הזאת של קשרי רדיו עשוי לכאוב, למשל? מנגד, הפונקציונליסטים עשויים לטעון כי לאור מורכבותו העצומה של הארגון הפונקציונלי של המערכת הפסיכולוגית האנושית, אין זה ברור שאמנם ייתכן שמערכת בעלת ארגון פונקציונלי כה מורכב (מערכת הזוהה מבחינה פונקציונלית למערכת הפסיכולוגית האנושית) תחסר qualia, וכי הטענה שהיא חסרה qualia נשענת לפיכך על אינטואיציה רעועה.<sup>82</sup> אנו נהיה בעמדה טובה יותר לבחון את הטענות הללו כשנדון באופן נרחב ומעמיק יותר בנושא של מודעות ו-qualia ביחידה 9.

Block, Ibid.

81

<sup>82</sup> טענה ברוח זו העלה הפילוסוף וויליאם לייקאן. ראו:

W. Lycan, *Consciousness*. Cambridge (Mass.): MIT Press, 1987, p. 24.

שימו לב שטיעוני הספקטרום המהופך וה-qualia הנעדרות שהצגנו הם טיעונים אנטי-פונקציונליסטים ולא אנטי-מטריאליסטים. על פיהם יש אי-התאמה בין התכונות ההתנסותיות של מצבים נפשיים לבין תכונותיהם הפונקציונליות; אין הם מתייחסים לתכונותיהם הפיזיקליות של מצבים נפשיים. כדי לטעון לאי-התאמה מן הסוג האחרון נדרשים טיעונים אחרים. טיעון אפשרות ההשגה וטיעון הידע שהוצגו ביחידה 2, וטיעונים אחדים שהוצגו ביחידה 4-5, הם כאלו.

לפחות על-פני הדברים, qualia מציבות אתגר קשה להשקפת הפונקציונליזם העממי יותר מאשר להשקפת הפונקציונליזם האמפירי. גם אם ניתן לספק ל-qualia הסבר פונקציונלי שלם – גם אם ניתן להראות שכל המאפיינים שלהן נקבעים על ידי קשרים פונקציונליים מסוימים של מצבים נפשיים, עדיין נותר קושי לעמדת הפונקציונליזם העממי. שכן הפונקציונליסטים העממיים מחויבים לטענה חזקה יותר, לטענה שהמושגים של qualia שונות הם מושגים פונקציונליים, וכך, מן הסתם, שהם מושגים יחסותיים. לפי השקפה טבעית ביותר, המושג של כאב, למשל, גם אם הוא כרוך באפיון פונקציונלי (מצב הנגרם על ידי פגיעה ברקמות הגוף, גורם לרצייה להיפטר ממנו, וכדומה), אינו מתמצה על ידי אפיון כזה. המושג של כאב נראה באופן טבעי כמצב הכרוך באופן מהותי גם (ואולי בעיקר) בטבע פנימי בלתי נעים (לעתים נורא), טבע פנימי שאינו כפוף לאפיון יחסותי כלשהו. הפונקציונליסטים האמפיריים אינם מכחישים זאת. הם יכולים לקבל עמדה זו באשר למושג הכאב, אולם לטעון שמושג זה אינו משקף נכונה את טבעם של כאבים – שמושג הכאב אומנם אינו יחסותי, אולם מכאיבות היא תכונה יחסותית, תכונה המכוננת על ידי יחסים בין מצבים נפשיים שונים ומצבי קלט ופלט למיניהם. פונקציונליסטים עממיים אינם יכולים לנקוט בגישה זו, משום שההנחה שמושג הכאב שלנו שגוי באופן רדיקלי אינה מתיישבת עם גישתם האנליטית, כפי שראינו. עליהם לטעון אפוא שלמרות מראית-העין, מושג המכאיבות שלנו גם הוא מושג יחסותי, אך אין זה קל לשכנע באמתותה של טענה זו. ביחידה 9 נשוב ונתייחס לשאלת היחסותיות או הפנימיות של qualia ושל מושגים התנסותיים.

## 10. מילות סיכום

אין ספק שהשקפה הפונקציונליסטית היא אחת ההשערות המרתקות שהוצעו בפילוסופיה של הנפש, וסביר מאוד שחלק מחידושיה מתארים נכונה את היסוד הנפשי. חשבו על אחדים מחידושים אלו: השקפה זו מציעה הסבר אלגנטי לתופעה של ריבוי מימושים; היא מתווה מודל מעניין המסביר את יחסי הגוף והנפש – המודל של יחסי החומרה והתוכנה; היא מתגברת על קשיים של קודמותיה (התורה של זהות-סוג פיזיקליסטית והביהיוריזם הפילוסופי); היא מראה כיצד ניתן לדחות את הטענה שתכונות נפשיות הן תכונות פיזיקליות מבלי להיקלע לדואליזם; היא מצליחה לאפיין מצבים נפשיים על ידי התייחסות (בין היתר) למצבים נפשיים ויחד עם זאת לעשות רדוקציה למצבים נפשיים, ועוד. הפילוסופיה של הנפש שלאחר פיתוחה של ההשקפה הפונקציונליסטית אינה מה שהייתה לפני פיתוחה של השקפה זו.

במהלך דיונו בפונקציונליזם הצגנו שיקולים וטיעונים אחדים בזכות השקפה זו. הצגנו את השיקול בזכות פונקציונליזם עממי המתבסס על טענתו של ארמסטרונג לקשר מושגי בין מונחים נפשיים לבין תפקידים סיבתיים בתוספת הטענה שעל פיה התפקידים הסיבתיים של המצבים הנפשיים חייבים לכלול התייחסויות לא רק לגירויים תפיסתיים ולתגובות התנהגותיות אלא גם למצבים נפשיים אחרים. הצגנו את השיקול שעל פיו הרעיון הפונקציונליסטי בדבר הנפש כתוכנת הגוף מסביר כיצד מבנים חומריים שאין ביניהם כל דומות פיזיקלית משמעותית עשויים להיות שקולים בכוחותיהם הסיבתיים הרלוונטיים למצביהם הנפשיים: אם המבנים הפיזיקליים השונים המממשים את המצבים הנפשיים הם מימושים של אותה תוכנה, הם שקולים בכוחותיהם הסיבתיים בדיוק כשם שמחשבים שוני חומרה המממשים את אותה תוכנה שקולים בכוחותיהם הסיבתיים. והצגנו את טיעונו של פטנאם בזכות סבירותה של ההשקפה הפונקציונליסטית כהשערה אמפירית. אף אחד מאלו איננו שיקול מכריע בזכות ההשקפה הפונקציונליסטית. סבירותו של השיקול המושגי תלויה, בין היתר, בסבירותה של הטענה שהמושגים של qualia שונות הם מושגים יחסותיים ולא פנימיים, אך השכנוע באמיתותה של טענה זו מהווה אתגר לא פשוט. השיקול השני מסתמך על הכוח ההסברי האמור לאפיין את ההשקפה הפונקציונליסטית, אולם היותה של השקפה בעלת כוח הסברי אינו שיקול מכריע בזכותה.<sup>83</sup> הסבר טוב של תופעה איננו בהכרח אמתי, ולו משום שייכתנו הסברים אחרים; ואף ההסבר הטוב ביותר הקיים איננו בהכרח הסבר אמתי, ולו משום שייכתן שיימצאו בעתיד הסברים אחרים. טיעונו של פטנאם מראה שלהשקפה המזהה תכונות נפשיות עם תכונות פונקציונליות ושעדיפות על-פני השקפות המזהות תכונות נפשיות עם תכונות פיזיקליות או עם תכונות התנהגותיות, אולם אין בכוחו של טיעון זה לשלול את האפשרות שתכונות נפשיות זהות לתכונות אחרות. ייתכן גם שיימצא כי ניתן לזהות רק חלק מן התכונות הנפשיות עם תכונות פונקציונליות.

<sup>83</sup> ראו את דיונו בטיעון מן ההסבר הטוב ביותר בזכות ההשקפה המטריאליסטית ביחידה 4-5.

הצגנו ביקורות חשובות כנגד ההשקפה הפונקציונליסטית. חלק מהביקורות כנגד השקפה זו נראות חזקות יותר כנגד הפונקציונליזם האמפירי (למשל, ביקורת זהות התוכנה), אחרות – כנגד הפונקציונליזם העממי (ביקורות ה-qualia). עניינן של שתיים מהביקורות שבהן דנו הן בשניים מהמאפיינים הבולטים של חיי הנפש – היותם בעלי מכוונות (אינטנציונליות) או תוכן, והיותם בעלי תכונות התנסותיות (qualia). לכל אחד מהמאפיינים הללו נידרש בהמשך. יחידה 8 תוקדש לנושא של תוכן נפשי, ויחידה 9 תוקדש למודעות, כשעיקר הדגש בה יושם על מודעות התנסותית. ביחידות אלו גם נציג את גישת התפקיד הפונקציונלי כלפי תוכן נפשי, ואת הגישה הפונקציונליסטית-קוגניטיביסטית כלפי מודעות התנסותית. גישות אלו, המטפלות במאפיינים חשובים ביותר של החיים הנפשיים, מהוות חלק מהשקפה פונקציונליסטית מקיפה. אם גישות אלו מוצלחות, משמעות הדבר היא לא רק שההשקפה הפונקציונליסטית נחלצת מביקורות, אלא שהיא מקבלת חיזוקים פוזיטיביים חשובים. אם נראה שהיא לא רק מתיישבת עם קיומם של מאפיינים בולטים של המצבים הנפשיים כמו מכוונות או מודעות התנסותית, אלא גם מצליחה להסביר אותם, הסבירות של אמיתותה תתחזק מאוד, חרף העובדה שהטיעונים בזכותה אינם מכריעים. אין ספק שמעלותיה של ההשקפה הפונקציונליסטית, שאותן מנינו לעיל, יגבירו סבירות זו.





## נספח

### נפש, מוח, ותוכנות מחשב\*

מאת ג'ון ר' סרל

איזו משמעות פסיכולוגית ופילוסופית עלינו לייחס לעבודה הנעשית כיום בסימולציות מחשב של כישורים קוגניטיביים אנושיים? כשאני עונה על שאלה זו, אני מוצא שמועיל להבחין בין מה שאני מכנה התזה של ב"מ (בינה מלאכותית) "חזקה" לבין התזה של ב"מ "חלשה" או "זהירה". לפי התזה של ב"מ "חלשה", הערך העיקרי של המחשב בחקר הנפש הוא בהיותו כלי עבודה רב עוצמה. כך למשל, הוא מאפשר לנו לנסח היפותזות ולבחון אותן בדרך חמורה ומדויקת יותר. ואילו לפי התזה של ב"מ חזקה, המחשב אינו רק כלי עבודה לחקר הנפש; המחשב המתוכנת באופן הנכון הוא נפש, במובן זה שאפשר לומר על מחשבים שבהם התוכנות הנכונות, שהם **מבינים**, פשוטו כמשמעו, ושיש להם מצבים קוגניטיביים נוספים. בב"מ חזקה, מאחר שלמחשב המתוכנת יש מצבים קוגניטיביים, אזי תוכנותיו אינן רק כלי עבודה המאפשרים לנו לבחון הסברים פסיכולוגיים; התוכנות עצמן הן ההסברים.

אין לי כל התנגדות לטענות של ב"מ חלשה, לפחות בכל הנוגע למאמר זה. הדיון הנוכחי יתמקד בטענות שהגדרתי כב"מ חזקה, ובמיוחד בטענה שלמחשב המתוכנת כראוי יש מצבים קוגניטיביים במשמעות הרגילה של הביטוי, וכך תוכנותיו מסבירות הכרה (cognition) אנושית. מעתה ואילך, כשאתייחס לב"מ, כוונתי תהיה לגרסה ה"חזקה", כפי שהיא מבוטאת על-ידי שתי הטענות הללו.

בחיבור זה אתייחס למחקריו של רוג'ר שנק ועמיתיו באוניברסיטת ייל (Schank and Abelson, 1977), משום שאני מכיר אותם טוב יותר ממחקרים דומים אחרים, ומאחר שהם מדגימים היטב את סוג המחקרים שברצוני לבחון. אולם כל האמור להלן אינו תלוי בפרטים באשר לתוכנות של שנק. אותם טיעונים אפשר להחיל על תוכנת SHRDLU של טרי וינוגרד (Winograd, 1973), על תוכנת ELIZA של ווייצנבאום (Weizenbaum, 1965), ולמעשה על כל סימולציה של תופעות נפשיות אנושיות הנעשית על-ידי מכונת טיורינג.

בקיצור רב, ותוך השמטת פרטים שונים, אפשר לתאר את התוכנה של שנק כדלקמן: תכלית התוכנה היא לחקות את הכושר האנושי להבנת סיפורים. הכושר האנושי להבנת סיפורים

\* "Minds, Brains, and Programs," by John R. Searle, from *The Behavioral and Brain Sciences*, vol. 3. Copyright © 1980 Cambridge University Press Reprinted by permission of Cambridge University Press. תרגום: אודי תגרי

מתאפיין בכך שאדם יכול לענות על שאלות הנוגעות לסיפור כאשר המידע שהוא מוסר כולל פרטים שמעולם לא אוזכרו בסיפור במפורש. כך, למשל, נניח שסיפרו לכם את הסיפור הזה:

"אדם נכנס למסעדה והזמין המבורגר. כאשר ההמבורגר הגיע לשולחן הוא היה חרוך כפחם. האדם יצא מהמסעדה בכעס רב, בלי לשלם בעבור ההמבורגר ובלי להשאיר תשר." וכעת, לו נשאלתם: "האם האיש אכל את ההמבורגר?" סביר להניח שתענו: "לא, הוא לא אכל אותו." באופן דומה, נניח שהיו מספרים לכם את הסיפור הזה: "אדם נכנס למסעדה והזמין המבורגר; כשההמבורגר הגיע לשולחנו הוא היה מרוצה ממנו; וכשעזב את המסעדה השאיר למלצרית תשר שמן לפני ששילם את החשבון." כעת, לו נשאלתם: "האם האדם אכל את ההמבורגר?" יש להניח שהייתם עונים, "כן, הוא אכל את ההמבורגר." בדומה לכך, המכונות של שנק מסוגלות לענות באופן כזה לשאלות הנוגעות למסעדות. לשם כך, קיים בהן "ייצוג" של מידע הנוגע למסעדות מן הסוג הקיים אצל בני אדם, מידע אשר מאפשר להם לענות על שאלות מהסוג שהוצג לעיל, ביחס לסיפורים דומים. כאשר המחשב מקבל את הסיפור והשאלה מוצגת לו, הוא מדפיס תשובות מהסוג שהיינו מצפים לקבל מאדם, לו היה נמסר לו סיפור דומה. תומכי ב"מ חזקה טוענים שבהשתלשלות זו של שאלות ותשובות המחשב לא רק מחקה את היכולת האנושית אלא:

1. אפשר לומר, פשוטו כמשמעו, שהמחשב **מבין** את הסיפור ומספק את התשובות על השאלות; ו-2. המחשב והתוכנה **מסבירים** את היכולת האנושית להבין סיפור ולענות על שאלות על אודותיו.

לא נראה לי שהמחקר של שנק תומך בדרך כלשהי בשתי הטענות האלה, וזאת אנסה להראות להלן.<sup>1</sup>

אחת הדרכים לבחינת תיאוריה בדבר הנפש היא לשאול את עצמי מה היה קורה לו נפשי באמת הייתה פועלת על בסיס אותם עקרונות שאותה תיאוריה מייחסת לפעולת הנפש. הבה נחיל מבחן זה על התוכנה של שנק באמצעות הניסוי המחשבתי (Gedankenexperiment) הזה: נניח שאני סגור בחדר וניתנת לי חבילה גדולה של ניירות הכתובים בסינית. נניח גם שאיני יודע סינית (ואמנם איני יודע), לא קרוא ולא כתוב, ושאפילו איני בטוח אם אדע לזהות כתב סיני ככתב סיני, להבדיל, למשל, מכתב יפני או מסתם שרבוט חסר פשר: מבחינתי, כתב סיני אינו אלא שרבוט חסר פשר. ונניח עוד, שאחרי שאני מקבל חבילה זו של ניירות הכתובים בסינית, אני מקבל חבילה נוספת של מסמכים הכתובים בסינית יחד עם מערכת כללים המקשרים בין החבילה הראשונה לשנייה. הכללים כתובים באנגלית, ואני מבין אותם כמו כל דובר אנגלית מלידה. הכללים מאפשרים לי להתאים קבוצה אחת של סמלים פורמליים לקבוצה אחרת של סמלים פורמליים, ובמילה "פורמלי" כוונתי היא שאני מסוגל לזהות את הסמלים אך ורק לפי צורותיהם. וכעת, נניח גם שאני מקבל חבילה שלישית של סמלים

1 איני טוען, כמובן, ששנק עצמו מחויב לטענות הללו.

בסינית יחד עם הוראות מסוימות, באנגלית, המאפשרות לי להתאים פריטים מסוימים בחבילה השלישית לשתי החבילות הראשונות, וכללים אלו מנחים אותי כיצד להחזיר לנותני החבילות שמחוץ לחדרי סמלים סיניים מסוימים בעלי צורות מסוגים מסוימים, בתגובה לצורות מסוגים מסוימים שניתנות לי בחבילה השלישית. שלא בידיעתי, האנשים שנותנים לי את כל הסמלים האלה מכנים את החבילה הראשונה שקיבלתי "כתב יד", את החבילה השנייה "סיפור", ואת החבילה השלישית "שאלות". יתר על כן, את הסמלים שאני מחזיר להם בתגובה לחבילה השלישית הם מכנים "תשובות לשאלות", ואת מערכת הכללים באנגלית שניתנה לי הם מכנים "התוכנה". וכעת – כדי לסבך מעט את העלילה – תארו לעצמכם שאותם אנשים גם נותנים לי סיפורים באנגלית, שאותם אני מבין, ולאחר מכן הם שואלים אותי שאלות באנגלית הנוגעות לאותם סיפורים, ואני מחזיר להם תשובות באנגלית. נניח גם שלאחר זמן מה אני משתפר במידה כזו בהבנת ההוראות למניפולציה בסמלים הסיניים, והמתכנתים משתפרים במידה כזו בכתיבת התוכנות, שמנקודת ראות חיצונית – כלומר, מנקודת ראותו של אדם הנמצא מחוץ לחדר שבו אני סגור – תשובותי לשאלות אינן נבדלות כהוא זה מתשובות לשאלות שיתנו דוברי סינית מלידה. איש לא יכול לדעת שאינני מדבר ולו מילה אחת בסינית אך ורק על סמך התבוננות בתשובות. ונניח גם – כפי שללא ספק יהיה – שתשובותי לשאלות שאני נשאל באנגלית אינן נבדלות כהוא זה מתשובותיהם של אחרים הדוברים אנגלית מלידה, מהטעם הפשוט שאני דובר אנגלית מלידה. מנקודת ראות חיצונית – מנקודת הראות של מישהו הקורא את "תשובותי" – תשובותי לשאלות בסינית ותשובותי לשאלות באנגלית טובות באותה מידה. אולם במקרה של הסינית, בשונה מהמקרה של האנגלית, אני מפיק את התשובות באמצעות מניפולציה של סמלים פורמליים חסרי פשר. ככל שזה נוגע לסינית, אני מתנהג כמו מחשב פשוט; אני מבצע פעולות חישוב על אלמנטים פורמליים. לצרכי הסינית אני פשוט מימוש של תוכנת המחשב.

ובכן, תומכי ב"מ חזקה טוענים שהמחשב המתוכנת מבין את הסיפורים ושהתוכנה, במובן מסוים, מסבירה את ההבנה האנושית. כעת אנו נמצאים בעמדה המאפשרת לנו לבחון טענות אלו לאור ניסויי החשיבה שלנו.

1. באשר לטענה הראשונה, בדוגמה דלעיל, נראה לי מובן מאליו שאני לא מבין מילה בסיפורים הסיניים. הקלטים והפלטטים שלי אינם נבדלים מאלו של דוברי סינית מלידה, ועדיין – תהא לי כל תוכנה פורמלית אשר תהא – לא אבין דבר. מאותו טעם, המחשב של שנק אינו מבין דבר באף סיפור – בין שהוא בסינית ובין שהוא באנגלית, או בכל שפה אחרת, מאחר שבמקרה של הסינית המחשב הוא אני, ובמקרים שבהם המחשב אינו אני, אין במחשב דבר נוסף מעבר למה שיש לי במקרה שבו אינני מבין דבר.

2. באשר לטענה השנייה, שהתוכנה מסבירה הבנה אנושית, אנחנו יכולים לראות שהמחשב והתוכנה אינם מספקים תנאים מספיקים להבנה, מאחר שהמחשב והתוכנה מתפקדים ולא מתקיימת כל הבנה. ואולם, האם הם מספקים את התנאי ההכרחי להבנה, או תרומה משמעותית להבנה? אחת הטענות שמשמיעים תומכי ב"מ חזקה היא שכאשר אני מבין סיפור

באנגלית, הריני עושה בדיוק את אותם הדברים שאני עושה בסיפור שבסינית – או עוד מאותו דבר: מניפולציה של סמלים. ההבדל הוא שבשונה מהסינית, שאיני מבין, הרי שבאנגלית, שאני מבין, אני מבצע יותר מניפולציות של סמלים פורמליים. נכון שלא הוכחתי שטענה זו שקרית, אולם על פי הדוגמה לעיל היא בהחלט לא מתקבלת על הדעת. הסבירות האפשרית של טענה זו מבוססת על ההנחה שאנו יכולים לבנות תוכנה שיהיו לה אותם קלטים ופלטטים כשל דוברי שפה מלידה, ובנוסף לכך אנו מניחים שעל הדוברים חלה רמת תיאור מסוימת שלפיה הם גם מממשים תוכנה. על בסיס שתי הנחות אלו אנו מניחים שאפילו אם התוכנה של שנק אינה מסבירה לחלוטין את הבנת הסיפור, הרי שהיא מספקת לכך הסבר חלקי. ובכן, אני מניח שזו אפשרות אמפירית, אולם עד עתה לא נמסר לנו שום נימוק שיגרום לנו שנאמין שזו האמת, מאחר שמה שמוצע (גם אם לא מוכח) על-ידי הדוגמה שהצגתי הוא שתוכנת המחשב פשוט אינה רלוונטית להבנת הסיפור. במקרה של הסינית, יש לי את כל מה שבינה מלאכותית יכולה לתת לי במושגים של תוכנה, אך איני מבין דבר; במקרה של האנגלית, אני מבין הכול, ועד עתה אין כל טעם להניח שהבנתי קשורה בדרך כלשהי לתוכנות מחשב, כלומר לפעולות חישוב המתבצעות על אלמנטים המוגדרים באופן פורמלי. כל עוד התוכנה מוגדרת במושגים של פעולות חישוב על אלמנטים המוגדרים באופן פורמלי טהור, דוגמה זו מציעה שלא למנותים אלו כשלעצמם אין שום קשר מעניין להבנה. הם בוודאי אינם תנאים מספיקים, ושום טעם לא הוצע בזכות ההנחה שהם תנאים מספיקים או שהם בכלל תורמים תרומה משמעותית להבנה. שימו לב שכוחו של הטיעון אינו נעוץ פשוט בטענה שמכונות שונות עשויות להיות בעלות אותן קלטים ופלטטים בעודן פועלות על פי עקרונות פורמליים שונים – זו אינה הנקודה כלל ועיקר. עיקר כוחו של הטיעון נעוץ בטענה שכל עקרונות פורמליים טהורים שהם שנצליח להכניס למחשב לא יהיה בהם די להבנה, מאחר שאדם יהיה מסוגל לפעול לפי עקרונות פורמליים אלו בלי להבין דבר. לא ניתן כל נימוק בזכות ההנחה שעקרונות אלו הכרחיים או אפילו בעלי תרומה כלשהי, מאחר שלא ניתן כל נימוק בזכות ההנחה שכאשר אני מבין אנגלית אני פועל באמצעות תוכנה פורמלית כלשהי.

ובכן, מהו הדבר שיש לי במקרה של המשפטים באנגלית ואשר אין לי במקרה של המשפטים בסינית? התשובה הברורה היא שאני יודע מה משמעותם של הראשונים, ואילו אין לי שמץ מושג מהי משמעותם של האחרונים. אולם מאיזה דבר מורכבת הבנה זו ומדוע – יהא אשר יהא – אין אנו יכולים להקנותו למכונה? אחזור לשאלה זו בהמשך, אך לפני כן ברצוני להמשיך בדוגמה.

הזדמן לי להציג דוגמה זו בפני כמה אנשים העוסקים בבינה מלאכותית, ומעניין לציין שאין ביניהם הסכמה באשר לתשובה ההולמת. אני מקבל מגוון רב להפתיע של תשובות, ובהמשך אתייחס לתשובות הרווחות ביותר שקיבלתי (שאציין עם מקורן הגיאוגרפי).

ראשית דבר, ברצוני לחסום מספר אי הבנות רווחות לגבי "הבנה": בחלק ניכר מהדיונים האלה נעשית עבודה מתוחכמת רבה באשר למילה "הבנה". מבקרי מציינים שקיימות דרגות שונות רבות של הבנה; ש"הבנה" אינה פרדיקט דו מקומי פשוט; שקיימים אף רמות וסוגים

שונים של הבנה, ושלעיתים עיקרון השלישי המוצא\* אף אינו חל בדרך ישירה על הבעיות שצורתן "א מבין את ב"; שבמקרים רבים ההכרעה בשאלה, האם א מבין את ב, היא עניין להחלטה ואינה עניין עובדתי פשוט; וכן הלאה. באשר לכל הנקודות הללו ברצוני לומר: בוודאי, בוודאי. אלא שכולן כאחת אינן קשורות לעניין שעל הפרק. ישנם מקרים מובהקים שבהם המונח "הבנה" כפשוטו, חל, ומקרים מובהקים שבהם אין הוא חל; ושני סוגי מקרים אלו הם כל מה שאני זקוק לו לצרכי טיעון זה.<sup>2</sup> אני מבין סיפורים באנגלית; במידה פחותה יותר, אני מסוגל להבין סיפורים בצרפתית; במידה פחותה עוד יותר, סיפורים בגרמנית; ובסינית, אני מבין כלל. מאידך גיסא, המכונת שלי והמחשבון שלי אינם מבינים דבר; הם אינם שייכים לעניין הזה של הבנה. פעמים רבות אנו מייחסים – באמצעות מטאפורות ואנלוגיות – "הבנה" או פרדיקטים קוגניטיביים אחרים למכונות, למכונות חישוב או לחפצים אחרים מעשה ידי אדם, אולם ייחוס זה אינו מוכיח דבר. אנו אומרים "הדלת **יודעת** מתי להיפתח בזכות התא הפוטו-חשמלי שלה", "מכונת החישוב **יודעת** איך (מבינה איך, **מסוגלת**) לחבר ולחסר, אבל לא לחלק", ו"מד החום חש שינויים בטמפרטורה". אנו מבצעים ייחוסים אלו מטעם מעניין למדי,<sup>3</sup> הקשור לכך שחפצים מעשה ידינו הם הרחבה של מטרותינו, ולכן אך טבעי לנו לייחס להם מכוונות באופן מטאפורי; אך לגבי ידי, דוגמאות אלה אינן משכנעות מבחינה פילוסופית. המובן שבו דלת אוטומטית "מבינה הוראות" של התא הפוטו-חשמלי שלה שונה לחלוטין מהמובן שבו אני מבין אנגלית. אם המובן שבו המחשב המתוכנת של שנק מבין סיפורים אמור להיות המובן המטאפורי שבו הדלת מבינה, ולא המובן שבו אני מבין אנגלית, הרי שהסוגיה כולה לא הייתה ראויה לדיון. אלא שניואל וסימון (Newell and Simon, 1963) כותבים שההכרה שהם טוענים לקיומה במחשב היא בדיוק ההכרה של בני אדם. הישירות של טענה זו מוצאת חן בעיני, והיא טענה מסוג הטענות שבהן אדון. אטען, שבמובן המילולי של "הבנה", תוכנת המחשב מבינה את מה שהמכונת ומכונת החישוב מבינות – קרי, בדיוק לא כלום. הבנת המחשב אינה רק חלקית או לא שלמה (כמו הבנת הגרמנית שלי): היא אפס. ועתה, לתשובות.

**1. תשובת המערכת (ברקלי).** "אמנם נכון שהאדם הנעול בחדר אינו מבין את הסיפור, אך העובדה היא שהוא איבר אחד בלבד ממערכת שלמה, והמערכת כן מבינה את הסיפור. מול אותו אדם פתוחה מחברת גדולה שבה כתובים הכללים, ברשותו דפי טיוטה רבים ועפרונות לעריכת חישובים, ברשותו 'מאגרי נתונים' שבהם קבוצות של סמלים סיניים. אלא מאי, ההבנה אינה מיוחסת לפרט בפני עצמו, אלא למערכת השלמה הזו שהוא מהווה חלק ממנה".

\* עקרון השלישי המוצא הוא העיקרון הלוגי שלפיו כל טענה היא אמתית או שקרית ואין אפשרות שלישית. (הערת המתרגם)

2 כמו כן, מ"הבנה" משתמע הן מצבים נפשיים (בעלי מכוונות) והן האמת (תקפות, הצלחה) של מצבים אלו. למטרת דיון זה אנו מגבילים עצמנו לעיסוק במצבים אלו בלבד.

3 מכוונות היא, לפי הגדרתה, אותו מאפיין של מצבים נפשיים מסוימים אשר בעטיו הם מכוונים על או אל או נסבים על אודות אובייקטים ומצבי עניינים בעולם. כך, האמונות, הרצונות והכוונות הם מצבים בעלי מכוונות, ואילו צורות של חרדה או דיכאון ללא אובייקט אינן כאלה. לדיון נוסף ראו Searle, 1979.

תשובתי לתיאוריית המערכת פשוטה למדי: נניח לפרט להפנים את כל המרכיבים של המערכת. הוא ישנן את הכללים הכתובים במחברת ואת מאגרי הנתונים של הסמלים הסיניים, והוא יבצע את כל החישובים בראשו. באופן זה, הפרט יטמיע את המערכת בשלמותה. אין במערכת דבר שאין הוא מקיף. אנו יכולים אפילו להיפטר מהחדר ולהניח שהוא עובד תחת כיפת השמים. ועדיין הוא אינו מבין סינית כהוא זה, ומן הסתם המערכת אינה מבינה דבר, מפני שאין במערכת הזאת דבר שאינו נמצא בתוכו. אם הוא אינו מבין, הרי שאין כל דרך שבה המערכת תוכל להבין, מפני שהמערכת אינה אלא חלק ממנו.

לאמתו של דבר, אני נבוכ מעט בשל כך שאני נאלץ לתת אפילו תשובה זו לתיאוריית המערכת, מפני שהתיאוריה נראית לי כה בלתי סבירה מלכתחילה. הרעיון הבסיסי שלה הוא שבעוד האדם אינו מבין סינית, איכשהו החיבור של אותו אדם וכמה פיסות נייר עשוי להבין סינית. לא קל לי לדמיין כיצד אדם שאינו לפות בכבלי אידאולוגיה מסוימת ייחס סבירות כלשהי לרעיון זה. ואף על פי כן, אני חושב שאנשים רבים אשר מחויבים לאידיאולוגיה של ב"מ חזקה יטו בסופו של דבר לומר משהו הדומה מאוד לטענה זו. לפיכך, הבה נוסיף ונעמיק בה. על פי אחת הגרסאות של השקפה זו, בעוד האדם בדוגמת הפנמת המערכת אינו מבין סינית במובן שבו דובר סינית מלידה מבין סינית (שכן, למשל, הוא אינו יודע שהסיפור מתייחס למסעדות ולהמבורגרים, וכדומה), "האדם כמערכת מניפולציה של סמלים פורמליים" **באמת מבין סינית**. אין לבלבל בין תת המערכת, שבתוך אותו אדם, המבצעת מניפולציה של סמלים פורמליים בסינית, לבין תת המערכת המטפלת בסמלים באנגלית.

כלומר, בעצם קיימות באדם שתי תת מערכות; האחת מבינה אנגלית, השנייה סינית, ו"פשוט מאוד, שתי המערכות כמעט ואינן קשורות זו לזו". אבל, ברצוני לענות, לא זו בלבד שהן כמעט ולא קשורות זו לזו, אלא שאין ביניהן כלל קווי דמיון. אם נרשה לעצמנו לדבר לרגע בעגה זו של 'תת מערכות', אזי תת המערכת המבינה אנגלית יודעת שהסיפור עוסק במסעדות ובאכילת המבורגרים, היא יודעת שהיא נשאלת שאלות על מסעדות ושהיא עונה לשאלות כמיטב יכולתה באמצעות היסקים מסוימים מתוכן הסיפור, וכן הלאה. אולם תת המערכת הסינית אינה יודעת דבר מכל זה. תת המערכת האנגלית יודעת ש"המבורגרים" מציין המבורגרים, ואילו תת המערכת הסינית יודעת שלאחר "שרבוט מקושקש" יבוא "קשקוש משורבט" ולא יותר. היא יודעת אך ורק שבצד אחד שלה מוצגים סמלים פורמליים מסוימים שבהם נעשית מניפולציה על פי כללים הכתובים בשפה האנגלית – ושבצד השני יוצאים סמלים אחרים. והרי לב הטיעון בדוגמה המקורית היה שמניפולציית סמלים מסוג זה כשלעצמה אינה מספיקה להבנת סינית במובן המילולי של "הבנת סינית", מפני שהאדם יכול לכתוב "שרבוט מקושקש" אחרי "קשקוש משורבט" בלי להבין סינית בכלל. והנחת תת המערכות בתוך האדם אינה עונה לטיעון זה מפני שמצב תת המערכות אינו טוב יותר ממצבו של האדם מלכתחילה; עדיין אין להן ולו קצה קצהו של הדבר הקיים באדם (או בתת המערכת) הדוברים אנגלית. ואמנם, במקרה כפי שהוא מתואר, תת המערכת הסינית היא פשוט חלק מתת המערכת האנגלית, חלק העוסק במניפולציה של סמלים חסרי משמעות בהתאם לכללים הכתובים באנגלית.

הבה נשאל את עצמנו מה אמור לעמוד מלכתחילה ביסוד תשובת המערכת; כלומר, מה אמור להיות הבסיס העצמאי לקביעה, שבתוך האדם חייבת להיות תת מערכת אשר מבינה, פשוטו כמשמעו, סיפורים בסינית? למיטב הבנתי, הבסיס היחיד הוא שבדוגמה יש לי קלט ופלט הזהים לאלו של דוברי סינית מלידה ותוכנה שמוליכה מן הראשון לאחרון. אולם כל עניין של הדוגמאות היה לנסות ולהראות שאין די באלה להבנה, באותו מובן שבו אני מבין סיפורים באנגלית, מפני שבאדם – ולפיכך בתת המערכות שמהן האדם מורכב – יכול להתקיים השילוב הנכון של קלטים, פלטים ותוכנה, והוא עדיין לא יבין שום דבר במובן שבו אני מבין אנגלית. הסיבה הבסיסית היחידה לאמירה ש**חייבת** להיות בי תת מערכת אשר מבינה סינית היא שיש בי תוכנה ואני יכול לעבור בהצלחה את מבחן טיורינג; אני יכול לשטות בדוברי סינית מלידה. אולם אחת מנקודות המחלוקת היא בדיוק זו – האם מבחן טיורינג בכלל תקף? הדוגמה מראה שעשויות להיות שתי "מערכות", אשר שתיהן יעברו בהצלחה את מבחן טיורינג, אבל רק אחת מהן מבינה; והטענה שמאחר ששתיהן עוברות את מבחן טיורינג, שתיהן מבינות, אינה מפריכה זאת, שכן היא אינה עונה לטענה, שאותה מערכת שבתוכי אשר מבינה אנגלית כוללת הרבה יותר מאשר המערכת אשר אך ורק מעבדת סינית. בקצרה, תשובות המערכת פשוט מניחה את המבוקש, בהתעקשה – ללא כל טיעון – שהמערכת חייבת להבין סינית.

יתר על כן, תשובת המערכת מובילה כמדומה להשלכות אבסורדיות כשלעצמן. אם עלינו להסיק שחייבת להיות בי הכרה, על בסיס הנימוק שיש בי סוג מסוים של קלטים ופלטים שביניהם תוכנה, אזי עלינו להחשיב גם כל מיני תת מערכות לא קוגניטיביות כקוגניטיביות. כך למשל, קיימת רמה של תיאור שבה הקיבה שלי מבצעת עיבוד מידע, ובכך היא מממשת תוכנות מחשב, אבל ברור שאיננו רוצים לומר שיש בה הבנה כלשהי (השוו: Pylyshyn, 1980). אולם אם נקבל את תשובת המערכת, אזי קשה יהיה לראות כיצד נוכל להימנע מהטענה שהקיבה, הלב, הכבד וכן הלאה הן כולן תת מערכות מבינות, היות שלא קיימת דרך עקרונית להבחין בין המניע היסודי לטענה שתת המערכת הסינית מבינה לבין הטענה שהקיבה מבינה. ובדרך אגב, הטענה שהקלטים והפלטים במערכת הסינית הם מידע ואילו הפלטים והקלטים בקיבה הן מזון ותוצרי מזון, אינה עונה לנקודה זו, מאחר שמנקודת ראותו של האדם, קרי מנקודת הראות שלי, אין מידע לא במזון ולא בסינית – הסינית אינה אלא שרבוטים חסרי משמעות. במקרה של הסינית, המידע קיים אך ורק בעיניהם של המתכנתים והמפרשים, ואם רק ירצו בכך, דבר לא ימנע מהם להתייחס לקלטים ולפלטים של איברי מערכת העיכול כאל מידע.

נקודה אחרונה זו נוגעת לכמה בעיות בלתי תלויות בב"מ חזקה, וכדאי לסטות לרגע כדי להסביר אותן. אם ב"מ חזקה עתידה להיות ענף של הפסיכולוגיה, הרי שעליה להיות מסוגלת להבחין בין מערכות שהן באמת נפשיות לבין מערכות שאינן נפשיות. עליה להיות מסוגלת להבחין בין העקרונות שעל בסיסם פועלת הנפש לבין עקרונות שעל בסיסם פועלות מערכות לא נפשיות; שאם לא כן, הם לא יוכלו להסביר לנו מה בנפש ייחודי לנפש. והאבחנה בין נפשי ללא נפשי אל לה להתקיים רק בעיניו של המתבונן, היא חייבת להיות מהותית למערכות;

שאם לא כן, כל מתבונן שיעלה הרצון מלפניו יוכל להתייחס לבני אדם כאל לא-נפשיים, ואילו לסערות הוריקן, למשל, כאל נפשיות. אולם בספרות של הב"מ אבחנה זו מטושטשת לעתים קרובות בדרכים שבטווח הארוך יחתרו תחת הטענה שב"מ היא תחום חקירה קוגניטיבי. מקארתי, לדוגמה, כותב: "אפשר לומר שלמכוונות פשוטות כמו מדי חום יש האמנות, והקיום של האמנות מאפיין כמדומה את רוב המכוונות המסוגלות לפתור בעיות" (McCarthy, 1979). מן הראוי שכל מי שחושב שלב"מ יש סיכוי כתיאוריה של הנפש ישקול את השלכותיה של הערה זו. אנחנו מתבקשים לקבל כתגלית של ב"מ חזקה שבגוש המתכת התלוי על הקיר ואשר משמש אותנו לוויסות הטמפרטורה יש האמנות בדיוק באותו מובן שלנו, לבני זוגנו, או לבנות זוגנו ולילדינו יש האמנות; ולא זו בלבד, אלא של"רוב" המכשירים האחרים בחדר – טלפון, מכשיר הקלטה, מכונת חישוב, מתג אור החשמל – גם להם יש האמנות במובן הזה. מטרת המאמר הנוכחי אינה לטעון כנגד הנקודה של מקארתי, ולכן פשוט אקבע את הדברים הבאים ללא טיעון. חקר הנפש מבוסס על עובדות כמו אלו שלבני אדם יש האמנות, ואילו לווסתי חום, לטלפונים ולמכוונות חישוב אין האמנות. אם מישהו מציג תיאוריה אשר שוללת את הנקודה הזו, פירוש הדבר שהוא מציג דוגמה נגדית לתיאוריה, ושהתיאוריה שקרית. מתקבל הרושם שאנשים העוסקים בב"מ ואשר כותבים דברים מעין אלו סבורים שביכולתם לחמוק ממסקנה זו, מפני שהם לא באמת מתייחסים לעניין ברצינות, והם גם אינם סבורים שמישהו אחר יתייחס לדבריהם ברצינות. אני מציע להתייחס לאמירה זו ברצינות לכל הפחות לכמה רגעים. הקדישו דקה של מחשבה לשאלה מה נדרש כדי לבסס את הטענה שלאותו גוש מתכת התלוי על הקיר יש האמנות אמתיות, האמנות עם כיוון התאמה, תוכן טענתי, ותנאי סיפוק; האמנות אשר יכולות להיות האמנות חזקות או האמנות חלשות; האמנות עצבניות, חרדות או בטוחות; האמנות דוגמטיות, רציונאליות או דעות קדומות; האמנות בלתי מעורערות או הרהורים מהוססים. מד החום אינו עומד בתנאים הנדרשים. גם לא הקיבה, הכבד, מכונת החישוב או הטלפון. אולם, מאחר שאנו מתייחסים לרעיון הזה ברצינות, ראוי לשים לב שאמתותו תחת תביעתה של ב"מ חזקה להיות מדע של הנפש. שכן לפי רעיון זה, הנפש נמצאת בכל מקום. ומה שרצינו לדעת הוא מה מבחין בין הנפש לבין וסתי חום או כבדים. ואם מקארתי צודק, אזי לב"מ חזקה לא היה שמץ סיכוי לספק את התשובה לשאלה זו.

**2. תשובת הרובוט (ייל).** "נניח שהיינו כותבים תוכנה שונה מהתוכנה של שנק. נניח שהיינו מתקינים מחשב בתוך רובוט, ומחשב זה לא רק היה מקבל קלטים של סמלים פורמליים ופולט פלטים של סמלים פורמליים, אלא גם היה מפעיל את הרובוט כך שהוא היה מבצע פעולות הדומות מאוד לתפיסה, הליכה, התנועעות, תקיעת מסמרים בקירות, אכילה, שתייה – כל מה שנרצה. כך, למשל, על גוף הרובוט תהיה מותקנת מצלמת טלוויזיה שתאפשר לו 'לראות', לרובוט יהיו ידיים ורגליים שיאפשרו לו 'לפעול', וכל זה יהיה בשליטה של ה'מוח' הממוחשב שלו. בשונה מהמחשב של שנק, לרובוט כזה תהיה הבנה ממשית, ומצבים נפשיים נוספים."



הדבר הראשון הראוי לציון בתשובת הרובוט הוא הודאתה במשתמע שקוגניציה אינה אך ורק עניין של מניפולציה בסמלים פורמליים, שכן תשובה זו מוסיפה מערך של יחסים סיבתיים עם העולם החיצוני (השוו Fodor, 1980). אולם תשובתי לתיאוריית הרובוט תהיה שהתוספת של כישורים "תפיסתיים" ו"מוטוריים" אלו אינם מוסיפים דבר לתוכנה של שנק במושגים של הבנה בפרט ומכוונות בכלל. כדי לראות זאת, הבה נחיל את ניסוי המחשבה הנדון על מקרה הרובוט. נניח שבמקום המחשב שבתוך הרובוט מכניסים אותי לתוך חדר, וכמו במקרה הסיני המקורי נותנים לי סמלים סיניים נוספים והנחיות נוספות באנגלית בדבר התאמת סמלים סיניים לסמלים סיניים אחרים ולמשוב החוזר של סמלים סיניים מהחדר החוצה. עוד נניח, כי שלא בידיעתי, חלק מהסמלים הסיניים שמגיעים אלי מקורם במצלמת טלוויזיה המחוברת אל הרובוט, ואילו סמלים סיניים אחרים שאני מוציא מן החדר גורמים למנועים שבתוך הרובוט שיניעו את רגליו וידיו. חשוב להדגיש שכל מה שאני עושה הוא מניפולציה של סמלים פורמליים: אין לי כל ידיעה בנוגע לעובדות הנוספות הללו. אני מקבל "מידע" ממנגנון ה"תפיסה" של הרובוט, ואני מפיק "הוראות" למנגנון ההנעה שלו בלי לדעת דבר הנוגע לעובדות אלו. אני ההומונקולוס – "האיש הקטן" – של הרובוט, אבל בשונה מההומונקולוס המסורתי אני יודע מה מתרחש. אינני מבין דבר פרט לכללים לביצוע המניפולציות בסמלים. במקרה זה, לטענתי, לרובוט אין שום מצבים בעלי מכוונות; הוא פשוט נע אנה ואנה כתוצאה של החיווט החשמלי שבגופו והתוכנה שלו. יתר על כן, בממשי את התוכנה, לא קיימים בי שום מצבים בעלי מכוונות מהסוג הרלוונטי. כל שאני עושה הוא מילוי הוראות פורמליות בדבר מניפולציה של סמלים פורמליים.

**3. תשובת סימולטור המוח (ברקלי ואם.איי.טי).** "נניח שאנחנו מתכננים תוכנה שאינה מייצגת מידע שיש לנו על אודות העולם, כמו המידע בכתבים של שנק, אלא מחקה את ההשתלשלות הממשית של פריקת מטעני הנורונים בסינפסות המוח של דובר סינית מלידה בשעה שהוא מבין סיפורים בסינית ועונה לשאלות על אודותיהם. המכונה מוזנת בקלטות של סיפורים בסינית ושאלות על אודותיהם, היא מחקה את המבנה הפורמלי של מוחות סיניים ממשיים המעבדים את הסיפורים הללו, ומפיקה תשובות סיניות כפלטות. אנחנו יכולים אף לדמיין שהמכונה אינה פועלת עם תוכנה סדרתית בודדת, אלא עם מערך שלם של תוכנות הפועלות במקביל, באותו אופן שמוח אמיתי של אדם פועל – אנו משערים – כשהוא מעבד שפה טבעית. והרי עתה בוודאי נהיה חייבים לומר שהמכונה מבינה את הסיפורים; ואם נסרב לומר זאת, האם לא נצטרך גם להכחיש שדוברי סינית מלידה מבינים את הסיפורים? וכי מה שונה – או עשוי להיות שונה – במישור הסינפטי בין תוכנת המחשב לבין תוכנת המוח הסינית?"

לפני שנתייחס למענה זה ברצוני לסטות מהנושא כדי לציין שמדובר בתשובה משונה ביותר בעבור חסיד של בינה מלאכותית (או פונקציונליזם, וכו'): לתומי חשבתי שכל הרעיון של ב"מ חזקה הוא שאיננו צריכים לדעת איך עובד המוח כדי לדעת איך עובדת הנפש. ההיפותזה הבסיסית – כך לפחות שיערתי – היא שקיימת רמה של פעולות נפשיות, המורכבת מתהליכי חישוב של אלמנטים פורמליים המכונים את מהות היסוד הנפשי, והניתנים למימוש

בתהליכי מוח שונים ומשונים, באותו האופן שבו תוכנת מחשב ניתנת למימוש בחומרות מחשב שונות: לפי ההנחות של ב"מ חזקה, היחס בין הנפש למוח הוא כמו זה שבין התוכנה לחומרה, וכך אנחנו יכולים להבין את הנפש בלי לעסוק בניורופיזיולוגיה. אם העיסוק בבינה מלאכותית היה מחייב אותנו לדעת כיצד המוח פועל, אזי לא היינו טורחים לפתח ב"מ. מכל מקום, אפילו קרבה רבה זו לפעולות המוח עדיין אינה מספיקה כדי להפיק הבנה. כדי לראות זאת, דמיינו שבמקום אדם חד-לשוני היושב בחדר ומתמך סמלים, יש לנו אדם המפעיל מערך מורכב של צינורות מים המחוברים באמצעות שסתומים וברזים. כאשר אותו אדם מקבל את הסמלים בסינית הוא מעיין בתוכנה, הכתובה באנגלית, שבה כתוב אילו שסתומים וברזים עליו לפתוח ואילו מהם לסגור. כל חיבור מים מקביל לסינפסה במוח הסיני, והמערכת כולה מתוכננת בצורה כזאת שאחרי סיום פריקת כל המטענים הנכונים, כלומר אחרי הפעלת כל השסתומים והברזים הנכונים, התשובות בסינית מופיעות בקצה שורת הצינורות בצד הפלט.

ובכן, היכן ההבנה במערכת זו? היא מקבלת קלטים בסינית, היא מחקה את המבנה הפורמלי של הסינפסות במוח הסיני, והיא מפיקה קלטים בסינית. אולם האיש אינו מבין סינית בכלל, וכמוהו גם צינורות המים, ואם אנו מתפתים לאמץ את ההשקפה האבסורדית (לדעת) שאיכשהו החיבור של האיש עם צינורות המים מבין סינית, זכרו שבאופן עקרוני האיש יכול להטמיע את המבנה הפורמלי של צינורות המים ולבצע בדמיונו את כל "פריקת המטען העצבי". הבעייתיות של סימולטור המוח היא שהוא מחקה את הדברים הלא נכונים במוח. כל עוד הוא מחקה רק את המבנה הפורמלי של פריקת המטען העצבי בסינפסות, הוא לא יחקה את הדבר החשוב במוח, קרי התכונות הסיבתיות שלו, יכולתו להפיק מצבים בעלי מכוונות. ואילו דוגמת צינורות המים מראה שהתכונות הפורמליות אינן מספיקות לתכונות סיבתיות: התכונות הפורמליות של המערכת הניורוביולוגית הרלוונטית יכולות להתממש בלעדי התכונות הניורוביולוגיות.

**4. תשובת הקומבינציה (ברקלי וסטנפורד).** "אמנם כל אחת משלוש התשובות הקודמות אינה משכנעת לגמרי כשלעצמה כהפרכה של הדוגמה הנגדית של החדר הסיני, אך כשהן מצורפות יחדיו משקלן המשותף משכנע הרבה יותר ואפילו מכריע. דמיינו רובוט, שמחשב-דמוי-מוח מותקן בחלל גולגלתו, דמיינו שהמחשב מתוכנת עם כל הסינפסות הקיימות במוח האנושי, דמיינו שהתנהגות הרובוט כולה אינה מובחנת מהתנהגותו של אדם, וכעת חישבו על כל אלה כמערכת מאוחדת ולא רק כמחשב עם קלטים ופלטים. במקרה כזה נצטרך, ללא ספק, לייחס למערכת מכוונות".

אני מסכים לגמרי שבמקרה דנן, כל עוד זה היה המידע היחיד המצוי בידינו, היפותזה זו, שלרובוט יש מכוונות, היתה נראית הגיונית ואף בלתי נמנעת. למעשה, פרט למראה ולהתנהגות, שאר מרכיבי השילוב אינם רלוונטיים. לו יכולנו לבנות רובוט שטווח רחב של התנהגותו אינו מובחן מהתנהגות אנושית, היינו מייחסים לו מכוונות, כל עוד לא מתקיים

טעם כנגד ייחוס כזה. לא היינו צריכים לדעת מראש שמוחו הממוחשב הוא אנלוג פורמלי של המוח האנושי.

אלא שבאמת איני רואה כיצד כל זה יכול לסייע לטענות של ב"מ חזקה, מהטעמים הללו: על פי תומכי ב"מ חזקה, מימוש תוכנה פורמלית בעלת פלטים וקלטים נכונים הוא תנאי מספיק – אכן, מרכיב מכונן – למכוונות. כפי שניסח זאת ניואל (Newell, 1979), מהות היסוד הנפשי היא פעולתה של מערכת סמלים פיזיקלית. אולם ייחוס המכוונות לרובוט בדוגמה זו אינם קשורים כלל לתכונות פורמליות. הם מבוססים פשוט על ההנחה שאם הרובוט נראה ומתנהג כמונו במידה מספקת, הרי שנשער, עד שיוכח אחרת, שיש לו מצבים נפשיים כשלנו, אשר גורמים להתנהגותו ומתבטאים בהתנהגותו, ושהוא חייב להיות בעל מנגנון פנימי המסוגל להפיק מצבים נפשיים כאלה. לו ידענו להסביר את התנהגותו ללא הנחות אלו, הרי שלא היינו מייחסים לו מכוונות, ובמיוחד לו ידענו שקיימת בו תוכנה פורמלית. וזו בדיוק הנקודה המרכזית בתשובתי הקודמת להתנגדות השנייה.

נניח שידענו שהתנהגותו של הרובוט מוסברת לגמרי על בסיס הנתונים הללו: אדם שנמצא בתוכו מקבל סמלים פורמליים בלתי מפורשים מהחיישנים של הרובוט ושולח סמלים פורמליים בלתי מפורשים למנגנוני התנועה שלו, ואותו אדם מבצע את מניפולציית הסמלים הללו על פי קובץ הוראות. יתר על כן, נניח שאותו אדם חף מכל ידיעה בדבר עובדות אלו הנוגעות לרובוט; הוא יודע אך ורק אילו פעולות עליו לבצע על אילו מן הסמלים נטולי המשמעות. במקרה זה היינו מתייחסים אל הרובוט כאל גולם מכני משוכלל. ההשערה שלגולם זה יש נפש הייתה חסרת הצדקה ולא הכרחית, מפני שכעת כבר אין כל טעם לייחס לרובוט או למערכת שהוא חלק ממנה מכוונות כלשהי (פרט, כמובן, למכוונות של האדם המבצע את המניפולציה בסמלים). מניפולציית הסמלים הפורמליים ממשיכה ללא הפסקה, הקלטים והפלטים מותאמים בצורה נכונה, אולם מוקד המכוונות הממשי היחיד הוא האדם, והוא אינו מכיר אף אחד מן המצבים המכוונים הרלוונטיים; כך למשל, הוא אינו **רואה** מה שמגיע לעיניו של הרובוט, הוא אינו **מתכוון** להניע את זרועו של הרובוט, והוא אינו **מבין** אף אחת מהאמירות שהרובוט אומר או שנאמרות לו; ומאותם טעמים שצויינו לעיל, גם המערכת שהאדם והרובוט הם חלק ממנה אינה רואה, אינה מתכוונת, ואינה מבינה.

כדי להבין נקודה זו, השוו מקרה זה למקרים שבהם נראה לנו טבעי לגמרי לייחס מכוונות למיני פרימאטים אחרים כמו קופים או לחיות בית כמו כלבים. הטעמים שבעטיים הדבר נראה לנו טבעי הם, בערך, הטעמים הללו: איננו יכולים למצוא פשר להתנהגות החיה מבלי לייחס לה מכוונות, ואנחנו רואים שאותן חיות עשויות מחומר שדומה לחומר שלנו עצמנו – זו עין, זה אף, זה העור של החיה, וכן הלאה. לאור ההתנהגות העקבית של החיה וההנחה שאותו חומר הוא הבסיס הסיבתי להתנהגותה, אנחנו מניחים הן שהתנהגותה חייבת להיות מבוססת על מצבים נפשיים, והן שהמצבים הנפשיים חייבים להיות תולדה של מנגנונים המורכבים מחומר הדומה לחומר שלנו. אין ספק שהיינו מניחים הנחות דומות לגבי הרובוט אלמלא היה לנו טעם שלא לעשות זאת, אולם ברגע שהיה נודע לנו שהתנהגותו היא תולדה של תוכנה

פורמלית, ושהתכונות הסיבתיות הממשיות של העצב החומרי אינן רלוונטיות, היינו נוטשים את הנחת המכוונות.

ישנן שתי תגובות נוספות לדוגמה שלי אשר מועלות לעתים קרובות (ולכן ראויות לדיון), אבל גם הן מחטיאות את הנקודה.

**5. תשובת הנפשות האחרות (ייל).** "איך אתה יודע שאנשים אחרים מבינים סינית, או כל דבר אחר? רק על פי התנהגותם. ובכן, המחשב יכול לעבור את מבחני התנהגות, כפי שהם יכולים, ולכן אם אתה מתכוון לייחס הכרה לאנשים אחרים, הרי שאתה חייב באופן עקרוני לייחס אותה גם למחשבים."

התנגדות זו ראויה לאמתו של דבר לתשובה קצרה בלבד. הבעיה בדיון זה אינה איך אני יודע שלאנשים אחרים יש מצבים קוגניטיביים, אלא בעצם מהו הדבר שאני מייחס להם כאשר אני מייחס להם מצבים קוגניטיביים. לב הטיעון הוא שלא ייתכן שמדובר אך ורק בתהליכי חישוב ובפלט שלהם מפני שתהליכי חישוב והפלט שלהם יכולים להתקיים ללא מצב קוגניטיבי. העמדת פנים של הימצאות במצב של היעדר תחושה אינה תשובה לטיעון זה.<sup>\*</sup> ב"מדעים הקוגניטיביים" אנו קדם-מניחים את מציאותו של המצב הנפשי ואת הינתנותו לידיעה באותו אופן שבו במדעים הפיזיקליים אדם חייב להניח את מציאותם של אובייקטים פיזיקליים את הינתנותם לידיעה.

**6. תשובת ריבוי המדורים (ברקלי)\*\*** "הטיעון שלך בשלמותו קדם-מניח שב"מ מתייחסת אך ורק למחשבים אנלוגיים ודיגיטליים. אבל זהו עניין מקרי הקשור למציאות הטכנולוגית הנוכחית. יהיו אשר יהיו התהליכים הסיבתיים שלדברך מהותיים למכוונות (בהנחה שאתה צודק), בבוא העת נצליח לבנות מכשירים שמתבצעים בהם תהליכים סיבתיים אלה, וזו תהיה בינה מלאכותית. כך שהטיעונים השונים שלך אינם מאיימים בשום צורה על יכולתה של בינה מלאכותית ליצור הכרה ולהסביר אותה."

באמת שאין לי כל התנגדות לתשובה זו, וברצוני רק לומר שבפועל היא מגמדת את הפרוייקט של ב"מ חזקה כי היא מגדירה ב"מ חזקה מחדש ככל דבר אשר מייצר הכרה באופן מלאכותי ומסביר אותה. הקביעה המקורית של בינה מלאכותית עוררה עניין מפני שהיא הייתה תזה ומוגדרת היטב: תהליכים נפשיים הם תהליכי חישוב על אלמנטים מוגדרים פורמלית. מטרתי היא לקרוא תיגר על תזה זו. אם הטענה מוגדרת מחדש כך שאינה עוד התזה הזו, אי אפשר להחיל את התנגדויותי שכן לא מתקיימת עוד היפותזה הניתנת לבחינה שעליה אפשר להחילן.

\* במשפט זה מרמז סרל לביקורת שהופנתה כנגד הביהביוריסטים שלפיה אפשר להאמין בהשקפתם רק אם מעמידים פנים שנמצאים במצב של היעדר תחושה. (הערת המתרגם)

\*\* שם זה לתשובה לקוח ממשפט מן הברית החדשה: "בבית אבי מדורים רבים" (הבשורה לפי יוחנן, פרק י"ד, 2). סרל בחר בשם זה לתשובה משום שלפיה ייתכן מגוון גדול של מחשבים ואילו טיעון החדר הסיני מתייחס רק לסוגים מעטים של מחשבים. (הערת המתרגם)

נחזור כעת לשאלה שעליה הבטחתי שאשתדל לענות: בהנחה שבדוגמה המקורית שנתתי אני מבין את האנגלית, אך אינני מבין את הסינית, וכך שהמכונה אינה מבינה אנגלית ואינה מבינה סינית, הרי עדיין חייב להתקיים בי משהו שבעטיו אני מבין אנגלית, וחייב להיות חסר בי משהו מקביל שבשל חסרונו אינני מצליח להבין סינית. ועתה, מדוע שלא נוכל להקנות את המשהו הזה, יהא אשר יהא, למכונה?

איני רואה כל טעם עקרוני מדוע לא נוכל להקנות למכונה את היכולת להבין אנגלית או סינית, מאחר שבמובן חשוב גופינו עם מוחותינו הם בדיוק מכונות כאלו. אולם אני כן רואה טיעונים חזקים מאד בזכות הטענה שלא נוכל להקנות דבר מעין זה למכונה, כאשר פעולת המכונה מוגדרת אך ורק במונחים של תהליכי חישוב על אלמנטים מוגדרים פורמלית; כלומר, כאשר פעולת המכונה מוגדרת כמימוש של תוכנת מחשב. אני מסוגל להבין אנגלית ולהיות בעל צורות אחרות של מכוונות לא מפני שאני המימוש של תוכנת מחשב (אני עצמי מממש, אני משער, אינספור תוכנות מחשב), אלא, ככל שידיעתנו מגעת, מפני שאני אורגניזם מסוג מסוים בעל מבנה ביולוגי (כלומר, כימי ופיזיקלי) מסוים, ומבנה זה, בתנאים מסוימים, מסוגל להיות סיבתן של תפיסות, פעולות, הבנה, למידה, ותופעות מכוונות אחרות. וחלק מהנקודה של הטיעון הנוכחי הוא שרק משהו שיש לו את הכוחות הסיבתיים הללו יכול להיות בעל מכוונות כזו. ייתכן שתהליכים פיזיקליים וכימיים אחרים יכולים להפיק בדיוק את אותן תולדות; ייתכן, למשל, שגם לבני המאדים יש מכוונות, אך מוחותיהם עשויים מחומר שונה. זו שאלה אמפירית, הדומה לשאלה האם חומר שהכימיה שלו שונה מזו של כלורופיל יכול לגרום פוטו-סינתזה.

אולם הנקודה העיקרית של הטיעון הנוכחי היא שמודל פורמלי טהור לעולם לא יהיה כשלעצמו תנאי מספיק למכוונות, מפני שהתכונות הפורמליות כשלעצמן אינן תכונות מכוונות של מכוונות, ולכן עצמן אינן כוח סיבתי פרט לאותו כוח לייצר, כשהן ממומשות, את השלב הבא של הפורמליזם בזמן שהמכונה פועלת. וכל תכונות סיבתיות אחרות המתקיימות במימושים מסוימים של המודל הפורמלי אינן רלוונטיות למודל הפורמלי, מפני שנוכל תמיד לממש את המודל הפורמלי באופן אחר, שממנו נעדרות בבירור אותן תכונות סיבתיות. גם אם, בדרך נס, דוברי סינית יממשו בדייקנות את התוכנה של שנק, הרי שנוכל להתקין את אותה תוכנה בדוברי אנגלית, בצינורות מים או במחשבים שאינם מבינים אנגלית, חרף התוכנה.

הגורם המשמעותי בפעולת המוח אינו הצל הפורמלי שמטיל רצף הסינפסות, אלא התכונות הממשיות של הרצפים. כל הטיעונים בזכות הגרסה החזקה של בינה מלאכותית שראיתי מתעקשים לשרטט קו סביב הצללים שמטילה ההכרה, ואז טוענים שהצללים עצמם הם הדבר האמיתי.

לסיכום, ברצוני לנסות ולציין במפורש חלק מהנקודות הפילוסופיות הכלליות הכלולות בטיעון במובלע. למען הבהירות, אנסה לעשות זאת במבנה של שאלה ותשובה, ואתחיל באותה שאלה שכבר אבד עליה כלח:

"האם מכונה יכולה לחשוב?"

התשובה היא בבירור: כן. אנו הננו בדיוק מכוונות כאלה.

"בסדר, אבל האם מכשיר, מכונה שאדם בנה, מסוגלים לחשוב?"

בהנחה שניתן לייצר באמצעים מלאכותיים מכונה בעלת מערכת עצבים, תאי עצב בעלי אקסונים ודנדריטים וכל השאר, הדומים לאלה שלנו במידה מספקת, אזי שוב התשובה לשאלה נראית ברורה מאלה: כן. אם אפשר לשכפל בדייקנות את הסיבות, הרי שאפשר לשכפל את התולדות. ואמנם, ייתכן שאפשר לייצר מודעות, מכוונות וכל השאר על בסיס עקרונות כימיים שונים מאלו שעליהם מבוסס המין האנושי. כפי שאמרתי, המדובר בשאלה אמפירית.

"בסדר, אבל האם מחשב דיגיטלי יוכל לחשוב?"

אם "מחשב דיגיטלי" אנו מתכוונים לכל דבר שקיימת לגביו רמת תיאור כלשהי שלפיה הוא מימוש של תוכנת מחשב, הרי שוב התשובה היא, כמובן: כן, מפני שאנו עצמנו מימושים של אינספור תוכנות מחשב, ואנו מסוגלים לחשוב.

"אבל האם משהו מסוגל לחשוב, להבין וכן הלאה **אך ורק** על בסיס היותו מחשב שבו מותקנת התוכנה הנכונה? האם מימוש של תוכנה, של התוכנה הנכונה, כמובן, הוא כשלעצמו תנאי מספיק להבנה?"

זו, לדעתי, השאלה הנכונה שצריכה להישאל, אף על פי שבדרך כלל מבלבלים בינה לבין אחת או יותר מהשאלות הקודמות, והתשובה עליה היא: לא.

"מדוע לא?"

מפני שלמניפולציות בסמלים הפורמליים אין מכוונות; הן חסרות משמעות; הן אפילו לא מניפולציות של **סמלים**, מפני שהסמלים אינם מסמלים דבר. בעגה הבלשנית, יש להן תחביר, אך לא סמנטיקה. המכוונות שלכאורה קיימת במחשבים נמצאת אך ורק בראשם של אלה שמתכננים אותם ושל אלה שמשתמשים בהם, של אלה השולחים פנימה את הקלט ושל אלה המפרשים את הפלט.

המטרה של דוגמת החדר הסיני היתה לנסות ולהראות זאת על ידי כך שמראים שברגע שאנחנו מכניסים למערכת גורם שבאמת יש לו מכוונות (אדם), ומתכנתים אותו עם התוכנה הפורמלית, אנו רואים שהתוכנה הפורמלית אינה מוסיפה שום מכוונות. היא אינה מוסיפה דבר ליכולתו של האדם להבין סינית, למשל.

בדיוק אותו מאפיין של ב"מ שנראה כה מצודד – האבחנה בין התוכנה לבין מימושה – הוכח כקטלני לטענה שחיקוי יכול להיות שכפול. האבחנה בין התוכנה לבין מימושה בחומרה נראית מקבילה לאבחנה בין רמת הפעולות הנפשיות לבין רמת הפעולות המוחיות. ולו יכולנו לתאר את רמת הפעולות הנפשיות כתוכנה פורמלית, הרי דומה שהיה ביכולתנו לתאר את מה שמהותי לנפש מבלי לעסוק בפסיכולוגיה אינטרוספקטיבית או בנוירו-פיזיולוגיה של המוח.

אולם המשוואה "היחס בין הנפש למוח הוא כמו היחס בין תוכנה לחומרה" קורסת בכמה נקודות, ביניהן שלוש הללו:

ראשית, ההבחנה בין תוכנה לבין מימושה מוליכה לכך שלאותה תוכנה עשויים להיות כל מיני מימושים מטורפים שאין בהם שום צורה של מכוונות. ווייצנבאום (Weizenbaum, 1976), פרק 2, למשל, מראה בפרוטרוט איך לבנות מחשב מגליל נייר טואלט וערימת אבנים קטנות. בדומה לכך, את תוכנת הבנת הסיפור הסיני אפשר לתכנת לסדרה של צינורות מים, מערך מכוונות רוח, או דובר אנגלית חד-לשוני – ואף אחד מאלה לא ירכוש באמצעות תכנות זה כל הבנה בסינית. מלכתחילה, אבנים, נייר טואלט, רוח וצינורות מים אינם החומר הנכון למכוונות – רק משהו שיש בו אותם כוחות סיבתיים כמו שיש למוח יכול להיות בעל מכוונות. ואף על פי שדובר האנגלית בנוי מהחומר המתאים לצורך מכוונות, אפשר להבחין בקלות ששינון התוכנה לא יקנה לו כל מכוונות נוספת, מאחר ששינון התוכנה לא ילמד אותו סינית.

שנית, התוכנה פורמלית לחלוטין, אולם המצבים בעלי המכוונות אינם פורמליים במובן זה. הם מוגדרים במונחי תכניהם, לא במונחים צורניים. ההאמנה שיורד גשם, למשל, אינה מוגדרת כצורה פורמלית מסוימת, אלא כתוכן נפשי מסוים שיש לו תנאי סיפוק, כיוון התאמה (ראו Searle, 1979), וכדומה. ואמנם, להאמנה כשלעצמה אין בכלל מבנה פורמלי במובן התחבירי הזה, מאחר שלאותה האמנה עצמה עשוי להיות מספר בלתי מוגדר של ביטויים תחביריים שונים במערכות לשוניות שונות.

שלישית, כפי שהזכרתי לעיל, מצבים ואירועים נפשיים הם, פשוטו כמשמעו, תולדה של פעילות המוח, אולם התוכנה אינה, במובן זה, תולדה של המחשב.

"ובכן, אם התוכנות אינן בשום מובן שהוא מרכיב מכונן של תהליכים נפשיים, מדוע כל כך הרבה אנשים סברו ההיפך? לכל הפחות, דבר זה מצריך הסבר מסוים."

אינני יודע מה התשובה לשאלה הזו. הרעיון שסימולציית מחשב יכולה להיות הדבר האמיתי היה אמור לעורר חשד מלכתחילה, מפני שהמחשב אינו מוגבל בשום דרך שהיא לסימולציות של פעולות נפשיות. אף אחד לא יניח שסימולציית מחשב של שריפת-ענק תשרוף כליל את השכונה או שסימולציית מחשב של סופת גשם תרטיב את כולנו כדבעי. מדוע בכלל שמישהו יחשוב שסימולציית מחשב של הבנה באמת תבין? לפעמים אומרים שיהיה קשה מאוד לגרום למחשבים להרגיש כאב או להתאהב, אבל אהבה וכאב אינם קלים או קשים יותר מאשר הכרה או כל דבר אחר. לצורך סימולציה, כל שאנו זקוקים לו הוא קלט ופלט נכונים ותוכנה ביניהם, אשר ממירה את הקלט לפלט. זה כל מה שיש במחשב לצורך כל מה שהוא עושה. הבלבול בין סימולציה לבין שכפול הוא אותה טעות, בין אם מדובר בכאב, באהבה, בהכרה, בשריפה או בסופת גשם.

ועדיין, קיימות מספר סיבות לכך שב"מ בוודאי נראתה בעבר – ובעבור אנשים רבים עדיין נראית – כמשכפלת בדרך כלשהי תופעות נפשיות ולכן גם מסבירה אותן, ודומני שלא נצליח לסלק אשליות אלו עד שנחשוף לחלוטין את הסיבות לקיומן.

ראשית, ואולי חשוב מכול, קיים בלבול באשר למושג של "עיבוד מידע": אנשים רבים במדעים הקוגניטיביים מאמינים שהמוח האנושי, על הנפש שלו, מבצע משהו הקרוי "עיבוד מידע", ובאופן אנלוגי, המחשב עם תוכנותיו מבצע עיבוד מידע; אולם שריפות וסופות גשם, מאידך גיסא, אינן מבצעות כל עיבוד מידע. לכן, אף על פי שהמחשב יכול לחקות את המאפיינים הפורמליים של כל תהליך שהוא, מתקיים יחס מיוחד בינו לבין הנפש והמוח, כי כאשר המחשב מתוכנת כראוי – אידיאלית, עם אותה תוכנה כמו זו שבמוח – עיבוד המידע זהה בשני המקרים, ועיבוד מידע זה הוא בעצם מהות היסוד הנפשי. אלא שהבעיה עם טיעון זה היא שהוא מבוסס על דו-משמעות של המושג "מידע". במובן שבו אנשים "מעבדים מידע" כאשר הם מהרהרים, למשל, בבעיות באריתמטיקה או כשהם קוראים שאלות המתייחסות לסיפורים ועונים עליהן, המחשב המתוכנת אינו "מעבד מידע" – הוא מבצע מניפולציה בסמלים פורמליים. העובדה שהמתכנת והמפרש של פלט המחשב משתמשים בסמלים המייצגים אובייקטים בעולם מצויה לחלוטין מחוץ לטווח המחשב. למחשב, אם לחזור על עצמנו, יש תחביר אבל לא סמנטיקה. ולכן, אם אני מקליד במחשב: "2 ועוד 2 שווה כמה?" הוא יענה "4". אבל אין לו שמץ מושג ש"4" פירושו 4 או שיש ל"4" משמעות כלשהי. והנקודה כאן אינה שחסר לו איזה מידע מסדר שני בנוגע לפשרם של הסמלים מהסדר הראשון, אלא שמבחינת המחשב לסמלים מהסדר הראשון אין פשר. כל מה שיש למחשב הם עוד סמלים. לפיכך, הצגת המושג של "עיבוד מידע" יוצרת דילמה: או שנפרש את המושג של "עיבוד מידע" בצורה כזו שמכוונת תשתמע כחלק מהתהליך, או שלא. במקרה הראשון, המחשב המתוכנת אינו מבצע עיבוד מידע, אלא רק מניפולציה של סמלים פורמליים. במקרה השני, אף על פי שהמחשב מבצע עיבוד מידע, הוא עושה זאת רק במובן שבו מכוונת חיבור, מכוונת כתיבה, קיבוצ, וסתי חום, סופות גשם והוריקנים מבצעים עיבוד מידע; קרי, יש רמת תיאור, שבה אנו יכולים לתאר אותם כקולטים מידע בצד אחד, ממירים אותו, ומייצרים מידע כפלט. אולם במקרה זה יש צורך במשקיף חיצוני אשר יפרש את הקלט והפלט כמידע במובנו הרגיל. ולפיכך לא בוססה שום דומות בין המחשב לבין המוח במושגים של עיבוד מידע.

שנית, בחלק גדול מהעיסוק בב"מ קיימים שרידים של ביהיוריזם או אופרציונליזם. מאחר שמחשבים המתוכנתים כראוי יכולים להיות בעלי דפוסי קלט-פלט הדומים לאלו של בני אדם, אנחנו מתפתים להניח שבמחשבים קיימים מצבים נפשיים הדומים למצבים נפשיים אנושיים. אולם ברגע שאנו רואים שאפשר – מבחינה מושגית ואמפירית כאחד – שמערכת תהיה בעלת כישורים אנושיים בתחום מסוים, אך ללא כל מכוונות, עלינו להיות מסוגלים להתגבר על הדחף הזה. למכוונת החישוב שעל שולחני יש יכולת לבצע חישובים, אולם אין לה מכוונות, ובמאמר זה ניסיתי להראות שמערכת יכולה להיות בעלת יכולות קלט-פלט אשר משכפלות את אלו של דובר סינית מלידה, ועדיין לא להבין סינית, ללא קשר לאופן התכנות שלה. מבחן טיורינג אופייני לאותה מסורת, בהיותו ביהיוריסטי ואופרציונליסטי בריש גלי,



ואני מאמין שאם העוסקים בב"מ ידחו מכול וכול את הביהביריזם והאופרציונליזם, אזי חלק גדול מהבלבול בין סימולציה ושכפול יעלם.

שלישית, שרידים אלו של אופרציונליזם חוברים למעין שריד של דואליזם; אכן, ב"מ חזקה היא בעלת מובן רק על בסיס הנחת מוצא דואליסטית, לפיה בכל הקשור לנפש אין למוח כל חשיבות. בב"מ חזקה (כמו גם בפונקציונליזם) מה שמשנה הוא התוכנות, ותוכנות הן בלתי תלויות במימושן במחשבים; אכן, ככל שהדברים נוגעים לב"מ, אפשר לממש אותה תוכנה במכונה אלקטרונית, בעצם נפשי קרטזיאני, או ברוח העולם ההגליאנית. הגילוי המפתיע ביותר שבו נוכחתי במהלך הדיון בסוגיות אלו הוא שרבים מהעוסקים בב"מ מזועזעים למדי מהרעיון שלי שתופעות נפשיות אנושיות ממשיות עשויות להיות תלויות בתוכנות הפיזיקליות-כימיות הממשיות של מוחות אנושיים ממשיים. אולם אם נקדיש לעניין זה מחשבה קצרה, נוכל לראות שלא הייתי צריך להיות מופתע; מפני שלב"מ חזקה אין שום סיכוי שבעולם, אלא במסגרת דואליסטית כזו או אחרת. הפרוייקט נועד לשכפל ולהסביר את היסוד הנפשי באמצעות תכנון תוכנות, אולם אפשר לבצע את הפרוייקט הזה אך ורק בתנאי שהנפש אינה תלויה במוח תלות מושגית או אמפירית, מפני שהתוכנה לחלוטין אינה תלויה במימוש כלשהו שלה. רק מי שמאמין שהנפש נפרדת מהמוח הן מושגית והן אמפירית – גרסה חזקה של דואליזם – יכול לקוות לשכפל את היסוד הנפשי באמצעות כתיבה והרצה של תוכנות, מאחר שתוכנות חייבות להיות בלתי תלויות במוח או בכל צורה אחרת של מימוש. אם פעולות נפשיות מורכבות מפעולות חישוב של סמלים פורמליים, נובע שאין להן שום קשר מעניין למוח; הקשר היחיד יהיה שבדרך מקרה המוח הוא מכונה אחת מני אינספור מכונות המסוגלות לממש את התוכנה. גרסה זו של דואליזם אינה הגרסה הקרטזיאנית המסורתית שלפיה ישנם שני סוגי **עצמים**, אולם היא קרטזיאנית במובן זה שהיא עומדת על כך שלדבר שבעטיו הנפש היא נפשית אין שום קשר מהותי עם התוכנות הממשיות של המוח. הנחת רקע דואליסטית זו מוסתרת מעינינו באמצעות העובדה שבספרות הב"מ רווחים גינויים חריפים לכול צורה של "דואליזם"; אלא שמחברי גינויים אלו אינם ערים, כמדומה, לכך שעמדתם קדם-מניחה גרסה חזקה של דואליזם.

"האם מכונה יכולה לחשוב?" להשקפתי, רק מכונה יכולה לחשוב; אדרבא – רק מכונות מסוג מיוחד מאוד, קרי מוחות ומכונות שיש להן אותם סיבתיים כמו שיש למוח. וזה הטעם העיקרי שבעטיו לב"מ חזקה כמעט ואין מה לומר לנו על חשיבה, מאחר שהיא אינה אומרת לנו דבר על מכונות. מעצם הגדרתה, היא עוסקת בתוכנות, ותוכנות אינן מכונות. מכונות – נוסף לכל המאפיינים האחרים שיש לה – היא תופעה ביולוגית, וסביר להניח שהיא תלויה מבחינה סיבתית בביוכימיה של מקורותיה, כמו שיצירת חלב, פוטוסינתזה או כל תופעה ביולוגית אחרת תלויות בביוכימיה של מקורותיהן הסיבתיים. איש לא יניח שאפשר לייצר חלב וסוכר באמצעות הרצת סימולציית מחשב של רצף התהליכים הפורמליים של ייצור חלב ופוטוסינתזה, אולם כשמדובר בנפש מוכנים רבים להאמין בנס מסוג זה, בגלל דואליזם עמוק ותמידי: הנפש שהם מניחים היא עניין של תהליכים פורמליים ואינה תלויה בסיבות חומריות ספציפיות, בשונה, למשל, מחלב ומסוכר.

להגנת דואליזם זה הם מבטאים לעתים את התקווה שהמוח הוא מחשב דיגיטלי (המחשבים המוקדמים, דרך אגב, נקראו 'מוחות אלקטרוניים'). אולם אין בכך כדי לסייע בידם. מובן שהמוח הוא מחשב דיגיטלי. מאחר שכל דבר הוא מחשב דיגיטלי, גם המוח הוא כזה. הנקודה היא שהיכולת הסיבתית של המוח לייצר מכוונות אינה יכולה להיות מורכבת מהעובדה שהוא מממש תוכנות מחשב, מאחר שבעבור כל תוכנה שהיא אפשר למצוא מכשיר שיממש אותה אף על פי שיהיה נטול מצבים נפשיים כלשהם. יהא אשר יהא הדבר שהמוח עושה כדי לייצר מכוונות, דבר זה לא יכול להיות מימוש של תוכנה, מאחר ששום תוכנה, כשלעצמה, אינה מספיקה למכוונות.

## ביבליוגרפיה

- Fodor, J.A. (1980) Methodological solipsism considered as a research strategy in cognitive psychology. *The Behavioral and Brain Sciences* 3:1.
- McCarthy, J.. (1979) Ascribing mental qualities to machines. In: *Philosophical perspectives in artificial intelligence*, ed. M. Ringle. Atlantic Highlands, N.J.: Humanities Press.
- Newell, A. (1979) Physical symbol systems. Lecture at the La Jolla Conference on Cognitive Science.
- Newell, A. & Simon, H.A. (1963) GPS: a program that simulates human thought. In: *Computers and thought*, ed. A. Feigenbaum & V. Feldman, pp. 279–93. New York: McGraw Hill.
- Pylyshyn, Z.W. (1980) Computation and cognition: issues in the foundations of cognitive science. *Behavioral and Brain Sciences* 3.
- Schank, R.C. & Abelson, R.P. (1977) *Scripts, plans, goals, and understanding*. Hillsdale, N.J.: Lawrence Erlbaum Press.
- Searle, J.R. (1979) What is an intentional state? *Mind* 88:74–92.
- Weizenbaum, J. (1965) Eliza – a computer program for the study of natural language communication between man and machine. *Communication of the Association for Computing Machinery* 9:36–45.
- (1976) *Computer power and human reason*. San Francisco: W.H. Freeman.
- Winograd, T. (1973) A procedural model of language understanding. In: *Computer models of thought and language*, ed. R. Schank & K. Colby. San Francisco: W.H. Freeman.

## הכרת תודה

אני חב תודות למספר רב של אנשים על שדנו עמי בנושאים אלה ועל נסיונותיהם הסבלניים להתגבר על בורותי בתחום הבינה המלאכותית. במיוחד ברצוני להודות לנד בלוק (Ned Block), יוברט דרייפוס (Hubert Dreyfus), ג'ון הוגלנד (Jhon Haugeland), רוג'ר שנק (Roger Schank), רוברט ווילנסקי (Robert Wilensky) וטרי ווינוגרד (Terry Winograd).