



פרטיות הצרכן ותובנות שיווקיות על נתוני המיקום של מכשירים ניידים: מצב של Win-Lose?



פיטר פאל זובצ'ק

ד"ר פיטר פאל זובצ'ק הוא חבר סגל בפקולטה לניהול ע"ש קולר באוניברסיטת תל אביב. את הדוקטורט סיים ב-INSEAD וכיום הוא חוקר ומרצה בתחום השיווק. במחקרו הוא מתמקד בתפקידן של רשתות חברתיות באינטראקציות עם הצרכן, באפליקציות לניהול קשרי לקוחות, בהפצה של מוצרים חדשים ובחקר קהילות. תחומי העניין שלו כוללים פרסום במכשירים ניידים ואסטרטגיה תחרותית. עבודותיו הופיעו ב-Journal of Marketing, Journal of Marketing Research, Journal of Personality and Social Psychology, Quantitative Marketing and Economics, Journal of Interactive Marketing and Social Networks.

תקציר

נתוני המיקום של מכשירים ניידים מציגים מידע מפורט על המיקום של צרכנים, וכך מספקים הזדמנות למשווקים להתאים לצרכנים הצעות רלוונטיות לפי הקשר המיקום. אולם נתוני המיקום יכולים לחשוף גם מידע רגיש על הצרכנים המבוסס על המקומות שבהם הם נמצאים. למשל, מתקני בריאות או מקומות הקשורים לדת. האם ניתן לספק בו-בזמן את רצון המשווקים לקבל תובנות מעשיות ואת ההעדפה של האנשים לפרטיות? באמצעות מידע שקיבלנו מספק של נתוני מיקום ברמת המכשיר, בדקנו את המידה שבה קיבוץ אנשים לפלחים הומוגניים המאפשר רמה גבוהה יותר של פרטיות, משפיע על תחזית הביצוע. מצאנו שרמה מסוימת של קיבוץ לפלחים מביאה לדיוק גבוה יותר בניבוי. יתרה מכך, הסתמכות על מיקומים של פעילות מסחרית מניבה תוצאות טובות לפחות כמו הסתמכות על מיקומי בתים. תובנות אלו מספקות הכוונה לספקי נתונים שצרכים לאזן בין שירות ללקוחות לבין ציפיות הצרכנים לפרטיות, וגם מספקות לרגולטורים תובנה על פירוט הנתונים הדרושים למשווקים לשם פעילותם.

תומכים באיסור מוחלט על פרסום ממוקד (Edelman 2020), תמיכה שהיא תגובת יתר בהתחשב בכך שרוב הצרכנים בעולם מוכנים לקבל סיכונים מסוימים לפרטיות הדיגיטלית שלהם מטעמי נוחות (Fisher 2019). אחרים טענו שניהול פרטיות הצרכן יכול להיות הזדמנות לבדל את חוויית המותג שלהן (Goldfarb and Tucker 2013). עם זאת, ההתפתחות האחרונה של נופך הפרסום הדיגיטלי מצביעה על כך שייכתן שהזדמנות זו הוערכה ביתר. ואכן, המניע החזק ביותר לשינוי התנהגות החברות היה כניסת התקנות להגנת מידע כללי של האיחוד האירופי (GDPR), המחייבות אתרים לקבל הסכמה מפורשת מהצרכנים לקובצי העוגיות שנאספו או שותפו, למעט אלו החיוניים הדרושים לפונקציונליות של האתר.¹ הביקורת המרכזית על ה-GDPR הייתה האכיפה החסרה, אך עדיין הייתה לו השפעה חזקה מאוד על הדיון העולמי בנושא פרטיות דיגיטלית של צרכנים (Joseph 2023; Skiera et al. 2022).

איזון יעדים מתחרים?

הן החשש לפרטיות והן הפתרונות המוצעים לניצול הנתונים של מיקומי הצרכנים מזכירים את בעיית הפרטיות בפרסום מקוון. לאור התעוררות החשש לפרטיות ובשל לחץ מוגבר מצד מחוקקים אירופאים ואמריקאים, החלו חברות טכנולוגיה לנקוט עמדה בכל הנוגע לפרטיות הצרכנים (Snider 2021). עם זאת, בכל הנוגע לניצול של נתוני מיקום למטרות מעקב אחר מגע בין אנשים במהלך מגפת הקורונה, אומנם היו כאלו שהביעו דאגה מאובדן הפרטיות (למשל, Bengio et al. 2020), אך רוב הצרכנים היו מוכנים לוותר על הפרטיות בתמורה לתועלת חברתית (Ghose et al. 2022). הדבר עולה בקנה אחד עם סקרים שבוצעו לאחרונה, שלפיהם הערך שצרכנים מייחסים למודעות פרסומית מותאמת אישית ורלוונטיות יכול לגבור על חששם מפני השימוש של משווקים בנתוני המיקום שלהם (Arora et al. 2021; Verhagen et al. 2022).

לאור הזמינות של נתוני מיקום של מכשירים ניידים ותשומת הלב הגוברת הנובעת מכך לבעיית הפרטיות, עולה השאלה כיצד אפשר לאזן בין האינטרסים של המשווקים לבין הרצון של הצרכנים לשמור על פרטיותם. בהקשר של פרסום במכשיר

האימוץ הנרחב של טכנולוגיות ניידות מאפשר לקמעונאים לקבל את נתוני המיקום של הצרכנים, וכך מאפשר דיוק גדול יותר התומך בשיווק ממוקד. נתוני המיקום בזמן אמת תומכים בשליחת מבצעים למכשיר הנייד כאשר הצרכנים נמצאים בקרבת מוקד קמעונאי, ומסלולי המיקום של אנשים נותחו כדי לחדד את הדיוק של חיזוי תגובת הצרכן לטקטיקות שיווקיות כמו פרסום או תמחור (Luo, Andrews, Fang and Phang 2014; Shoshani, Zubcsek, and Reichman 2024).

אולם למרות היתרונות הפוטנציאליים של שימוש בנתוני המיקום של מכשירים ניידים, מתעורר החשש לפגיעה בפרטיות הצרכן הנגרמת מאיסוף וניתוח של נתונים אלו. למשל, בהתבסס על דפוס המיקומים שבהם ביקר מכשיר אחד מתוך יותר מ-10 מיליון, הסיקו Thompson & Warzel (2019a) כי המכשיר שייך למישהו שנסע עם נשיא ארה"ב דונלד טראמפ, מה שמרמז כי האימונים שמציב המידע על מיקום המכשירים הניידים עשויים להגיע גם לביטחון הלאומי. נתוני המיקום של מכשירים ניידים אפשרו גם זיהוי של אנשים שהשתתפו במהומה ב-6 בינואר 2021 בבניין הקפיטול בארה"ב (Warzel and Thompson 2021).

הפרטיות הדיגיטלית של הצרכן: המקרה המתמשך של פרסום מקוון

החששות של צרכנים לגבי הפרטיות הדיגיטלית נחקרו כבר יותר מעשור (Goldfarb and Tucker 2012; Tucker 2014), אך לראשונה הם נחשפים בהקשר של קובצי עוגיות (cookies) של צד שלישי בפרסום דיגיטלי. קובצי העוגיות של צד שלישי מאפשרים למפרסמים ולסוחרים נתונים לעקוב אחר משתמשים באתרי אינטרנט מרובים ולבנות פרופילים מפורטים של הפעילויות, ההעדפות וההתנהגויות המקוונות שלהם. המעקב הזה יכול להיות חוזרני מפני שהוא עשוי להתרחב מעבר לאתר אינטרנט בודד, ויכול לאפשר לא רק פרסום ממוקד אלא גם לחשוף מידע רגיש, כמו ביקור באתרים של ארגונים דתיים או של ספקי שירותי בריאות.

הפתרונות המוצעים להתמודדות עם איום הפרטיות בהקשר של פרסום דיגיטלי נפרסים על פני קשת רחבה. חלקם

¹ בעוד שחוק פרטיות הצרכן של קליפורניה (CCPA) אינו מחייב את הסכמתם המפורשת של צרכנים לאיסוף הנתונים האישיים שלהם באמצעות קובצי עוגיות, הוא דומה ל-GDPR של האיחוד האירופי בכך שגם הוא רואה בנתונים הנאספים על ידי קובצי עוגיות כמידע אישי.

ארבעה מיקומים בציון זמן (מגדלים סולריים ברשת תקשורת סולרית במדינה קטנה) כדי לזהות 95% מהאנשים. כדי להתגבר על כך, צברנו בנתונים שלנו ביקורים במיקום לפי שבוע ולפי מותג (כלומר על פני חנויות שונות של אותו מותג).

בסוג השני של הצבירה קיבצנו מכשירים דומים לפלחים הומוגניים, מתוך מחשבה שכך יתאפשר למשווקים לזהות פלחים רצויים לטרגוט, ללא יכולת לקשר מכשיר מסוים לאדם ספציפי. נציין כי בתגובה לחקיקת ה-GDPR, גוגל בוחנת אמצעים דומים כדי לשמור על פרטיות הצרכן ועדיין לאפשר פרסום ממוקד (Bohn 2021; Mehta 2023).

שקלנו דרכים שונות לקיבוץ הפרטים לפלחים הומוגניים בגדלים שונים. היסטוריית מיקומים מפורטת ברמת הפרט עשויה לחזות טוב יותר התנהגות עתידית, אך פירוט מה מסכן את פרטיות הצרכן. לעומת זאת, צבירה של צרכנים ל"פלחים" הטרוגניים גדולים עלולה לפגוע בערך הניבוי של הנתונים. הניתוח שלנו מציג את ההשפעה של חלוקת צרכנים לפלחים בגדלים שונים, ומאפשר לנו להעריך את התחליפיות (טרייד-אוף) בין פרטיות הצרכן לבין דיוק הניבוי.

אומנם לא הערכנו את הרגישות של כל מותג או קטגוריית מותג כדי להסיר את אלו המציבים את הסיכון הגדול ביותר לפרטיות הצרכנים, אך הנחנו כי צוברי הנתונים (המוצגים באיור 1) מעבירים נתונים (מצטברים) רק על ביקורים באזורי עסקים. יישום של הגבלה זו הוא פשוט ומעניק הגנה רבה יותר על כתובות הבתים של אנשים.

מסגרת אמפירית

את הנתונים עבור המחקר שלנו השגנו מחברה המתמחה באיסוף ובניתוח של נתוני מיקום באמצעות רשת של אפליקציות סולריות. הנתונים מכילים פעילות במכשירים ניידים במדינת ניוז'היב בארה"ב בתקופה של עשרה שבועות, החל מינואר 2020 ועד תחילת מרץ 2020. ההתמקדות בפרק זמן זה אפשרה לנו להימנע מהשפעה פוטנציאלית של מגפת הקורונה, שכן בסוף מרץ 2020 חוקקו ברחבי המדינה תקנות המשפיעות על תנועת הצרכנים.

הנתונים נאספו מאפליקציות של מכשירים ניידים על בסיס ההרשאות שקבע המשתמש במכשיר. כל רשומה מכילה את

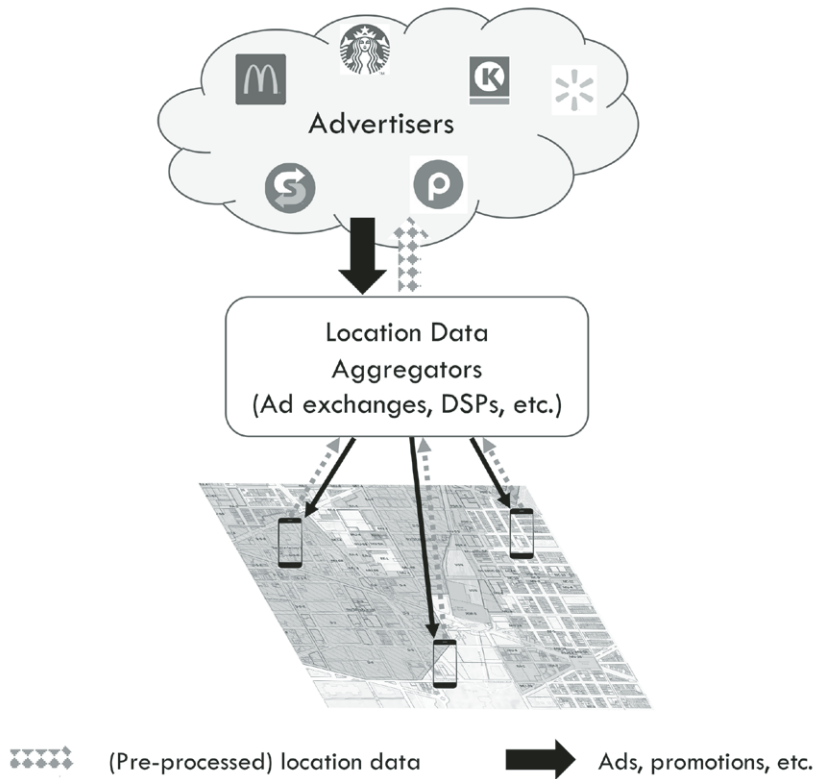
הנייד, הראו (Rafieian -I Yogaranasimhan 2021) שככל שרמת המיקוד גדלה והפרטיות פוחתת, כך המפרסמים מרוויחים יותר. הדבר מצביע על כך שפרטיות הצרכן והאינטרס של המשווקים עומדים בניגוד עניינים. בנייר עבודה שנכתב לאחרונה (Shoshani et al. 2022), בדקנו אני ועמיתיי את השאלה אם הצורך של צרכנים בפרטיות והיכולת של משווקים להפיק תובנות בעלות יכולת ניבוי מנתוני המיקום של מכשירים ניידים עומדים בניגוד זה לזה, ואם כן, מהן האפשרויות העומדות בפני הרגולטורים.

חשבו על מותג שרוצה לטרנט צרכנים בפרסום. הוא יכול לשלב שירותים מבוססי מיקום באפליקציה משלו למכשיר הנייד, מה שיאפשר לו לטרנט את הצרכנים שהורידו את האפליקציה – אך ייתכן שהוא מעוניין להגיע לקהל רחב יותר. לשם כך הוא עשוי להחליט לזהות ולטרנט את הלקוחות הפוטנציאלים עבורו (וייתכן שגם לקוחות של מתחרים וצרכנים שהם חדשים בשוק המותג) באמצעות שימוש בשירותיו של ספק נתונים של צד שלישי. נתונים אלו עשויים לאפשר למשווק לשקול טרגוט בפרסום לשוק רחב בהרבה. איור 1 מציג את המסגרת המאפשרת למותג לעסוק בשיווק ממוקד, תוך שימוש בנתוני מיקום המעוררים חשש לפרטיות מפני שהם יכולים לאפשר זיהוי בתים של אנשים או מיקומים שבהם הם ביקרו ואולי אינם רוצים שאחרים ידעו על כך, כמו שירותי בריאות או מקומות הקשורים לדת (Macha et al. 2023). האם אפשר להשיג את מטרת המותג לזהות את הלקוחות הפוטנציאלים שלו לטרנט בלי ליצור סיכון לזיהוי אנשים מתוך הנתונים? זו השאלה הבסיסית שהניעה את המחקר שלנו.

הפחתת הסיכונים לפרטיות בפרסום מבוסס מיקום במכשירים ניידים

בעבודה זו בחנו כמה אמצעים שבעזרתם מותגים וספקי נתונים יכולים להפחית את הסיכון לפרטיות הנובע מבחירה של קהלי פרסום על בסיס מיקום. המפתח לכך היה בחינת שני סוגים של צבירת נתונים. הסוג הראשון של צבירת נתונים שימש כדי להפחית במידה ניכרת את ממדיות הנתונים. יותר מאשר היסטוריית נלישה עשויה לזהות אנשים, נתוני המיקום בציון הזמן מציבים את הצרכנים במרחב בעל ממדיות גבוהה במיוחד, והופכים את זיהוי המכשיר האנונימי למשימה קלה יחסית. (De Montjoye et al. 2013) הראו שמספיקים

איור 1: מודל של אקוסיסטם לפרסום המבוסס על מיקום. הנתונים הנאספים מהמכשירים הניידים של הצרכנים נצברים על ידי מתווך (צובר של נתוני מיקום, למשל, פלטפורמה בצד הביקוש), וכך יכולים המפרסמים להתאים אישית את קהל הפרסום שלהם באמצעות שימוש בנתוני המיקום המעובדים (בלבד).



של התנהגות העבר. לבסוף, ארבעת השבועות האחרונים של הנתונים היוו את שלב ה"בדיקה". כאן השתמשנו בנתוני הביקורים במותג ברמת הפלחים כתשומות, והערכנו את דיוק הניבוי של המודלים שלנו. השווינו את הביצועים, המוערכים באמצעות מדידת השטח שמתחת לעקומה (AUC), על פני אופנים שונים של יצירת הפלחים מהנתונים (בשלב הקיבוץ), עבור חלוקות בגודל שונה של צרכנים.

כדי למנוע ספירה כפולה, מיוזגו תצפיות עוקבות באותו מיקום לתצפית אחת לכל מכשיר. התייחסנו לכל תצפית "ממותגת" במערך הנתונים שנוצר כביקורים במותג התואם. הגבלנו את ההתייחסות לביתוח שלנו לצרכנים שביקרו לפחות במיקום אחד מבין 200 המותגים הפופולריים ביותר בנתונים שלנו. במהלך כל שבוע בשלב קיבוץ הנתונים של הניתוח שלנו. פעולה זו הגבילה את המדגם שלנו לכ-9 מיליון ביקורים במותג (שנצפו ביותר מ-1.6 מיליארד תצפיות גולמיות) על ידי 180,674 מכשירים.

אמצעי הזיהוי ומיקום ה-GPS של המכשיר הנצפה, יחד עם ציון הזמן של התצפית. כמו כן, בתצפיות הנמצאות בסביבה הקרובה של קמעונאים, הרשומה כוללת את המותג שהמכשיר היה הקרוב ביותר אליו.

במחקר שלנו חיקינו את המשימות של צובר הנתונים באקוסיסטם הפרסום המאפשר את המידע על המיקום כמו באיור 1: עבור כל קיבוץ של מכשירים לקבוצות הומוגניות שונות זו מזו, הערכנו את יכולתם של המשווקים לחזות במדויק את הביקורים העתידיים של הצרכנים במותג, ובמקביל את רמת הפרטיות שמאפשרת החלוקה הנתונה. לשם כך חילקנו את הנתונים לשלושה שלבים עוקבים. בשבועיים הראשונים של הנתונים – שלב ה"קיבוץ" – השתמשנו כדי לחלק את המכשירים לפלחים על בסיס הקריטריונים המתוארים להלן. שבועות 3-6 של הנתונים היו שלב ה"אימוץ". נתוני הביקור במותג בתקופה זו שימשו להכשרת מודל המנבא את ביקורי המותג העתידיים של כל מכשיר על סמך צבירה ברמת הפלחים

איור 2: האיור ממחיש כיצד המשקל המוטל על מיקום הבית משנה את הרכב הפלחים. בכל פאנל חילקנו את 18 אלף המכשירים במחקר שלנו לעשרה פלחים גדולים באותה מידה, שאותם אנו מציגים כנקודות על אותה סקאלת צבעים התואמת למיקומי הבית המשוערים שלהם. באיור מצד שמאל השתמשנו ב- $w=1$, (כלומר מידע על ביקור במותג בלבד), בפאנל האמצעי $w=0.5$, ומצד ימין $w=0.2$. הפחתת המשקל המוטלת על תכונות הביקור במותג (לעומת מיקום הבית) מחלקת את המכשירים לפלחים סמוכים יותר זה לזה.



קיבוץ פלח

במספר הקבוצות שיווצרו), עבור כל המחלקים K של N , וכך הבטחנו כי כל הקבוצות שיתקבלו יהיו בעלות אותו גודל.

בדקנו שלוש גישות לפילוח המכשירים הניידים, תוך שימוש בנתונים שנאספו בשלב הקיבוץ. ראשית, קיבצנו את האנשים על סמך המותגים שבהם ביקרו בעבר. בחרנו את 25 המותגים המובילים (אנו מציינים סט זה באות B), וביצענו את האלגוריתם של K-means תחת אילוץ על מרחב התכונה שהוגדר על ידי תדירות הביקור של יחידים במותגים מסוג B במהלך שלב הקיבוץ. בגישה השנייה עשינו שימוש במיקום ה"בית" של המכשיר, שאותו הסקנו מהמיקום הנפוץ ביותר בשעות הלילה (8PM-4AM) (Bettini et al. 2005). הכללנו את קואורדינטות ה-GPS של מיקום הבית במרחב התכונות הבסיסי, וכך קיבצנו את האנשים על סמך הדמיון בהתנהגויות הביקור במותג שלהם ו/או הקרבה של בתיהם זה לזה. כדי לקחת בחשבון סקאלה שונה של קואורדינטות מיקום הבית תוך שמירה על הפרופורציות של ספירת הביקורים במותג, ערכנו איטרציות על פני טווח של משקלים שונים של w (בין 0 ל-1) שהצבנו בנתוני הביקור במותג ומיקום הבית בהתאמה. איור 2 מדגים את התוצאה של תהליך זה עבור מספר נמוך של (עשרה) פלחים. כאשר $w=1$, המכשירים מקובצים רק על סמך התנהגות הביקור במותג שלהם. ככל ש- w פוחת, הפלחים המתקבלים הופכים בהדרגה סמוכים יותר. הגדרה של $w=0$ (לא מוצגת באיור 2) תבצע פילוח של K-means

כדי לחקור את הטרייד-אוף הפוטנציאלי בין יכולת הניבוי של משווקים לבין פרטיות הצרכן, הגדרנו את הפרטיות במונחים של גודל הפלח המינימלי $z > 0$. הגדרה זו היא לצורך מדידה ותואמת את המושג של k -אנונימיות (Sweeney 2002), שלפיו לא ניתן להבחין בין הפרט לבין לפחות $k-1$ אנשים אחרים בנתונים. שינינו את גודל הפלחים (לפי התפיסה שבה הגדלה של פלח z מגדילה את רמת הפרטיות של הצרכן המתקבלת בחלוקה הנתונה), והערכנו את תחזית הביצוע של המסווגים שאימנו קודם לכן כפונקציה של רמת הפרטיות המתאפשרת ברזולוציה נתונה. כדי ליצור קבוצות בגודל מאוזן, השתמשנו באלגוריתם של K-means תחת אילוץ (Bradley et al. 2000), המאפשר לנו להגביל את גודל הקבוצה המינימלי לכל $0 \leq z \leq N$ במדגם של N מכשירים. זהו יתרון חשוב על פי שיטות פילוח חלופיות, כמו (H)DBSCAN (=Density-Based Spatial Clustering of Applications, Ester et al. 1996; Li and Xi 2011), שעשויות ליצור קבוצות המכילות אפילו מכשיר אחד בלבד. קבוצות קטנות מאוד יכולות לאפשר זיהוי של אנשים ספציפיים, ומכך ניסינו להימנע.

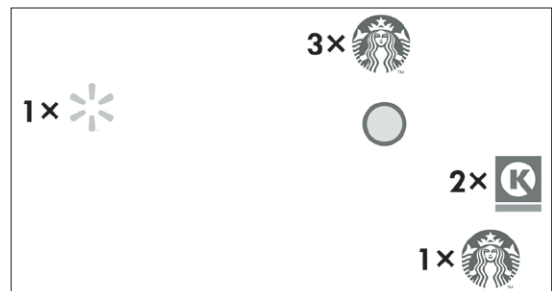
משכנו מדגם אקראי של $N=18,000$ מכשירים מהמדגם המלא שלנו שתואר קודם לכן, והגבלנו את גודל הקבוצה המינימלי z להיות שווה ל- N/K (כלומר מספר המכשירים במדגם מחולק

תחת אילוץ על הקואורדינטות של מיקום הבית המשוער של כל מכשיר.

הגישה השנייה שלנו אפשרה להעריך אמפירית כמה תבנות ניבוי נוספות מספק השילוב של מיקום הבית מעבר לנתוני הביקור במותג, אך מגבלה עיקרית שלה היא ההסתמכות על מיקום הבית. למרות שהצבירה של מכשירים לפלחים הומוגניים מאפשרת מידה מסוימת של פרטיות, עצם השימוש בנתוני המיקום של הבית מהווה איום על פרטיות הצרכן: יש לאסוף את נתוני המיקום של הבית מכיוון שהם משמשים בסיס ליצירה של פלחים הומוגניים. בעוד שאיסוף של החברה לנתוני הביקור במותג עשוי שלא לחשוף את זהותם של אנשים, אפשר להסיק את מיקום הבית מתוך נתוני המיקום של המכשיר הנייד וכך לזהות את בעל המכשיר (Macha et al. 2023).

כדי להפיג חששות מסוג זה, הצענו גישה שלישית ליצירת פלחים באופן ששומר יותר על הפרטיות. בגישה זו השתמשנו באותו הליך המתואר לעיל, אך במקום להסיק את מיקום הלילה הנפוץ ביותר עבור כל מכשיר, הסקנו מיקום של "בית מדומה" שאותו הגדרנו כמרכז של כל נקודות העניין שזוהו (כלומר מיקומים ממותגים שאינם למגורים) שבהם נמצא המכשיר, בשקלול לפי תדירות הביקור. איור 3 מדגים את תהליך בניית המיקום של הבית המדומה.

איור 3: בניית מיקום "הבית המדומה" (העיגול החלקי) עבור מכשיר שביקר בארבע חנויות קמעונאיות בסך הכול 7 פעמים. כלומר, אחת. שימו לב כיצד נגזר מיקום הבית המדומה קרוב יותר לחנויות שבהן ביקר המכשיר בתדירות גבוהה מאשר מיקום במרכז של ארבע החנויות שהיה נגזר ללא שקלול.



בניגוד למיקום הבית המשוער שהתבסס על "המקום שבו ישנים", אנו מבססים את מיקום הבית המדומה על "המקום שבו קונים". ההבדל אולי דק, אך זוהי הבחנה מרכזית. בעוד

שקיוץ המבוסס על מיקום הבית המשוער משקף קרבה בין פרטים המתגוררים באותה שכונה, השימוש במיקום הבית המדומה הוא בסיס לגיבוש קבוצות של אנשים שלעיתים קרובות נמצאים במיקומים פיזיים סמוכים זה לזה באותו הפלח. באמצעות ההסתמכות על מעקב אחר המיקומים המסחריים ולא אחר מיקום הבית, קיווינו לקבל את אותו המידע שנתון מיקום הבית. לדוגמה, הנוחות והמשיכה לקניות במותגים מסוימים, תוך כדי הימנעות מהחשש לפרטיות הקשור לזיהוי בעלי המכשירים.

ניתוח

בשלבי האימון והבדיקה חישבנו ממוצע של ספירת הביקורים השבועיים במותג - על פני כל המכשירים בכל קבוצה, והקצינו את הממוצעים שהתקבלו לפלח שאליו משתייך המכשיר כמשתנים בלתי תלויים התואמים למכשיר זה. בעזרת נתוני האימון, עבור כל אחד מ-200 המותגים בעלי הביקורים הרבים ביותר (כלומר גם מותגים מחוץ לקבוצה B המשמשת ליצירת הפלחים), אימנו מודל שישתמש בתכונות הממוצעות האלו ברמת הקבוצה במשך שבוע t, כדי לחזות האם מכשיר נתון יבקר או לא (0 או 1) (בכל חנות של) מותג נתון בשבוע t+1. (במילים אחרות, אימנו 200 מסווגים בינאריים - אחד לכל מותג - כל אחד משתמש באותם 200 מנבאים רציפים). לפיכך, בעוד שהמנבאים ששימשו בנייתו שלנו היו זהים עבור כל האנשים בפלח נתון, המשתנה התלוי היה ההתנהגות של האדם בעל המכשיר, ולא הממוצע של הקבוצה. בחנו שימוש במגוון של שיטות למידת מכונה, כולל גרסיה לוגיסטית, מודלים מבוססי עצים ומכונות תמך וקטורי (SVM). כל אחת מהשיטות סיפקה דפוסיים דומים של תוצאות, וכאן אנו מתמקדים בתוצאות שהתקבלו בגרסיה לוגיסטית.

הערכנו את ביצועי המודל שלנו שאומן בנתוני האימון באמצעות המדגם שהתקבל בשלב הבדיקה. כלומר, עבור שבועות 7, 8 ו-9 חישבנו את ממוצע הביקורים במותג עבור כל פלח מכשירים, והשתמשנו במודל שלנו כדי לחזות את התנהגות ברמת המכשיר של הביקור בכל מותג בשבועות 8, 9 ו-10. לכל אחד מ-200 המותגים שכללנו בנייתו זה, הערכנו את דיוק הניבוי של המודל שלנו באמצעות גודל השטח שמתחת לעקומה (AUC). עבור כל אחת מגישות הפילוח שנידונו בסעיף הקודם, חישבנו ממוצע AUC על פני 200 המותגים המובילים עבור כל רמה של z בנייתו שלנו, כדי לקבוע את הטרייד-אוף

ה-AUC מהמודל המשלב נתוני ביקור במותג ומיקום הבית אינו יורד באופן מונוטוני עם z . כאשר הפלחים של המכשירים נוצרים תוך שימוש בשילוב שבין ביקור במותג ונתוני מיקום הבית, קיים טווח של ערכי z ($24 \leq z \leq 400$), שמסוגל גם לאפשר פרטיות באמצעות צבירה וגם להציג תחזיות ביצוע מעולות ביחס למודל הביקור במותג העושה שימוש בנתונים ברמת המכשיר.

”המקום שבו קונים” לעומת ”המקום שבו ישנים”

הניתוח שלעיל ממחיש כיצד אפשר להשיג ביצועים מעולים תוך שימוש בפלחים הומוגניים במקום בנתונים ברמת המכשיר. עם זאת, כפי שצוין קודם לכן, האיום על פרטיות הצרכן הנובע מאיסוף ומשימוש בנתוני מיקום הבית, מקוזז ככל הנראה את הרווחים של פרטיות המתקבלים מצבירה של המכשירים לפלחים הומוגניים. כדי לעקוף בעיה זו, נגשת הקיבוץ השלישית שלנו החליפה את מיקומי הבית במיקום הבית המדומה, הנגור כמרכז של כל ביקורי המותג במהלך שלב הקיבוץ עבור מכשיר נתון. כפי שמוצג באיור 4, בעוד שה-AUC לערכים שונים של z התואמים לגישה זו דומים מאוד לאלו שהתקבלו באמצעות מיקום הבית האמיתי של אנשים, מצאנו שקיבוץ המכשירים על בסיס מיקום הבית המדומה שלהם וספירת הביקורים במותג, שיפרו את תחזית הביצוע של המודל שלנו לכל $1 < z < N$. הניתוח שלנו מוכיח כי אין צורך לאסוף את מיקום הבית של מכשיר כדי להשיג את דיוק הניבוי הגבוה ביותר האפשרי. ההסתמכות על נתונים שנאספו ממיקומים מסחריים מאפשרת לקבץ מכשירים באופן מה שנוצרים פלחים הומוגניים הן במונחים של קרבה גיאוגרפית והן של העדפות מותג. כך אנו יכולים לוותר על איסוף ואחסון של נתוני מיקום הבית ולהפחית את החשש לפרטיות.

כדי להדגיש עוד יותר את טיב השמירה על הפרטיות בשימוש בבית מדומה, הערכנו את הדיוק שבו אפשר לזהות את מיקום הבית האמיתי של המכשיר מתוך הנתון של מיקום הבית המדומה. באיור 5 אנו מדגימים את התפלגות המרחקים שבין מיקום הבית האמיתי המשוער לבין מיקום הבית המדומה של כל מכשיר. המרחק החציוני על פני כל 180 אלף המכשירים הוא 10.71 ק"מ, גבוה בהרבה מדיוק הניבוי שנמדד על ידי Macha et al. (2023), שהעריכו רדיוס ממוצע של 2.5 מיילים (כ-4 ק"מ) כדיוק הניבוי הממוצע של מיקומי בית המתקבל

שבין רמת הפרטיות לבין דיוק הניבוי כפונקציה של גודל כל פלח. כדי לחקור את המידה שבה שילוב נתוני מיקום בשלב הקיבוץ שיפר את יכולת הניבוי של המודלים שלנו, ביצענו תרגיל זה לכל ערך w של המשקל שהוצב על תכונת ביקור המותג בשלב קיבוץ הפלח $\{0, .05, .1, \dots, 1\}$ לשתי גישות הפילוח המסתמכות על נתוני המיקום. לאחר מכן, עבור כל ערך של $z < N < 1$, חישבנו את ה-AUC המקסימלי התואם לערכים של $w \in \{0, .05, .1, \dots, 1\}$.

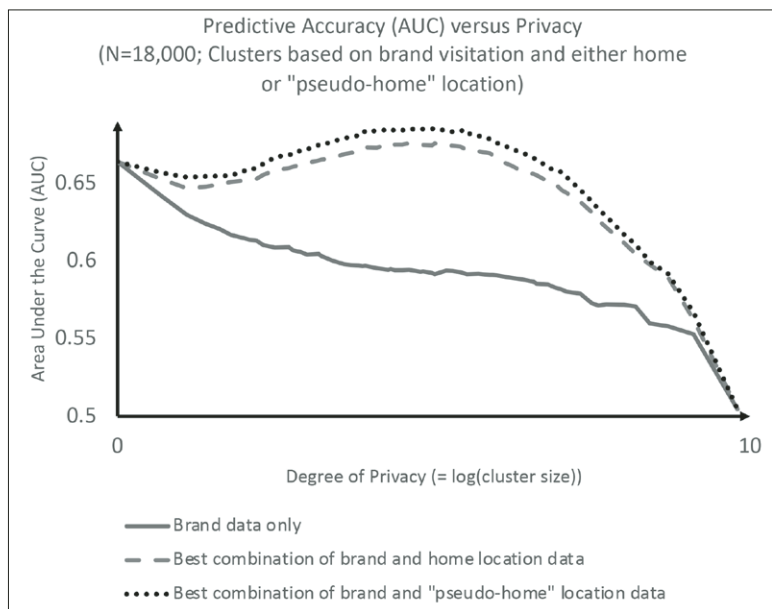
הערך הפוטנציאלי של נתוני מיקום

הדיון בתוצאות הוא במונחים של מידת הפרטיות z המתקבלת בחלוקה של המכשירים לקבוצות הומוגניות. נזכיר ש- z מקבל ערכים בין 1 ל- N , כך ש- z הוא מחלק של N , ו- N הוא מספר המכשירים במדגם. כדי לשמור על הפרטיות נדרשת צבירה של מספר גדול יותר של מכשירים ניידיים לכל מספר קטן יותר של קבוצות הטרונגיות יותר. כלומר, ככל ש- $(z-1)$ -אנונימיות גדל, מספר הפלחים $K=N/z$ פוחת. עם זאת, אפשר לצפות לכך ששימוש בערך גדול יותר של z , מביא לכוח ניבוי חלש יותר להתנהגות עתידית. לכן חברות צריכות להחליט אם העדפה של פרטיות מוגברת לצרכנים מצדיקה את הירידה בתחזית הביצוע. נותרת השאלה האמפירית האם הגדלת מספר הפלחים תשפר תמיד את תחזית הביצוע, או האם יש נקודה שבה תוספת של עוד פלחים משפיעה לרעה על תחזית הביצוע.

התוצאות של ניתוח זה מוצגות באיור 4. כצפוי, תחזית הביצוע מתדרדרת ככל שהפלח גדול יותר ומספק פרטיות מוגברת לצרכנים. כאשר מסתמכים רק על נתוני הביקור במותג (הקו האחיד), רואים כי התובנות השיווקיות עומדות בניגוד לפרטיות הצרכן. עם זאת, התמונה משתנה עבור גישות קיבוץ הלוקחות בחשבון גם את נתוני המיקום. השילוב של נתוני מיקום הבית (קו מקווקו) חושף שתי תובנות מרכזיות. ראשית, עבור ערכים שאינם טריוויאליים של z (כלומר, $1 < z < N$), השילוב של נתוני מיקום הבית ליצירת פלחים מספק תחזיות ביצוע מעולות בהשוואה למודל המסתמך רק על נתוני ביקור במותג. שנית, בניגוד ל-AUC המתקבל במודל המסתמך רק על ביקור במותג ליצירת פלחים, עבור ערכים של $3 \leq z \leq 100$,

2 עבור ערכים של $z=1$ ו- $z=N$, הפילוח הוא טריוויאלי.

איור 4: כיצד השילוב של "המקום שבו ישנים" לעומת "המקום שבו קונים" משפיע על הטרייד-אוף שבין רמת הפרטיות לבין דיוק הניבוי.



אחד להגנה על הפרטיות של צרכנים רבים. משווקים יכולים לעמוד ביעדים שלהם בעזרת שימוש רק בנתונים הנוגעים למיקומים שאינם מגורים (ושימוש למשל במרכז הניאוגרפי של ביקורי המותג של הצרכן כתשומה ליצירת קבוצות). התוצאות שלנו מציגות דרך אפשרית לשיפור של פרטיות הצרכן בהקשר של איסוף ושימוש בנתוני המיקום של מכשירים ניידים.

משימוש בנתוני המכשירים הניידים. תוצאה זו מראה כי המחקר שלנו, שמתחשב רק בנתוני הביקור במוטג ובקואורדינטות של קווי הרוחב והאורך של מיקום החנות, יכול לספק מידה רבה של פרטיות לצרכן, יותר מאשר מחקר המשתמש בנתוני מיקום גולמיים של המכשיר הנייד (Ghose 2018; Macha et al. 2023).

דברים שכדאי לזכור

בעוד שאקוסיסטם הפרסום המותאם למיקום המוצג באיור 1 סייע למקד את הניתוח שלנו בטרייד-אוף שבין פרטיות הצרכן לבין דיוק הניבוי, נותרו כמה אתגרים חשובים ליישום.

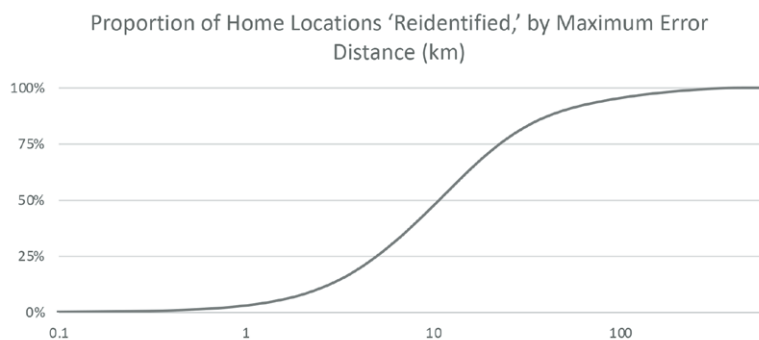
בניגוד לתקנת ה-GDPR שלפיה צרכנים צריכים להסכים לשיתוף הנתונים שלהם עם כל חברה, המודל שלנו מניח תרחיש קרוב יותר לאקוסיסטם של היום של נתוני מיקום, שבו מיקומי הביקורים בחנויות של מותגים משודרים באופן פסיבי לאחר הסכמה ראשונית כללית של הצרכנים לאיסוף מידע על המיקום שלהם. (Thompson & Warzel 2019b) הצביעו על חסרונות מערכתיים של גישה זו שעשויים להופיע בשווקים מסוג זה שאינם מוסדרים כהלכה.

השלכות ניהוליות

מכלול התוצאות שלנו מצביע על כך שהטרייד-אוף בין פרטיות הצרכן לבין תובנות שיווקיות עשוי להיות תפיסה דיכוטומית שגויה. קיבוץ של מכשירים על בסיס הדמיון בהתנהגות ביקור המותג שלהם יכול בו-בזמן לשפר את פרטיות הצרכן וגם לשפר את הדיוק של תובנות שיווקיות לגבי שימוש בנתונים ברמת המכשיר. חשוב לציין כי תובנות אלו חלות בתרחיש הנפוץ יותר ויותר, שבו הממדיות של נתוני המיקום של מכשירים ניידים המשמשת ליצירת פרופילים של קהלי פרסום מצטמצמת באמצעות צבירה זמנית.

יתרה מכך, הוכחנו שמיקום הבית, שיתכן שאותו הצרכנים לא מעוניינים לשתף, היה מיותר ביישום האמפירי שלנו. הגבלה של איסוף הנתונים למיקומים מסחריים יכולה להיות אמצעי

איור 5: ההתפלגות המצטברת של המרחק בין המיקומים המשוערים של הבתים האמיתיים לבין מיקומי הבתים המדומים.



בעבודה זו הותוו קווים מקבילים לדיון המתמשך על פרטיות הצרכן בפרסום מקוון, שנדונו גם בכך בנייר עבודה עם עמיתיי (Shoshani et al. 2022), שבו בדקנו האם רצונם של המשווקים לתובנות בנות פעולה והצורך של הצרכנים בפרטיות רבה יותר יכולים להתמלא בו-בזמן. לשם כך למדנו את המסגרת הנפוצה שבה פלטפורמות נתוני מיקום יוכלו להגביר את פרטיות הצרכן באמצעות קיבוץ מכשירים לקבוצות בעלות אפיונים דומים, ושיתוף רק של התכונות הממוצעות ברמת קבוצה (כולל הסבירות המשוערת שצרכן יבקר בחנויות קמעונאיות לא מקוונות של מותג נתון במהלך שבוע נתון).

התוצאות שלנו מראות כי ייתכן שמשווקים לא יצטרכו לבחור בין תובנות ניבוי לבין פרטיות הצרכן. במקום זאת, באמצעות קיבוץ המכשירים לקבוצות המבוססות על קווי דמיון בין המקומות שבהם הם קונים, כולל מספר הפעמים שהם מבקרים כל מותג, בתוספת המרכז הגיאוגרפי של כל ביקורי המותג שלהם – המפרסמים משיגים דיוק ניבוי גבוה ביחס למצב שבו הם משתמשים בנתונים ברמת המכשיר ו/או המיקום המשוער של בית הצרכנים. מכאן, שהסתמכות בלעדית על נתונים שנאספו ממקומות שאינם מגורים יכולה להתאים למטרות המשווקים ללא פגיעה בפרטיות המשתמש במידה בלתי סבירה. למרות שחלק מהאתגרים הארגוניים נופלים מחוץ לתחום הניתוח שלנו, אנו מקווים שתובנות אלו יעדכנו את התנהגות הרגולטורים ואת כל השחקנים בתעשיית נתוני המיקום של מכשירים ניידים.

אחד הפתרונות לבעיות העולות מהתחרות הקשה בין שחקנים קטנים ורגולציה חסרת מעוף יכול להיות הקמה של ארגון אחד או כמה ארגונים נדולים שיהיו אחראים על קביעת הסטנדרטים ויחייבו את החברות על כל הפרה של פרטיות הצרכן. למשל, אפשר ליישם בטלפונים של הצרכנים הסרה ו/או הצפנה של נתונים רגישים של צרכנים – על ידי אפל וגוגל, יצרניות מערכות הפעלה הפופולריות ביותר בשוק של מכשירים ניידים. ניסיון ריכוזיות כזה עשוי לזכות באמון הצרכנים בזכות נטילת האחריות המוגברת, וגם להעניק כוח תמחור עצום לחברות המעטות הפועלות כשומרות סף לכל נתוני מיקום הצרכנים.

בדומה לכך, בעוד שהניתוח שלנו התמקד בדיוק הניבוי הממוצע למקרה, הוא התעלם מההשלכות של הצורך לרכוש הופעות פרסומיות בכמויות גדולות. אומנם מצאנו שבממוצע חברות קטנות מרוויחות יותר מכול מצבירת נתוני המכשירים הניידים של הצרכנים, אבל לא בדקנו אם חברות אלו ימצאו כדאיות כלכלית בהגדלת נפח הפרסום שלהן לרמה שבה הן יוכלו לממש את הרווחים הממוצעים האלו.

סיכום ומסקנות

נתוני המכשירים הניידים של צרכנים שנאספו מטלפונים חכמים, הביאו ערך כלכלי עצום באמצעות האפשרות של המשווקים לטרגט את מסרי קידום המכירות בדיוק רב יותר. עם זאת, מכיוון שנתוני מיקום עשויים לחשוף לא רק העדפות מותג אלא גם מידע רגיש על צרכנים, עלה החשש לגבי המידה שבה איסוף ואחסון של נתוני מיקום דינמיים יפלוש לפרטיות הצרכן.

ד"ר פיטר פאל זובצ'ק peterz@tauex.tau.ac.il

- Arora, N., Liu, W.W., Robinson, K., Stein, E., Ensslen, D., Fiedler, L., & Schüler, G. (2021), "The value of getting personalization right or wrong is multiplying," *McKinsey & Company*, November 12. Accessed at: <https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/the-value-of-getting-personalization-right-or-wrong-is-multiplying>.
- Bohn, Dieter (2021), "Privacy and ads in Chrome are about to become FLoCing complicated," *The Verge*, March 30. Accessed at <https://www.theverge.com/2021/3/30/22358287/privacy-ads-google-chrome-floc-cookies-cookieapocalypse-finger-printing>.
- Bradley, P. S., Bennett, K. P., & Demiriz, A. (2000), "Constrained K-means clustering," Technical report, Microsoft Research, Redmond, WA, USA.
- De Montjoye, Yves-Alexandre, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel (2013), "Unique in the crowd: The privacy bounds of human mobility," *Scientific Reports*, 3 (1), 1-5.
- Edelman, G. (2020). "Why Don't We Just Ban Targeted Advertising?" *Wired*. Accessed at <https://www.wired.com/story/why-dont-we-just-ban-targeted-advertising/>.
- Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996), "A density-based algorithm for discovering clusters in large spatial databases with noise," In *proc. KDD* 96 (34), 226-231.
- Fisher, L. (2019), "Digital Marketing in Today's Privacy-Conscious World," *Insider Intelligence*, July 9. Accessed at <https://www.insiderintelligence.com/content/digital-marketing-in-todays-privacy-conscious-world>.
- Ghose, A. (2018). *TAP: Unlocking the mobile economy*. MIT Press.
- Ghose, A., Li, B., Macha, M., Sun, C., & Foutz, N. Z. (2022), "Trading privacy for public good: How did America react during COVID-19?," Working paper, NYU Stern School of Business.
- Goldfarb, Avi, and Catherine Tucker (2012), "Shifts in privacy concerns," *American Economic Review*, 102 (3), 349-53.
- Goldfarb, A., & Tucker, C. (2013), "Why managing consumer privacy can be an opportunity," *MIT Sloan Management Review*, 54(3), 10.
- Joseph, S. (2023), "Five years in, the GDPR has had a double-edged impact on the ad market," *Digiday*, May 25. Accessed at <https://digiday.com/marketing/five-years-in-the-gdpr-has-had-a-double-edged-impact-on-the-ad-market/>.
- Li, L., & Xi, Y. (2011), "Research on clustering algorithm and its parallelization strategy," In *proc. 2011 International Conference on Computational and Information Sciences*, (pp. 325-328). IEEE.
- Luo, X., Andrews, M., Fang, Z., & Phang, C. W. (2014), "Mobile targeting," *Management Science*, 60(7), 1738-1756.
- Macha, M., Foutz, N. Z., Li, B., & Ghose, A. (2023), "Personalized privacy preservation in consumer mobile trajectories," *Information Systems Research*, forthcoming.

- Mehta, I. (2023), "Google flips the switch on interest-based ads with 'Privacy Sandbox' rollout," *TechCrunch*, September 8. Accessed at <https://techcrunch.com/2023/09/08/google-flips-the-switch-on-interest-based-ads-with-privacy-sandbox-rollout/> .
- Rafieian, O., & Yoganarasimhan, H. (2021), "Targeting and privacy in mobile advertising," *Marketing Science*, 40 (2), 193-218.
- Shoshani, T., Zubcsek, P. P., & Reichman, S. (2024). "Analyzing Purchase Decisions Using Dynamic Location Data," *Journal of Interactive Marketing*, forthcoming.
- Shoshani, T., Zubcsek, P. P., & Schweidel, D. A. (2022), "Balancing Consumer Privacy with Marketing Insights in Mobile Location Data," *Marketing Science Institute Working Paper #22-102*.
- Skiera, B., Miller, K., Jin, Y., Kraft, L., Laub, R., & Schmitt, J. (2022). *The impact of the General Data Protection Regulation (GDPR) on the online advertising market*. Frankfurt.
- Snider, M. (2021), "Apple's privacy changes to iOS have arrived. What do you do with your Facebook app?" *USA Today*, April 26, accessed at <https://www.usatoday.com/story/tech/2021/04/26/facebook-apple-iphone-ipad-privacy-ios-software-update/7368248002/>.
- Sweeney, Latanya (2002), "K-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10 (05), 557-570.
- Thompson, S. A., & Warzel, C. (2019a), "How to Track President Trump," *The New York Times*, December 20, accessed at <https://www.nytimes.com/interactive/2019/12/20/opinion/location-data-national-security.html>.
- Thompson, S. A., & Warzel, C. (2019b), "Twelve Million Phones, One Dataset, Zero Privacy," *The New York Times*, December 19, accessed at <https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html> .
- Tucker, Catherine E. (2014), "Social networks, personalized advertising, and privacy controls," *Journal of Marketing Research*, 51 (5), 546-562.
- Verhagen, T., Meents, S., Merikivi, J., Moes, A., & Weltevreden, J. (2022), "How location-based messages influence customers' store visit attitudes: an integrative model of message value," *International Journal of Retail & Distribution Management*, 50(7), 781-798.
- Warzel, C., & Thompson, S. A. (2021), "They Stormed the Capitol. Their Apps Tracked Them," *The New York Times*, February 5, accessed at <https://www.nytimes.com/2021/02/05/opinion/capitol-attack-cellphone-data.html>.