https://pubsonline.informs.org/journal/isre

INFORMATION SYSTEMS RESEARCH
Articles in Advance, pp. 1–20
ISSN 1047-7047 (print), ISSN 1526-5536 (online)

Learning Personalized Privacy Preference from Public Data

Wen Wang,[a,*] **Beibei Li**[b]

[a] University of Maryland at College Park, Information System, College Park, Maryland 20742; [b] Carnegie Mellon University, Information Systems, Pittsburgh, Pennsylvania 15213

*Corresponding author

Contact: wenw@umd.edu, https://orcid.org/0000-0001-5983-3224 (WW); beibeili@andrew.cmu.edu, https://orcid.org/0000-0001-5466-7925 (BL)

Received: May 25, 2023
Revised: November 8, 2023; March 15, 2024
Accepted: May 12, 2024
Published Online in Articles in Advance: June 13, 2024

https://doi.org/10.1287/isre.2023.0318

Abstract. Learning consumers' personalized privacy preferences is crucial for firms and policymakers to establish trust and compliance and guide effective policymaking. Existing approaches rely mostly on private information such as proprietary user behavior data and individual-level demographic and socio-economic factors, or require explicit user input, which can be invasive and burdensome, potentially leading to user dissatisfaction. Nowadays, individuals generate and share vast amounts of information about themselves in the public domain, which can provide a valuable multifaceted view of their behaviors, attitudes, and preferences. This information thus has the potential to provide valuable insights into individuals' privacy preferences. In this study, we propose a novel framework to predict personalized privacy preference by leveraging a ubiquitous source of public data—social media posts. Deeply rooted in psychological and privacy theories, we use deep learning model and natural language processing algorithms to learn theory-driven psychosocial traits such as lifestyle, risk preference, personality, privacy-related economic preferences, linguistic styles, and more from social media posts. Interestingly, we find that psychosocial traits from public data provide greater predictive power than private information. Furthermore, we conduct multiple interpretability analyses to understand what drives the model's performance. Finally, we demonstrate the practical value of our model and show that our framework can assist platforms and policymakers in forecasting the consequences of privacy policies. Overall, our framework provides managerial implications for enhancing consumer privacy control and trust, optimizing platform data management, and informing policymakers about better data privacy regulations.

History: Ravi Bapna, Senior Editor; Heng Xu, Associate Editor.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/isre.2023.0318.

Keywords: personalized privacy preference • public data source • deep learning • natural language processing • psychosocial traits

1. Introduction

In today's digitally connected world, the collection of consumer data has grown rapidly, and with that growth comes an increasing need to respect and protect consumers' privacy preferences. Firms and policymakers have recognized the importance of understanding consumers' personalized privacy preferences and have made it a top priority to ensure that consumer data are collected, stored, and used in a responsible and ethical manner (Acquisti et al. 2015, 2022; Steed et al. 2022; Xu and Dinev 2022). By understanding consumers' personalized privacy preferences, companies and policymakers can enhance consumer trust and improve user experience through better privacy policies. This understanding also ensures compliance with evolving privacy laws and fosters a digital environment that respects user privacy.

Existing approaches of learning an individual's privacy preference heavily lean on either seeking explicit user input (Xu et al. 2012, Jiang et al. 2013, Acquisti et al. 2015, Xu and Zhang 2022) or accessing private information such as proprietary user behavior data and individual-level demographic and socio-economic factors (Dong et al. 2015, Balapour et al. 2020, Serramia et al. 2023). However, these strategies can be seen as intrusive and burdensome, and they encounter scalability and accessibility issues when dealing with millions or even billions of users. Their lack of automation capability hinders practical deployment on a large scale, obstructing online platforms and policymakers in real-world applications.

In this study, we propose a novel framework to automatically predict personalized privacy preference using public data, without the need for users' input or their private information. With the widespread use of digital technologies, individuals are generating and sharing vast amounts of data about themselves in the public domain, including information such as social media activity, blog activity, online reviews, and publicly

available records. This public data can provide a valuable multifaceted view of individuals' behaviors, attitudes, and preferences and thus can provide fruitful signals about individuals' privacy preferences; this in turn makes it possible to predict how those individuals may respond to privacy-related decisions and to develop more effective privacy policies and practices. For example, social media posts, one of the most common types of public data, can reveal different aspects of an individual's life. People usually post their lifestyle activities and emotional states on social media. In addition, what they post and how they post also reveal their intrinsic values and preferences, such as interests, personalities, and economic preferences. Even without knowing a person's private information such as occupation and demographic characteristics, the latent traits social media posts reveal can provide valuable insights about that person's privacy preference. In fact, the influence of such psychosocial traits (e.g., personalities, habits, etc.) on an individual's privacy perceptions and decision making has been widely documented in the literature (Phelps et al. 2000, Lu et al. 2004, Xu 2007, Quinn 2016, Barth and De Jong 2017).

It's worth noting that public data, using social media posts as an example, have two significant and unique advantages for understanding privacy preference. First, they are publicly accessible, enabling large-scale learning. Second, such data represent "organic data," generated by users themselves without overt research design components and constantly recorded by digital devices (Xu et al. 2020). This organic nature enables users to articulate their preferences in an authentic manner without external distortions. It's important especially when considering that data gleaned from surveys and laboratory experiments, commonly employed in prior privacy research, may be susceptible to noise and research design biases[1] (Marreiros et al. 2017, Xu and Zhang 2022).

Automatic prediction of personalized privacy preference from public data holds immense potential to inform individuals, organizations, and policymakers about data privacy and security decisions. For individuals, learning their unique privacy preferences would help platforms serve them better, enhancing their online experience and sense of control. For platforms and policymakers, implementing such an approach could refine data handling practices and guide regulatory action and platform guidelines, enabling them to shape policy decisions and platform designs that account for each individual's unique preferences and uphold the rights of the digital citizen.

In this study, we predict personalized privacy preference from a ubiquitous source of public data—social media posts (e.g., Twitter). In recent years, social media has become an essential part of people's daily routines, and its usage has significantly increased. According to a recent report (IBISworld 2020), more than half of the world now uses social media (59%) and the average daily time spent using social media is 2 h 31 min. Individuals actively post their opinions, thoughts, and daily activities on the Internet, making these posts widely accessible to anyone. As users typically recognize that their social media posts are shared publicly, these posts, especially those publicly available posts, are considered public data because of their broad accessibility and the implied consent granted by users when choosing to share such information. To collect a comprehensive data set, we conduct an online experiment to collect individuals' privacy preference, their private information (e.g., demographic characteristics), and their public data (i.e., publicly available social media posts) under their agreement. To capture the individual's privacy preference, we adopt the Internet Users' Information Privacy Concerns (IUIPC) Survey, a scale widely used in privacy literature (Malhotra et al. 2004).

Drawing on the theory of privacy and psychology, we identify five categories of theory-driven psychosocial traits that can influence individuals' privacy preferences. These traits include lifestyle activities and habits, personality, risk preference, privacy-related economic preferences, and emotional states. Then we adopt multiple natural language processing algorithms and deep learning algorithms to learn these psychosocial traits from social media posts. We next use a widely utilized machine learning model, LightGBM (Ke et al. 2017) (Light Gradient Boosting Machine), to predict the individual privacy preference based on the extracted traits. LightGBM is an open-source, gradient boosting framework that uses tree-based learning algorithms to perform machine learning tasks such as classification, regression, and ranking. We conduct extensive experiments to explore the predictive power of public data compared with private information. Furthermore, we demonstrate the practical value of our model for businesses and society by performing multiple in-depth interpretability analyses and constructing a decision support showcase to assist platforms and policymakers in forecasting the consequences of any privacy policy.

Our study yields some interesting findings. First, results with cross-validation show that both private information (i.e., demographic characteristics) and public data (i.e., psychosocial traits encoded in social media posts) can provide significant predictive power. Interestingly, the psychosocial traits from public data provide more predictive power than private information. This highlights the feasibility that one can predict individuals' privacy preferences without access to their private data, relying solely on insights gained from their social media activities.

Second, we conduct multiple in-depth interpretability analyses to understand the model performance. Results show that lifestyle, risk preference, and emotional states are the most important traits in public data for predicting

individual privacy preference; age, employment status, and gender are the most important private information. Moreover, we explore how the importance of psychosocial traits varies among different population groups. We find that compared with other subgroups, psychosocial traits are more important than private information for young people (i.e., age 18–34), high-income individuals (i.e., over $50,000 (50k)), and full-time employees.

Finally, we demonstrate the practical value of our model through decision support showcases. Through simulation, we show that our model can assist platforms and policymakers in forecasting the consequences of privacy policies ahead of time. By simulating user bases and the policy shocks of different privacy risk levels, we are able to quantify shifts in the platform user base in response to changes in policy. Interestingly, we observe that policies resulting in higher privacy risk (e.g., liberal and loose policies including extensive data collection, sharing with third parties, and long-term data retention) can lead to a significant increase in the number of users on the platform who display diverse lifestyles and risk-seeking traits, compared with those policies associated with lower risk (e.g., conservative and strict policies including limiting data collection to only what's necessary, restricting third-party sharing, and implementing stringent data retention and deletion policies). Additionally, we find that these policies leading to high privacy risk could also lead to a significant decrease in a platform's female user base. By utilizing our tool, platforms can modulate their risk levels and strike a balance to avoid potential discrimination. This kind of foresight allows platforms to evaluate the possible adverse outcomes of policy changes, prepare for shifts in the user base, and avoid unanticipated consequences, thereby fostering a more thoughtful and responsive approach to privacy policy development.

In summary, our study makes a fourfold contribution. First, our paper is among the first to investigate the predictive power of public data for personalized privacy preference. Notably, we find that in determining privacy preference, publicly available tweets possess greater predictive power than private information such as demographic details. Second, we demonstrate the feasibility of using various natural language processing algorithms to learn a broad range of psychosocial traits from these publicly available tweets. Third, our model provides an automated tool that estimates an individual's privacy preference without requiring user input. It has the potential for platform deployment for large-scale inference. Fourth, we demonstrate the practical value of our model. We show that it can serve as a decision support system, aiding platforms and policymakers in assessing the consequences of privacy policies. Together, our model and findings provide individuals, platforms, and policymakers with new tools and insights to enhance consumer privacy control and trust, optimize data management, and inform better data privacy regulations.

2. Literature Review

2.1. Learning Individual Privacy Preference

Our study is closely related to the learning of individual privacy preferences in literature. Individual privacy preference encapsulates personal beliefs, attitudes, and perceptions toward privacy, and it has been shown to influence users' behavioral responses (Belanger et al. 2002, Xu and Teo 2005, Metzger 2007), stances on regulation (Milberg et al. 2000, Smith 2001, Pavlou 2011), technology adoption (Vijayasarathy 2004, Easwara Moorthy and Vu 2015), consumer choices and brand loyalty (Lee 2008, Jai and King 2016), policy development (Goodwin 1991, Hochheiser 2002), etc. Much of the prior literature often relies on direct user inquiries (e.g., survey or experimental responses) to obtain users' privacy preferences (Xu et al. 2012, Jiang et al. 2013, Acquisti et al. 2015, Adjerid et al. 2019, Alashoor et al. 2022, Xu and Zhang 2022). Individuals are asked to rank, rate, or detail their specific preferences related to a variety of topics such as preferences on personal data usage, sharing practices, and comfort levels with different privacy protection measures, aiming to paint a detailed picture of how users perceive and value their online privacy.

However, this approach tends to be time-consuming and challenging to scale up when dealing with millions or even billions of users. Furthermore, it lacks the capacity for automation, impeding its practical deployment to aid online platforms and policymakers in real-world scenarios. Our study contributes to this stream of literature by offering an alternative method to automatically assess users' privacy preferences: predicting them from a publicly available data source, namely, social media posts. It has the potential to be embedded with other variables of interest to better understand privacy decision making in the future.

2.2. Privacy Decision Modeling

Our study is also related to the literature on privacy decision modeling, which involves learning and predicting users' privacy decisions in specific contexts. For example, Serramia et al. (2023) and Barbosa et al. (2019) model privacy decisions related to smart devices (e.g., smart home and smart speakers), and they predict "allow" or "deny" preferences based on the user's current information privacy inclinations as well as personal and home attributes. Balapour et al. (2020), Brandão et al. (2022), Alshehri and Alotaibi (2019), and Mendes et al. (2022) model consumers' mobile app privacy decisions to predict whether users will grant or deny an application access to their devices' data and sensors. Dong et al. (2015), Naini et al. (2015), and Bigwood et al. (2012) investigate users' information-sharing behavior (i.e.,

deciding whether to share information with all contacts or just part of their contact list) and request behavior (i.e., whether to accept friend requests) in the context of online social networks.

This stream of literature tends to rely heavily on private information such as proprietary user behavior data (e.g., app usage behavior, home attributes) and individual-level demographic and socio-economic factors. However, these types of information are generally not accessible and cannot be widely generalized to other scenarios. In addition, they tend to focus on specific privacy decisions in different contexts instead of the general and intrinsic privacy preference. In contrast with existing work, we aim to learn a general intrinsic privacy preference using publicly available data without any private information. Our learned intrinsic privacy preference should be generalizable to and informative in many different contexts to inform policy decisions and platform designs.

3. Proposed Framework to Learn Theory-Driven Psychosocial Traits from Public Data with Deep Learning Algorithms

In this study, we focus on a ubiquitous and readily accessible data source—social media posts, with emphasis on platforms such as Twitter. In the contemporary era, the role of social media in daily life has become significant. It is commonplace for individuals to publicly express their thoughts, emotions, and daily activities on the Internet, rendering these posts available to anyone with Internet connectivity.

Many social media platforms, as part of their terms of service and privacy policies, inform users that anything they post publicly can be seen and accessed by anyone. For example, Twitter's privacy policy explains that most of the information users provide is information the users are asking Twitter to make public. In Twitter's own words, "Any registered user of Twitter can send a Tweet, which is public by default. Twitter is primarily designed to help you share information with the world. Most of the information you provide us through Twitter is information you are asking us to make public." In addition, social media data have been widely used as a source of public data in the literature aimed at understanding user behaviors and promoting marketing efforts (Acquisti and Gross 2009, Ghose et al. 2012, Liu et al. 2016, Adamopoulos et al. 2018, Zhang and Moe 2021, Wang et al. 2022, Oh et al. 2023, Schoenmueller et al. 2023).

Social media posts are capable of offering a valuable and diverse perspective on individual behaviors, attitudes, and preferences. This wealth of information is invaluable for comprehending patterns related to privacy preference. In this section, we first build on psychological and privacy theory to identify the psychosocial traits that can determine one's privacy preference and then describe how to operationalize and extract such theory-driven traits from social media posts using deep learning algorithms.

3.1. Theoretical Foundations for Psychosocial Traits

We now outline the theoretical foundations of factors that affect humans' privacy preferences and decision making. Two main types of human characteristics exist: psychosocial traits and demographic characteristics (Bongers et al. 1993, Hoogendoorn et al. 2000, Everson-Rose and Lewis 2005, Bonde 2008, Maree 2021). The former denotes the psychological and social features of an individual, including personality traits, attitudes, beliefs, values, and interests. These factors are not necessarily intrinsic to an individual's biological makeup or demographic profile, but are instead shaped by a variety of environmental and situational factors. Demographic characteristics refer to the statistical attributes of a population, such as age, gender, race, education, income, and occupation. They are commonly used to classify people into different groups for research or marketing purposes, and they can provide insights into the behaviors and preferences of different populations. Both psychosocial traits and demographic characteristics are informative in helping us understand human behavior and develop communication, marketing, or public policy strategies.

We describe the theoretical foundations underlying the factors that have been documented in the literature as contributing to privacy preference. We divide these factors into the two broad categories of psychosocial traits and demographic characteristics.

3.1.1. Psychosocial Traits

3.1.1.1. Personality. Prior literature has documented that personality (e.g., introversion versus extroversion) influences people's perceptions (Lu et al. 2004, Xu 2007). For instance, Bansal et al. (2010) and Smith et al. (2011) document the role of the "Big 5" personality traits in influencing individuals' perceptions of health information sensitivity. These five dimensions of personality are openness, conscientiousness, extroversion, agreeableness, and neuroticism.

3.1.1.2. Risk Preference. Users' risk preferences can be informative regarding their privacy preferences. For instance, Hong et al. (2021) suggest that risk avoidance plays a prominent role in individual privacy decision making. Individuals who are more risk averse may be more protective of their personal information and privacy, whereas those who are more risk-seeking may be more willing to share personal information with others.

3.1.1.3. Lifestyle Activities and Habits. Prior literature has documented that life activities and habits such as habitual pastimes of relaxing and entertainment and

shopping habits are likely to be related to privacy decisions (Phelps et al. 2000, Quinn 2016, Barth and De Jong 2017). Habits and lifestyle activities are a holistic reflection of an individual's life, encompassing their behaviors and daily choices. This covers a variety of aspects ranging from social activities to work habits. Habits or life choices usually reflect a person's deep and intrinsic values, beliefs, and motivations, which collectively play a significant role in shaping one's privacy preference.

3.1.1.4. Privacy-Related Economic Preference. Extant literature has been largely predicated on the notion that privacy decision making is largely a rational process driven by what we may refer to as "normative" factors (Adjerid et al. 2018). Such factors may include the objective benefits and costs of information disclosure, and the agent's stable, coherent economic preference. Normative theories of consumer choice are those consistent with the classical economic view of consumers as deliberative, utility-maximizing, rational agents who possess reasonably stable, and therefore predictable, preferences for goods (Mullainathan and Thaler 2000). From this perspective, a privacy calculus view of consumer decision posits that privacy is subject to interpretation in "economic terms" (Klopfer and Rubenstein 1977) and privacy decisions can be construed as the result of an economic calculus that weighs the expected benefits of privacy allowances against their resulting costs. Therefore, understanding individuals' economic preference is important to learn their privacy preference.

3.1.1.5. Emotional States. People rely on feelings to make strategic decisions. Emotion plays an important role in shaping our attitudes and behaviors, including those related to privacy. Its powerful role in explaining privacy preference has been documented in the literature. For example, Li et al. (2017) and Berendt et al. (2005) show that online users with high privacy concerns disclosed their most private information to a website that they found entertaining. Zhang and Xu (2016) show that the feeling of creepiness mediates the relationship between nudging and the user's attitude toward privacy. When people feel strong emotions, such as fear, anger, or anxiety, they may be more likely to prioritize their privacy and take steps to protect it. On the other hand, when people feel positive emotions, such as trust or excitement, they may be more willing to share personal information with others. Therefore, understanding a person's emotions can provide valuable insight into their privacy concerns.

3.1.2. Demographic Characteristics. Various studies have also investigated how demographic differences affect the degree of stated privacy concern (Bartel Sheehan 1999, Culnan and Armstrong 1999, Sheehan and Hoy 2000, Chen and Rea 2004). For example, Bartel Sheehan (1999) shows that women have generally been found to be more concerned than men about the collection of their private information. In addition, consumers who were less concerned about privacy have been found to be more likely to be less educated and to be African-American (Culnan 1995). All of these demographic characteristics are generally considered to be private information.

3.2. Operationalizing Psychosocial Traits with Deep Learning Algorithms

The extraction of information from unstructured data has received considerable attention in the field of information systems (Li et al. 2023; Wang et al. 2023, 2024). After defining the theory-driven psychosocial traits, we show how we can use deep learning algorithms to learn such traits from a user's social media posts.

3.2.1. Personality. We adopt the Mphasis HyperGraf Big 5 Trait Analyzer from AWS Marketplace to take input text from a user and assign a score based on the Big 5 personality traits, which are openness, conscientiousness, extroversion, agreeableness, and neuroticism. The back-end algorithm utilizes natural language processing and computational psycholinguistics to assign the scores. To provide input, we merge all tweets from each user into one long document. Each person has a 5-dimension personality score and each dimension ranges from one to five.

3.2.2. Risk Preference. Risk preference is determined by one's inclination to take risks. Activity diversity can serve as an indicator of risk preference. For instance, someone who partakes in diverse activities in their daily life, such as traveling, visiting bars, shopping, and engaging in outdoor pursuits, might be viewed as more exploratory. In contrast, those who mainly focus on a single activity, like working, are more likely to be seen as more conservative. Although activity diversity focuses on the variability of high-level types of activities, it does not capture the semantic meaning of each activity. Thus, another dimension of risk preference could be the risk encoded in the semantics, that is, the propensity for high-risk activities, such as extreme sports. When someone has a high propensity for such high-risk activities, this demonstrates risk-seeking traits. Therefore, we measure risk preference using two criteria: *activity diversity* and *participation in high-risk activities*. These two metrics complement each other.

3.2.2.1. Activity Diversity. First, to measure the activity diversity, we need to detect the activity from social media posts. Social media posts encode rich information about where the user is performing an activity or living an experience in point of interest (POI). For example, a post "currently visiting my dream school!" indicates the

user had an experience in *College & University*; a post that reads "Its a humid day @EncoreBeachClub so get ready for @Alesso to take the decks and get the party #turnt #beachclub #ladies" indicates the user was in *Nightlife Spot*; a post of "Came to get an old fashioned tape measure and a button for my coat" indicates the user is engaging in *Shop & Service*.

Villegas et al. (2020) demonstrate the feasibility of accurately predicting semantic location information, specifically POIs, from tweets, using the Bidirectional Encoder Representations from Transformers (BERT). They construct a large-scale data set combining tweets with their corresponding POIs, which are extracted from Foursquare and tagged by users. This approach offers the capability to infer users' activities based on their tweets. Although the original researchers do not release the trained model and the actual tweet texts, they provide tweet identifiers (IDs) and corresponding POI labels. Consequently, we collect the same tweets using these IDs and then implement and fine-tune a BERT model on our own. Out of the original data set, which contained 196,235 tweets, some were no longer accessible by the time we attempted retrieval. This reduces our data set to 101,560 tweets. We then implement and fine-tune our BERT model based on the training strategies described by Villegas et al. (2020), using the data we collected. We run our fine-tuning on an Nvidia A100 GPU. Beyond BERT, we also test several other benchmark models. The results show that our trained BERT model not only surpasses the performance of all benchmarks but also performs on par with the results presented by Villegas et al. (2020). Further details on the data set, training procedures, and model performance are available in the Online Appendix.

We use the trained BERT to predict the POI for each tweet. Subsequently, we aggregate each user's POI sequence and calculate the entropy of the POI category to measure activity diversity. A high entropy value indicates high activity diversity, suggesting that the individual is of an exploratory type. Conversely, low entropy signifies low activity diversity, indicating a more conservative individual.

$$H(x) = -\sum_{i=1}^{N} p(x_i)\log\, p(x_i) \tag{1}$$

where N is the number of the POI category.

3.2.2.2. Participation in High-Risk Activities.

Second, we assess the users' propensity for participation in high-risk activities. Each activity can have a distinct risk level. For example, the post "Just jumped out of a plane 10,000 feet up and lived to tweet about it! The world looks so different from above. Bucket list item #Skydiving #FreefallFeels" indicates that a user participated in the high-risk activity of skydiving.

To quantify participation in high-risk activities, we first use GPT-4 to curate a list of such activities. We identify a total of 104 activities, which cover a wide range of categories. These include

- Extreme sports and outdoor activities: for example, skydiving, bungee jumping.
- Transportation risks: for example, motorcycle speeding, illegal car racing.
- Work-related risks: for example, high-altitude construction work, deep-sea fishing in hazardous conditions.
- Hazardous recreational activities: for example, recklessly handling fireworks, shooting at a gun range without ear protection.
- Miscellaneous risks: for example, high-stakes gambling, wing suit flying, etc.

Subsequently, we quantify the propensity of each tweet that referenced a high-risk activity. This quantification is based on the cosine similarity between a Twitter post and identified high-risk activities. For this, we employ the sentence transformer (Reimers and Gurevych 2019), a cutting-edge framework for transforming sentences into continuous vectors or embeddings. Both tweets and high-risk activities are converted into embeddings to capture their semantic meanings. We then compute the similarity between each tweet and each high-risk activity. A high similarity implies a high likelihood that a tweet pertains to participation in a high-risk activity. From each user, we extract two features: "ever participate," determined by the highest probability among all tweets, and "mean participation probability," which is calculated using the average probability from all tweets.

3.2.3. Habit and Lifestyle Activity.

We quantify lifestyle and habit by using the unsupervised topic models on inferred daily points of interest of each user. A full image of a user's POI (i.e., where they have visited) and when they visited (i.e., weekends or weekdays, morning, afternoon, or night) can show a person's lifestyle. After we have each user's full POI sequence, we can construct an activity profile for each user. An activity profile D_i of an individual i is defined as mapping POI and timestamp to activities that exhibit a pattern of behavior. D_i is a set of tuples $D_i = \{d_i^i, \ldots, d_{n_i}^i\}$, $d_j^i = (a_j^i, c_j^i)$, where a_j^i is the POI (i.e., activity), c_j^i is the coarser timestamp of this activity, and n_i denotes the number of activities for individual i. To abstract away variations of the exact time in day-to-day activities, a coarser timestamp (timestamp associated with an individual's location) is associated with each activity: 0–7, 7–9, 9–11, 11–14, 14–17, 17–19, 19–21, 21–23. Automatic discovery of individual lifestyles from location data is a nontrivial problem given the massive scale and high dimensionality. Also, the differences in an individual's activities across days and the differences from other individuals' activities add further complexity.

We take an unsupervised topic modeling approach that has shown potential for uncovering complex temporal and behavioral patterns in individuals' daily routine on textual semantic location data. Specifically, we leverage the concept of latent Dirichlet allocation (LDA) designed for text documents to model an individual's day-to-day activities. LDA is a probabilistic, unsupervised learning model of a bag of words and of hidden discrete variables called topics. For text modeling, we may view each document as a mixture of various topics, where each topic is characterized as a distribution over words. To identify lifestyles, we make an analogy between text documents and day-to-day activities, authors, and individuals. We view each activity d_j^i in D_i, and the mapped activity profile as a word w. We represent each day's activities of an individual (author) as a bag of words—document d.

3.2.3.1. Weekend Lifestyles. We detect users' lifestyles on weekends and weekdays separately. We represent each lifestyle as the top activities ranked by their relevance. The detected lifestyle for a weekend is shown in Table 1. The top activities of weekend topic 0, working on the weekend, are dominated by "professional & other" during most times of a day. For the weekend topic 1, nighttime entertainment, the top activities are dominated by "art & entertainment" at 21–23 and 19–21; for the weekend topic 2, daytime entertainment and outdoors, the top activities are art & entertainment at 11–14, 9–11, and 7–9. The weekend topic 4, late-night diverse lifestyle, has a very diverse activity portfolio including professional & other, art & entertainment, "outdoor," "travel & transport," and "shop & service."

3.2.3.2. Weekday Lifestyles. In addition to the weekend lifestyles, we also learn about the weekday lifestyle, which is summarized in Table 2. On the weekdays, topics 0 and 1 are dominated by working and entertainment, respectively. They are relatively nondiverse lifestyles. Weekday topics 2 and 3 are diverse lifestyles with a diverse set of activities. Weekday topic 2 is a diverse daytime style where all activities happen during the daytime (e.g., 11–14, 9–11). The weekday topic 3 is a diverse nightlife style where all activities happen during the late nights such as 0–7 and 21–23.

3.2.4. Economic Thinking. Language, particularly in tweets about financial matters or when using quantitative expressions, mirrors individuals' economic mindset and preferences. On the one hand, the use of financial and economic terminology can signal an individual's knowledge of and interest in economic matters. This can be an indicator of economic preference, as those who are more financially literate or interested are likely to make different economic choices than those who are not. On the other hand, quantitative expressions, particularly those related to money or numbers, are concrete and can be directly associated with economic thinking or sensitivity to value and cost. We measure a person's economic thinking, as expressed in their tweets, using two sets of features.

3.2.4.1. Financial-Economic-Oriented Word Usage. We first use an economics and finance lexicon to quantify economic thinking. To compile a set of financial-economic-oriented words, we combine two sources: (i) *Glossary: Economics and Personal Finance Terms*, published

Table 1. Weekend Lifestyles with Their Top Activities and Corresponding Timestamp

Activity (timestamp)			
Weekend topic 0: Working on weekend	Weekend topic 1: Nighttime entertainment	Weekend topic 2: Daytime entertainment and outdoors	Weekend topic 3: Late-night diverse lifestyle
Professional & other (11–14)	Art & entertainment (21–23)	Art & entertainment (11–14)	Professional & other (0–7)
Professional & other (9–11)	Art & entertainment (19–21)	Art & entertainment (9–11)	Art & entertainment (0–7)
Professional & other (14–17)	Art & entertainment (14–17)	Art & entertainment (7–9)	Outdoor (0–7)
Professional & other (17–19)	Art & entertainment (17–19)	Art & entertainment (14–17)	College & university (0–7)
Professional & other (7–9)	Art & entertainment (11–14)	Outdoor (11–14)	Travel & transport (0–7)
Professional & other (19–21)	Outdoor (21–23)	Outdoor (14–17)	Shop & service (0–7)
College & university (11–14)	Professional & other (21–23)	Outdoor (9–11)	Professional & other (21–23)
Art & entertainment (11–14)	Outdoor (14–17)	Outdoor (7–9)	Food (0–7)
Professional & other (21–23)	Outdoor (19–21)	Professional & other (11–14)	Art & entertainment (21–23)
Outdoor (11–14)	College & university (21–23)	Professional & other (14–17)	Outdoor (21–23)

Table 2. Weekday Lifestyles with Their Top Activities and Corresponding Timestamp

Activity (timestamp)			
Weekday topic 0: Working dominant	Weekday topic 1: Entertainment dominant	Weekday topic 2: Diverse daytime style	Weekday topic 3: Diverse nightlife style
Professional & other (14–17)	Art & entertainment (21–23)	Professional & other (11–14)	Professional & other (0–7)
Professional & other (11–14)	Art & entertainment (19–21)	Art & entertainment (11–14)	Art & entertainment (0–7)
Professional & other (17–19)	Art & entertainment (14–17)	Professional & other (9–11)	Outdoor (0–7)
Professional & other (19–21)	Art & entertainment (11–14)	Outdoor (11–14)	College & university (0–7)
Professional & other (9–11)	Art & entertainment (17–19)	Professional & other (7–9)	Travel & transport (0–7)
Professional & other (21–23)	Outdoor (21–23)	Outdoor (9–11)	Shop & service (0–7)
Art & entertainment (14–17)	Outdoor (19–21)	Art & entertainment (9–11)	Professional & other (21–23)
Professional & other (7–9)	Outdoor (14–17)	Art & entertainment (7–9)	Art & entertainment (21–23)
College & university (11–14)	Professional & other (21–23)	Outdoor (7–9)	Food (0–7)
College & university (14–17)	Outdoor (11–14)	Shop & service (11–14)	Outdoor (21–23)

by the Federal Reserve Bank of St. Louis, MO, and (ii) *Glossary of Financial and Business Terms*, published by the *New York Times*. Examples of identified words include "bond yield," "capital gains," and "liquid asset." We calculate the ratio of financial-economic-oriented words for each user by taking the number of financial-economic words detected divided by the total number of words. A higher ratio of financially and economically oriented word usage suggests that a user has a stronger inclination to think from an economic and financial perspective in their daily lives. That said, the user is more likely to have a keen sense of economic calculus and thus to weigh the expected benefits of privacy allowances against their associated costs.

3.2.4.2. Quantitative Expression. In addition, we detect the quantitative expression in tweets, that is, usage of money, quantity, and number words using named entity detection with NLTK and SpaCy. People's propensity to mention money and numbers in their social media posts can serve as a measure of economic thinking because it reflects a quantitative orientation and a sensitivity to value and cost. If a user regularly references numbers and monetary values, this can indicate an analytical approach to understanding the world, an awareness of trade-offs, and a practical perspective on daily decisions. This attention to detail and focus on quantifiable metrics align with the core principles of economic thinking.

3.2.5. Emotional States. We use a pretrained emotion detection model for tweets (Hartmann 2022) to detect seven emotions including anger, disgust, fear, joy, neutrality, sadness, and surprise. The model was trained on a comprehensive and diverse data set of emotions that included emotion labels for texts from Twitter and Reddit, student self-reports, and utterances from TV dialogue. For each tweet, we extract a probability vector for 7-dimension emotion. We then extract three sets of user-level emotion status.

3.2.5.1. Average Emotion Status. We take the average of all tweets for each person to represent a user's average emotion status.

3.2.5.2. Time-Sensitive Emotion Status. In addition, we capture the sequential patterns or dynamics of a user's emotional status over time. We quantify the exponential moving average (EMA) of the emotion sequence. In this way, we assign more weight to recent tweets and less weight to older tweets. We calculate the EMA for each feature of the emotion sequence as follows:

$$\text{(EMA)} \quad y_t = \alpha * x_t + (1-\alpha) * y_{t-1} \tag{2}$$

where y_t denotes the EMA up to t; α is the weight to the most recent tweet emotion x_t at time step t and its value lies between zero and one (in our case, we set $\alpha = 0.5$ as it is a moderate level of priority to recency); $(1-\alpha)$ is the weight to all previous tweets' EMA up to $t-1$; $t \in [0, k]$ with 0 representing the earliest tweet time and k representing the most recent tweet time.

3.2.5.3. Variability of Emotional Status. Furthermore, we extract the standard deviation (std) from each emotion sequence. A larger standard deviation suggests that an individual experiences greater emotional variability over time.

3.2.6. Other Control Features. We also capture the following features to represent users' social interactivity and their basic linguistic style.

3.2.6.1. Interdependent Self. The distinction between an interdependent self and an independent self has been found to influence individual privacy concerns (Xu 2007). The interdependent self emphasizes social roles, relationships, and the beliefs, values, and norms of the in-group. On the other hand, the independent self emphasizes personal goals, characteristics, achievements, and desires. A high level of social interactivity is indicative of the interdependent self. We use metrics such as the number of comments, retweets, likes, and followers to represent a user's social interactivity and their alignment with the interdependent self.

3.2.6.2. Verbal Cues. We also include features related to users' linguistic styles. A user's basic linguistic style can provide insights into various aspects of their real life, such as their education level and personality. This, in turn, has potential to shed light on their privacy preferences. We extract various linguistic features using the package TAALES (Kyle and Crossley 2015). Specifically, we extract 11 feature sets that cover diverse linguistic features such as concreteness, contextual distinctiveness, contextual diversity, unigram and *n*-gram frequency, etc. Please refer to the Online Appendix for a comprehensive list of the verbal cues that we extracted and detailed descriptions of them.

3.2.6.3. Part of Speech. In addition to the aforementioned verbal cues, we also detect a user's part-of-speech tagging (POS tagging). POS tagging is a natural language processing task where each word in a sentence is assigned a part of speech (e.g., noun, verb, adjective, etc.). This provides insights into a user's high-level language habits. We identify nouns, verbs, adjectives, adverbs, and "wh-" words and then calculate the usage ratio of each POS for every user.

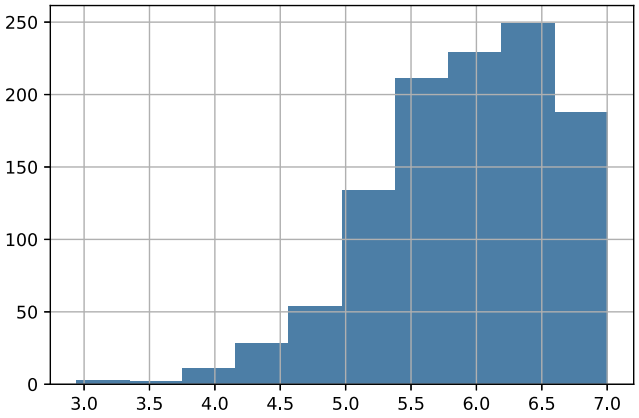
4. Data Collection with Online Experiment

To collect a comprehensive data set, we conduct an online experiment to collect individuals' privacy preference, their private information (e.g., demographic characteristics), and their public data (i.e., publicly available social media posts) under their agreement. We run the experiment on Prolific, which is similar to Amazon Mechanical Turk in that it is an on-demand platform that enables large-scale data collection by connecting researchers to participants around the globe. Multiple strategies and attention checks are employed, such as requiring a minimum approval rating of 97%, to ensure that our participants deliver high-quality responses.

To capture the individual's privacy preference, we adopt a widely used and reliable privacy preference scale—the IUIPC scale (Malhotra et al. 2004). The IUIPC scale has been employed in numerous studies and across various fields such as marketing, information systems, and psychology, indicating its widespread acceptance for its high reliability and cross-cultural applicability. This privacy preference construct consists of three first-order dimensions—namely, collection, control, and awareness—exhibiting desirable psychometric properties in the context of online privacy. Each dimension has three or four statements, and we ask users to rate these statements using seven-point scales anchored with "strongly disagree" and "strongly agree." In addition, we collect their Twitter account username and their full publicly available social media posts under their agreement. Finally, we also collect private information such as age, gender, race, educational background, household income, marital status, employment status, what industry they work in, etc. Please refer to the Online Appendix for a full description of our experiment design.

We aggregate three second-order privacy preferences into overall preference metrics. Specifically, we take the average of three second-order privacy preferences and use the mean as the overall preference. The distribution of privacy preferences is shown in Figure 1. We observe a diverse distribution of collected privacy preferences, which indicates a broad distribution of people with varying levels of privacy preference. Some people exhibit high privacy concerns, whereas others have low concerns. The privacy preference score ranges from 2.94 to 7, with 2.94 representing an individual with relatively low privacy concerns, and 7 representing an individual with high privacy concerns. We collected data from 1,109 users, encompassing a total of 402,400 tweets publicly available on their accounts.[2] Overall, we extracted 65 features for demographic characteristics and 251 features for psychosocial traits.

Figure 1. (Color online) Privacy Preference Distribution in Collected Data Set



Note. The *x*-axis represents the privacy preference score and the *y*-axis represents the number of users.

5. Experimental Analysis to Investigate the Predictive Power of Public Data

We consider the extracted psychosocial traits from social media posts as public data and collected demographic characteristics as private information. In this section, we employ predictive analytics to examine the predictive capabilities of public data and compare them with those of private information.

We use a widely used machine learning model, LightGBM (Ke et al. 2017), which is an open-source, gradient boosting framework that uses tree-based learning algorithms to perform machine learning tasks such as classification, regression, and ranking. It is designed to be highly efficient, scalable, and fast. We split the data to 70% as training set and 30% as test set. We conduct five-fold cross-validation on the training set to tune hyperparameters, and we report the performance on the test set as a model comparison. We experiment with three conditions: (i) demographic characteristics only, (ii) psychosocial traits only, and (iii) demographic characteristics plus psychosocial traits. We measure predictive performance using the mean squared error (MSE):

$$\text{MSE} = \frac{1}{M}\sum_{i=1}^{M}(y_i - \hat{y}_i)^2 \tag{3}$$

where M denotes the number of users.

The model comparison results are summarized in Table 3. Overall, the feature set of demographic characteristics plus psychosocial traits performs the best with an MSE of 0.4542. In addition, psychosocial traits performs better than the demographic characteristics. This indicates that the public data features exhibit higher predictive power than private information. We further illustrate this pattern by quantifying the power of each feature group. First, private information provides a +7.02% MSE gain when we compare models (iii) and (ii). Moreover, public data provide a +9.63% MSE gain when we compare models (iii) and (i). To test the *statistical significance*, we conduct the paired t-test.[3] Both private information and public data provide significant performance gain ($p < 0.005$).

To summarize, this analysis suggests that we can predict an individual's privacy preference without access to their private data, relying solely on insights gained from their social media activity. The potential reason for the effectiveness of our approach is that compared with the kind of demographic characteristics obtained from private data, the psychosocial traits obtained from social media are more directly related to an individual's psychological makeup and personal experiences, which can have a greater influence on their attitudes and behaviors related to privacy. Demographic characteristics, such as age, gender, and income, may be useful indicators of privacy preference in some cases, but they do not capture the full complexity of individual differences in privacy attitudes and behaviors. Psychosocial traits, on the other hand, can provide deeper insights into an individual's motivations, values, and beliefs, which can help to explain their privacy preferences more accurately.

6. Uncovering the Model Performance with Interpretability Analysis

In this section, we conduct multiple in-depth interpretability analyses to understand what drives our model's performance. We perform these analyses at various levels of granularity, including the population level, feature group level, and individual level. This allows us to deeply understand what drives the predictions and also provides valuable insights to practitioners for future use.

6.1. Population-Level Analysis: How Does the Importance of Psychosocial Traits Vary for Different Population Subgroups?

Whereas our previous model comparisons primarily display the average predictive power for all populations, this analysis delves into the importance of psychosocial traits across different population subgroups or scenarios, such as different age groups. Understanding how these traits play out across various subgroups can lead to a nuanced understanding of privacy preferences and model performance.

SHapley Additive exPlanations (SHAP) are used to explain supervised learning models, by quantifying the impact of a particular feature or group of features on the model outcome (Lundberg and Lee 2017).[4] By aggregating SHAP values of a feature across observations (mean of absolute values), one can explain the global impact of the features across all data. To obtain a measure of the relative global impact of all psychosocial traits or demographic characteristics, we aggregate SHAP values not only across observations but also across features within the group in a specific trained model. We compute the mean of feature importance within the feature group to control for the feature size difference in various groups.

We then analyze the relative importance ratio of psychosocial traits to demographic characteristics under

Table 3. Model Performance Comparison for Different Feature Sets

Model	Privacy preference MSE	Predictive power	p-value
Demographic characteristics only	0.5026	+7.02%	$p < 0.005$
Psychosocial traits only	0.4885	+9.63%	$p < 0.005$
Demographic characteristics + psychosocial traits	0.4542	—	—

different subgroups, like age, gender, and marital status. The results are shown in Table 4. We find that psychosocial traits are more important for predicting privacy preference in young people (i.e., age 18–34) compared with relatively older people (i.e., over 35). The possible reason for this dynamic is that younger generations have grown up in a digitally interconnected world where self-expression and identity exploration are paramount. Therefore, intrinsic characteristics like personality, risk preference, and lifestyle potentially play a significant role in shaping their views on privacy. In contrast, those over 35, who grew up with less digital exposure, are more likely to be influenced by demographic factors and societal norms from their formative years.

What's more, psychosocial traits are more important for predicting privacy preference in high-income people (i.e., over 50k) compared with low-income people (i.e., less than 50k). High-income individuals, who are often more educated or technologically engaged, might see privacy as a nuanced extension of personal autonomy. Conversely, low-income individuals, possibly preoccupied with more immediate concerns, might have their privacy preferences more closely tied to basic demographic characteristics. This distinction underscores the complex interplay between economic status and psychosocial factors in shaping digital privacy perceptions.

In addition, psychosocial traits are more important for predicting privacy preference in full-time employed users compared with non-full-time-employed users. Full-time workers are often deeply entrenched in professional settings that demand constant self-presentation, negotiation, and boundary setting, both offline and online. Their psychosocial traits, such as their personality traits, risk preferences, and lifestyles, can significantly impact how they manage their professional image and personal boundaries, which in turn affects their privacy preferences. On the other hand, non-full-time employees might not be as exposed to such professional settings, so their privacy preferences may be influenced more by broad demographic features than nuanced psychosocial traits.

6.2. Feature Group-Level Analysis: Which Psychosocial Trait Features Drive the Prediction?

In this section, we uncover the importance of each group of psychosocial traits. Whereas our earlier models presented the overall predictive power of these traits, this analysis dives deeper to specify the importance of each trait group. This can help us understand which traits are most important in terms of predicting privacy preference in our model. To obtain the importance of different trait groups, we take the mean of SHAP feature importance within each individual trait group, and we report the impact of feature groups in Figure 2.

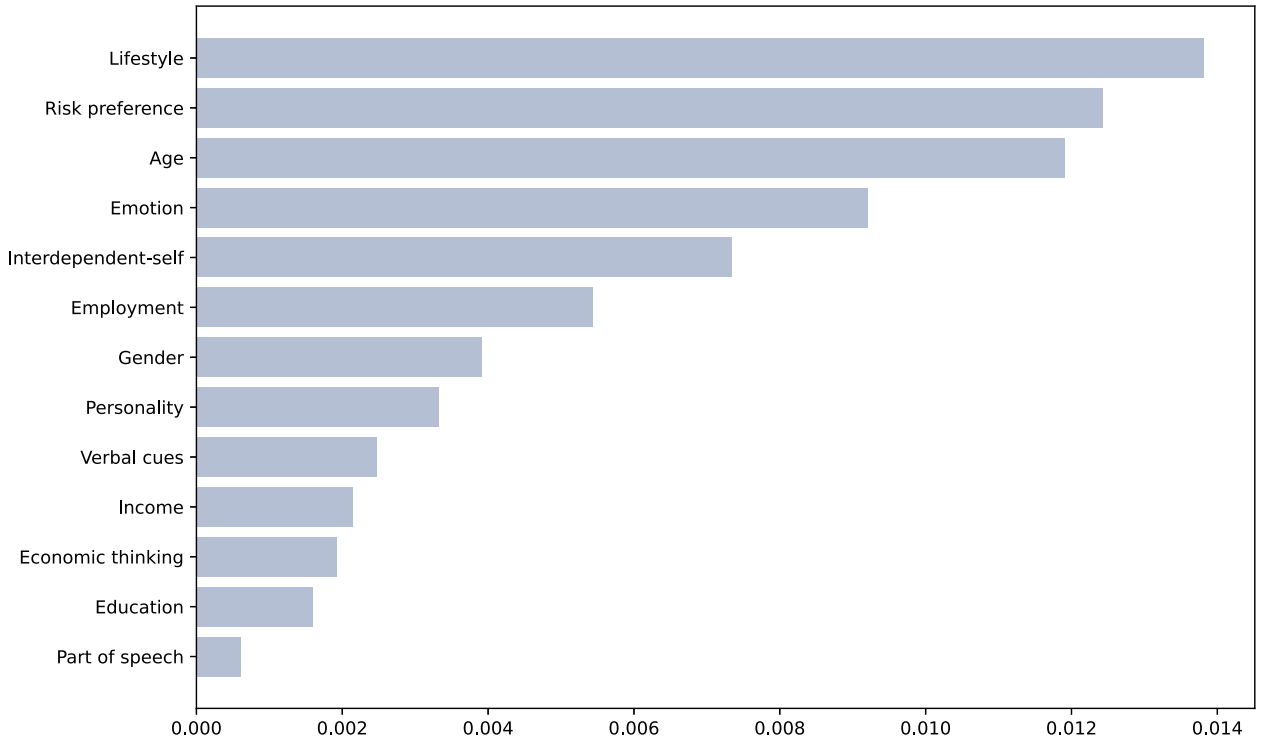
We find that lifestyle, risk preference, and emotion states are the most important psychosocial traits for predicting privacy preference in our trained model. Lifestyle is a holistic reflection of an individual's life, encompassing their habits, behaviors, and daily choices. This covers a variety of aspects ranging from social activities to work habits. The potential reason for the high importance of lifestyle is that lifestyle or life choices usually reflect a person's deep and intrinsic values, beliefs, and motivations, which collectively play a significant role in shaping one's perspective and preferences regarding privacy.

Risk preference also plays a pivotal role in shaping an individual's privacy preference. The high importance of risk preference can possibly be attributed to its direct impact on risk perception and estimation in privacy decisions. Risk-averse individuals often perceive the possible dangers of sharing personal information with heightened sensitivity, prioritizing the protection of their data because of fears of misuse or breaches. For them, the

Table 4. The Variations of Importance of Psychosocial Traits Under Different Population Subgroups

Feature	Subgroup	Relative importance ratio (psychosocial/demographic)
Age	18–34	2.3076
	Over 35	0.8130
Gender	Female	2.5924
	Male	4.4673
Marriage	Single	7.3010
	Have ever married	1.1920
Education	Less educated	6.9128
	Highly educated	2.4411
Household income	Less than 50k	3.0486
	Over 50k	6.2234
Employment	Full-time	5.8390
	Non-full-time	1.9790
Working sectors	First and second	1.6505
	Tertiary	3.2463
	Quaternary	3.4569

Figure 2. (Color online) Feature Importance of Each Group of Psychosocial Traits



risks of sharing often overshadow the immediate benefits. Conversely, risk-seeking individuals tend to be more open to sharing personal information, prioritizing immediate benefits like convenience or personalized experiences over potential future hazards. They might view the consequences of data misuse as distant or unlikely, leading them to exercise less-stringent control over their data.

Emotional states also significantly influence individuals' privacy preferences, shaping people's attitudes and behaviors. Intense emotions, like fear, anger, or anxiety, often prompt individuals to prioritize and safeguard their privacy. Conversely, positive feelings, such as trust and excitement, can make people more open to sharing personal information. Therefore, including emotions in a model can offer deep insight into individual decision-making processes.

In addition, the interdependent self (i.e., social interactivity) is an important set of psychosocial traits for predicting privacy preferences. If a user shows a high level of social interaction, as evidenced by numerous comments, retweets, likes, and followers, this may indicate their real-life social network or friendships and reflect their social openness. Therefore, levels of social interactivity can be insightful for understanding privacy preferences. Users who frequently interact with others may hold a more open view on privacy. Conversely, those with low interactivity might adopt a more reserved stance on privacy.

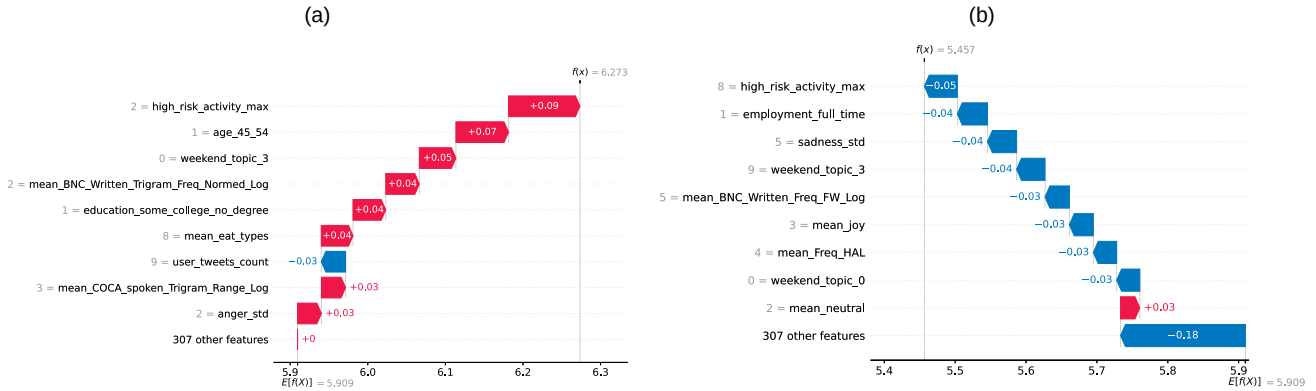
In comparison, demographic features are generally less important than psychosocial traits. Out of the top five important feature groups, only one (age) belongs to the demographic group, whereas the other four are psychosocial traits. Among the demographic features, age, employment status, and gender stand out as the most important. Age is likely to be associated with varying comfort levels with technology, which in turn shapes users' online privacy expectations. Employment can determine one's sensitivity to information exposure, with some professions demanding greater discretion. Furthermore, gender plays a role and women have often been more susceptible to online threats, which affects their privacy concerns.

6.3. Individual-Level Analysis: Understanding Individual Users for Practical Applications

In this section, we delve deeper into individual users and show practitioners how to leverage SHAP analysis to understand each individual. This can help platforms better serve each individual, enhancing their online experience and sense of control. We randomly sample two users from our data and plot their corresponding SHAP values in Figure 3. It shows features each contributing to pushing the model output from the base value to the model output. Essentially, it quantifies the amount and direction in which each variable impacts the predicted privacy preference. Features pushing the prediction higher are shown in red, and those pushing the prediction lower are in blue.

Example 1 for a specific user shows that the individual's low value of ever participating in high-risk activities (i.e., high-risk activity max) pushes the model to

Figure 3. (Color online) SHAP Value Plot Examples for Two Users



Notes. (a) Example 1. (b) Example 2.

predict +0.09 higher privacy concern. A low value of diverse nightlife style propensity (i.e., weekend topic 3 probability) pushes the model to predict +0.05 higher privacy preference. Being aged 45–54 pushes the model to predict +0.07 higher privacy concern. These findings suggest that for this specific user, risk aversion, a nondiverse lifestyle, and age primarily drive the predicted high privacy concern. Individuals with this set of characteristics usually have a limited range of activities, a smaller social circle or community, and a stronger attachment to routine and familiarity. These factors make them more concerned about the potential negative consequences of sharing personal information and make them more resistant to change.

Example 2 for another user shows that a high value of ever participating in high-risk activities (i.e., high-risk activity max) pushes the model to predict −0.05 lower privacy concern. Being employed full-time pushes the model to predict −0.04 lower privacy concern. A high value of sadness variability (i.e., sadness std) pushes the model to predict −0.04 lower privacy concern. A high value in diverse nightlife style propensity (i.e., weekend topic 3 probability) pushes the model to predict −0.04 lower privacy concern. These findings suggest that for this specific user, a risk-seeking trait, full-time employment, high variability in sadness, and a diverse nightlife style are the primary drivers of the predicted low privacy concern. The diversity of nightlife activities and risk-seeking propensity suggest openness beliefs. Being employed full-time implies a stable environment where the user might feel more secure about their personal data. High variability in sadness might indicate a willingness to share feelings and experiences, leading to a more relaxed attitude toward privacy. When combined, all these factors make the individual less apprehensive about potential repercussions of sharing personal information and more adaptable to change.

7. Decision Support Showcase: Forecasting Consequences of Privacy Policies for Online Platforms

In this section, we demonstrate the practical value and impact of our model for both businesses and society. We demonstrate how our model can assist platforms and policymakers in forecasting the consequences of privacy policies ahead of time, thereby ensuring regulatory compliance, fostering user trust and transparency, and averting potentially detrimental effects.

In particular, our model can gauge the influence of privacy policies on a platform's shift in user base ahead of time, thus addressing a major concern for online platforms. Specifically, we simulate policy shocks and see how the platform reacts through counterfactual prediction. We simulate a platform with 10,000 users using a Gaussian mixture model (GMM). The users are generated based on the actual user distribution we collected through our online experiment. We choose to do this because many features exhibit latent relationships. For instance, a specific age group might be inclined toward certain lifestyles, or a distinct personality type may have a particular risk preference. Neglecting to account for these latent associations could lead to flawed or unrealistic simulations, especially if certain combinations of features don't actually exist in the real world. Hence, we employ GMM to learn the joint distribution of all features from real users in our data set, ensuring that all latent relationships between features are captured. Only with this method can our simulations be meaningful and accurately reflect the real world.

GMM is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters.[5] We train a GMM with our collected data. Then we simulate 10,000 users on a platform according to our data distribution. Given the simulated users' features,

we use our trained model to predict all users' privacy preferences.

Then, we simulate three policy shocks with increasing levels of privacy risk (low, medium, and high) as follows:

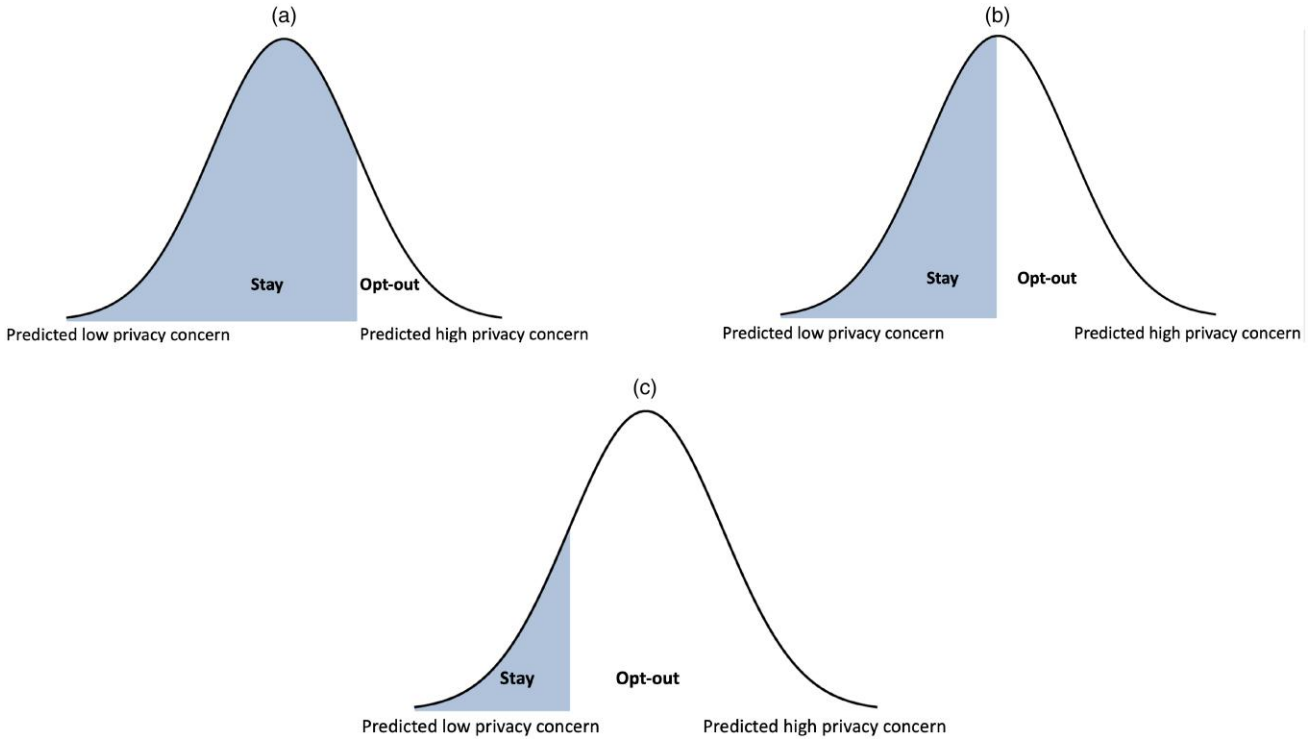
- Policies leading to low privacy risk: These policies generally incorporate more conservative and stringent guidelines. These may include limiting data collection to only what's essential, restricting the sharing of information with third parties, and enforcing strict data retention and deletion policies.
- Policies leading to high privacy risk: These policies involve more liberal and less restrictive protocols. For instance, they may encompass broad data collection practices, the sharing of information with third parties, and extended periods of data retention.
- Policies leading to medium privacy risk: These policies involve intermediate protocols situated between high-risk and low-risk policies.

We then propose the hypothetical platform response to these three policies, as summarized in Figure 4. In the case of a policy shock leading to low privacy risk, only the users with predicted high privacy concern will opt out, and the majority of users will stay. For a policy shock leading to medium privacy risk, half of users will opt out and half will stay. For a policy shock leading to high privacy risk, only the users with low privacy concern will stay, and most of the users will opt out. Under these three scenarios, we quantify the consequences of privacy policy shock and observe the platform's corresponding user shifts. We simulate three levels of risk by designating risk levels that correspond with the 25%, 50%, and 75% thresholds of the user population's privacy preference scores.

First, we visualize the platform's user shifts from a psychosocial trait perspective as shown in Figures 5 and 6. We find that when policy risk increases, the platform has more users with diverse weekend and weekday lifestyles (e.g., late-night diverse lifestyle) and fewer users with nondiverse lifestyles (e.g., weekend topic 0: working dominant). We find a consistent phenomenon with regard to risk-preference metrics: when the policy risk increases, the platform will have more exploratory consumers (i.e., consumers with high activity diversity) and risk-seeking consumers (i.e., consumers with high propensity to participate in high-risk activities). The possible reason for this is that people with diverse lifestyles and risk-seeking traits tend to have lower privacy concern, as they are inherently more accepting of potentially negative outcomes in exchange for potential rewards. That being said, these users might downplay the risk of a privacy breach in favor of the immediate benefits a platform or service provides.

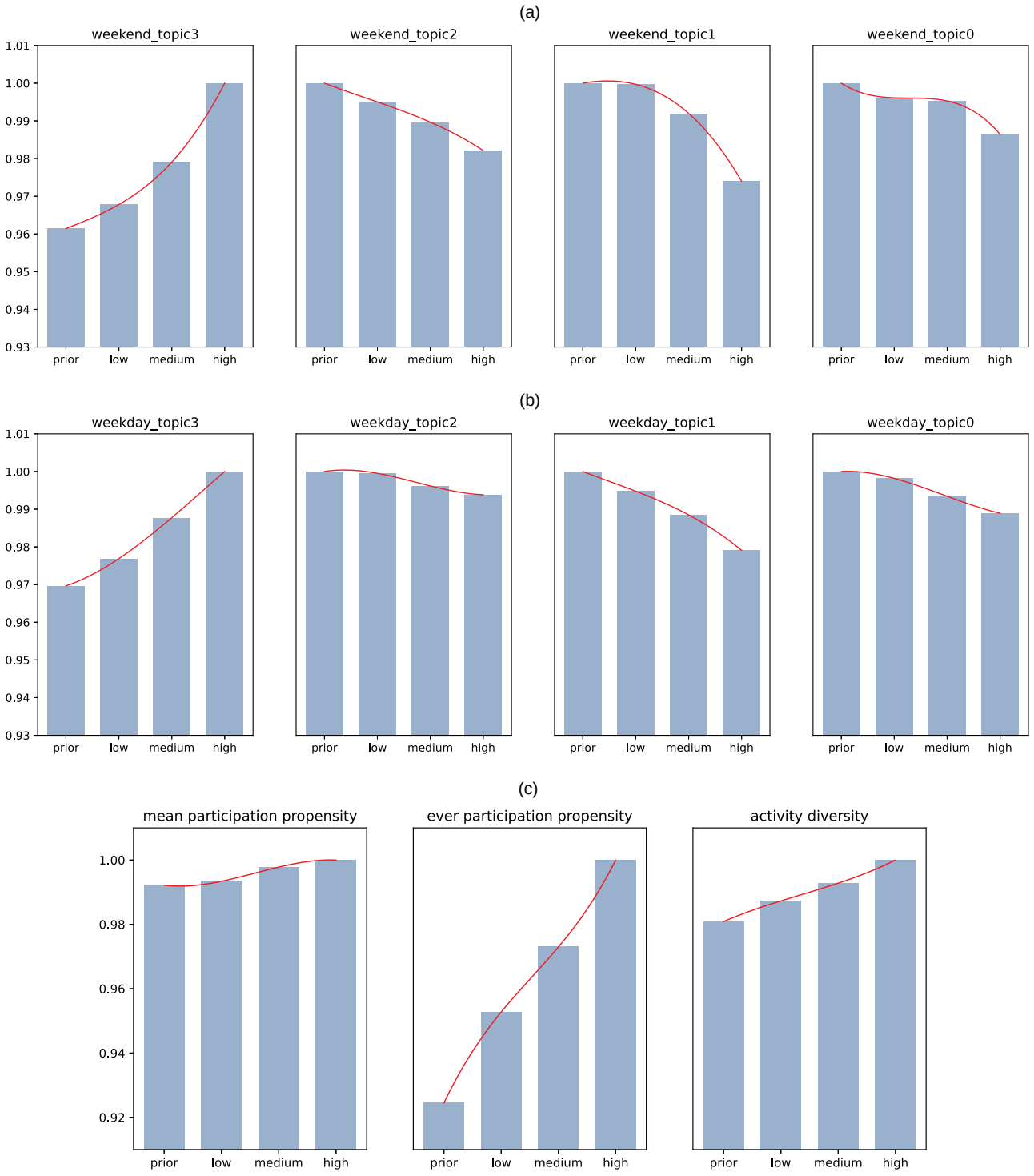
Second, regarding personality, as the privacy risk increases, the platform is likely to have more consumers characterized by agreeableness and openness and fewer with traits of neuroticism and conscientiousness. This is

Figure 4. (Color online) Platform Population Shift Under Various Policy Shocks with Low, Medium, and High Risk



Notes. (a) Low-risk policy shock. (b) Medium-risk policy shock. (c) High-risk policy shock.

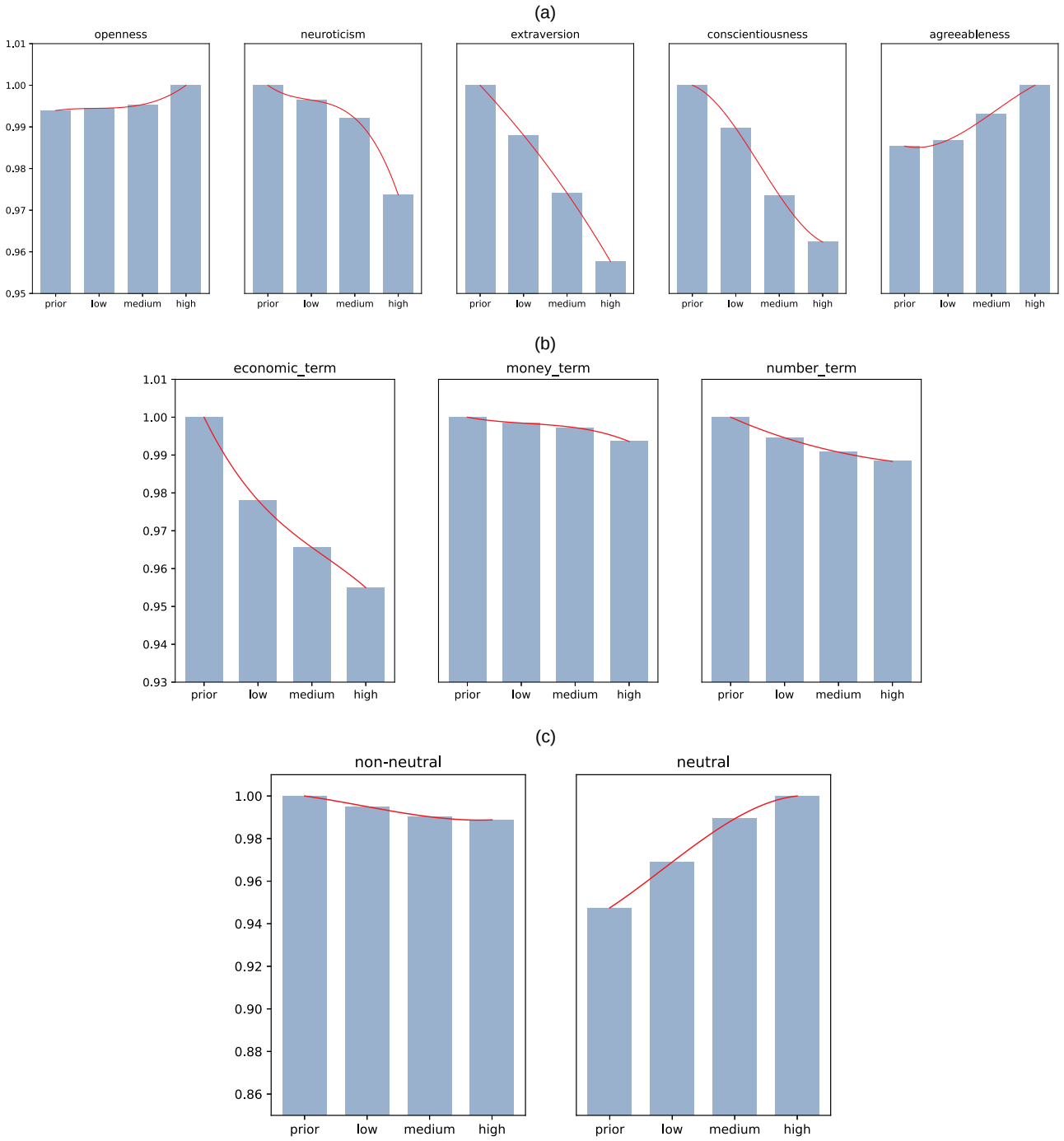
Figure 5. (Color online) Platform Population Shift Under Various Policy Shocks—Psychosocial Trait Perspective Part 1



Notes. *y* axis refers to percentage of users normalized by max value. (a) Weekend topics. (b) Weekday topics. (c) Risk preference.

because individuals with high agreeableness and openness often have lower privacy concern. Such individuals are typically characterized by trust, altruism, kindness, and affection. Their belief in the goodwill of others makes them less concerned about the potential misuse of their information. Conversely, individuals with high neuroticism experience emotions like anxiety, worry, and fear. Their heightened sensitivity to negative stimuli

Figure 6. (Color online) Platform Population Shift Under Various Policy Shocks—Psychosocial Trait Perspective Part 2



Notes. *y* axis refers to percentage of users normalized by max value. (a) Personality. (b) Economic thinking. (c) Emotion status.

might make them more concerned about their privacy. Similarly, conscientious individuals, being organized, responsible, and detail oriented, are more aware of the possible risks when sharing personal information online, leading to heightened privacy concern.

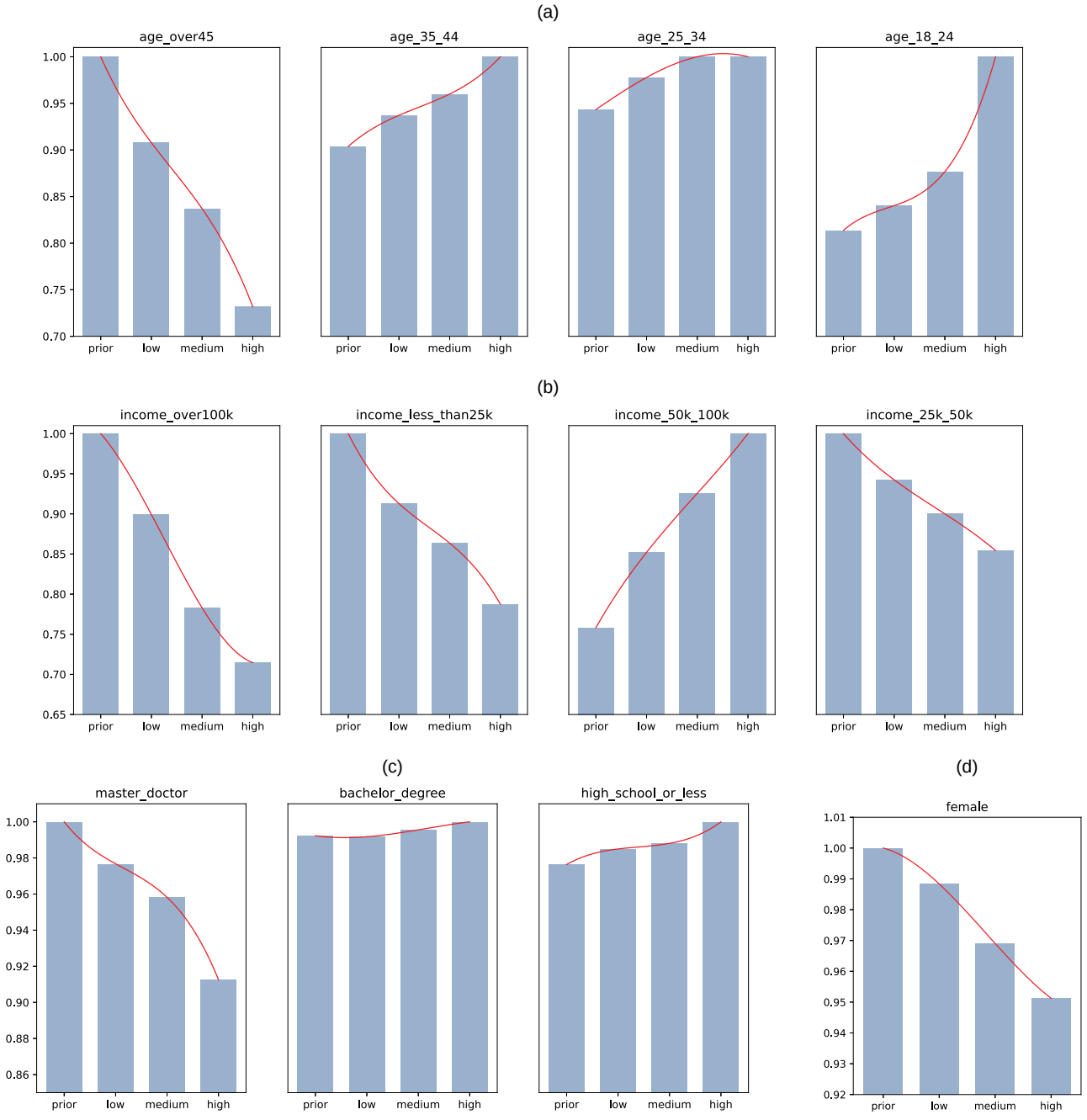
As privacy risk increases, platforms tend to observe more neutral emotions and fewer nonneutral emotions among users. One potential explanation is that individuals exhibiting neutral emotions may have lower privacy concern. They might process information more rationally, maintain a realistic assessment of risks, and trust in their ability to manage these risks. On the other hand, individuals showing nonneutral emotions might be more attuned to their feelings and personal experiences, making them more sensitive to privacy concerns. Their strong emotional responses

could be influenced by past breaches of trust, or they may have deep-seated feelings about the value of personal privacy.

Furthermore, as privacy risks increase on the platform, there appears to be a decline in its economically minded user base. This observation is based on the reduced usage of financial-economic-oriented words and quantitative expressions, such as "money" and "numbers." One possible explanation is that individuals with strong economic thinking might have heightened privacy concern. These individuals tend to be more analytical and reflective in their decision-making processes, which may lead them to be more sensitive to the potential risks and consequences of sharing personal information.

We then quantify the platform's user shift from the demographic characteristic perspective in Figure 7. Interestingly, as privacy policy risk increases, we observe a significant decrease in the platform's female user base.

Figure 7. (Color online) Platform Population Shift Under Various Policy Shocks with Low, Medium, and High Risk—Demographic Characteristic Perspective



Notes. *y* axis refers to percentage of users normalized by max value. (a) Age. (b) Income. (c) Education. (d) Gender.

This suggests that a policy with high privacy risk could potentially lead to discriminatory outcomes and significantly decrease the female user base on a platform. By utilizing our tool, platforms can modulate their risk levels and strike a balance to avoid possible discrimination.

We also note that as privacy risk increases, platforms have fewer older users (i.e., age over 45) and more young users. Older users often lack familiarity with digital privacy practices and are more concerned about the potential misuse of personal data. In contrast, younger users, although generally more tech savvy, may overlook privacy risks because of the perceived benefits of platform usage.

In terms of education, we observe that with increased privacy risk, the platform shows a sharp drop in users with master's and doctorate degrees and a mild increase in users with a high school education or lower. The reason for this may be that individuals with more education tend to exhibit heightened privacy concern because of increased awareness and understanding of data misuse and security risks. Their academic exposure often includes topics related to digital literacy and the nuances of personal information vulnerability in the digital age. On the other hand, those with less formal education may lack this specific awareness or may place other immediate concerns over data privacy, perceiving the benefits from freely sharing information without fully recognizing the possible long-term implications.

To summarize, the kind of foresight our model offers would allow platforms to evaluate the potential outcomes of policy changes, prepare for shifts in user base, and avoid unanticipated consequences, thereby fostering a more thoughtful and responsive approach to privacy policy development.

Additionally, we conduct a real user study to complement our simulation and validate the practical relevance of our model. This study demonstrates that the learned privacy preference score from our model can be informative to real users' perception of privacy policy, reinforcing its practical utility. For details, please see the Online Appendix.

8. Conclusion

This study proposes a novel framework to predict personalized privacy preferences using public data. We have demonstrated the feasibility of learning a broad range of theory-driven psychosocial traits from publicly available tweets using various natural language processing algorithms. We show that these psychosocial traits derived from public data offer greater predictive power than private information. This suggests that even without access to an individual's private details or explicit user input, we can accurately predict their privacy preferences based solely on public data. Furthermore, we highlight the practical value of our model for both businesses and society at large by conducting multiple in-depth interpretability analyses and providing a decision support showcase to assist platforms and policymakers in forecasting the consequences of privacy policies.

Our study carries fruitful managerial implications. First, we take a small step in offering an alternative method to automatically assess users' privacy preferences by predicting them from a publicly available organic data source: Twitter posts. Existing methods commonly used to learn about user privacy preferences, such as surveys or experimental designs, are time-consuming and difficult to scale. Our proposed approach makes it possible to automatically learn about privacy preferences from users' organic posts, where these users can proactively articulate their true preferences without external distortions. This alternative approach could also be used with other variables of interest to better understand privacy decision making on a broader scale in the future.

Second, our framework possesses high generalizability and opens a new door for many publicly available and organic data sources in the field of privacy research. In fact, our study is not limited to Twitter alone. It can easily be generalized to other platforms such as Reddit, LinkedIn, and Facebook, where users share their thoughts and experiences publicly.

Third, our model offers an automated tool for estimating an individual's privacy preferences without the need for user input. This tool can help platforms serve individuals better, enhancing users' online experience and sense of control. It also gives platforms and policymakers the capability to anticipate the impact of privacy policies before their implementation.

Our study also has some limitations, and it points to fruitful avenues for future research. Our method operates on users who have Twitter accounts and actively post tweets publicly. However, it does not cater to users without a Twitter account or those with minimal Twitter activity, as our framework relies on Twitter posts to extract psychosocial traits. Future research could develop a more generic approach that takes into account these missing users, including those without a Twitter account and those with low levels of activity, when designing the framework.

Endnotes

[1] For instance, some recent studies, such as those by Marreiros et al. (2017) and Xu and Zhang (2022), have identified the "now you mention it" effect. This phenomenon occurs when the mere mention of "privacy" in an experiment or survey instantly increases the participants' privacy concerns, potentially skewing the data. In contrast, Twitter data represent an "organic" form of data, generated spontaneously by users without the influence of explicit research methodologies. This organic nature ensures that the data reflect users' authentic privacy preferences without the distortion that might come from the artificial conditions of designed studies.

[2] We remove all tweets starting with "RT" as this indicates a retweet. Instead, we focus on organic tweets written by the users

themselves because such tweets can better reflect users' intrinsic traits. We retain users with over 50 tweets in their accounts. This ensures we have ample data about each user over a relatively long period, providing a more comprehensive view of each individual. All extracted psychosocial traits are normalized by KBinsDiscretizer.

[3] The paired *t*-test has been widely used in computer science literature to compare two models. It is specifically designed for situations where the data points in two groups can be paired in a meaningful way. In our case, the error from each sample when using the baseline model can be directly paired with the error from the same sample when using our proposed model. This means that any variability between samples can be controlled for, allowing for a more precise estimate of the difference in performance between the two models.

[4] Please note the learned feature importance is model dependent. That said, it assesses the importance of a feature to the prediction model being analyzed.

[5] One important hyperparameter of GMM is the number of components, which represents the number of Gaussian distributions that are used to model the data. The best number of components is 14, so we train a GMM with this selected hyperparameter.

References

Acquisti A, Gross R (2009) Predicting social security numbers from public data. *Proc. Natl. Acad. Sci. USA* 106(27):10975–10980.

Acquisti A, Brandimarte L, Hancock J (2022) How privacy's past may shape its future. *Science* 375(6578):270–272.

Acquisti A, Brandimarte L, Loewenstein G (2015) Privacy and human behavior in the age of information. *Science* 347(6221): 509–514.

Adamopoulos P, Ghose A, Todri V (2018) The impact of user personality traits on word of mouth: Text-mining social media platforms. *Inform. Systems Res.* 29(3):612–640.

Adjerid I, Acquisti A, Loewenstein G (2019) Choice architecture, framing, and cascaded privacy choices. *Management Sci.* 65(5): 2267–2290.

Adjerid I, Peer E, Acquisti A (2018) Beyond the privacy paradox: Objective vs. relative risk in privacy decision making. *MIS Quart.* 42(2):465–488.

Alashoor T, Keil M, Smith HJ, McConnell AR (2022) Too tired and in too good of a mood to worry about privacy: Explaining the privacy paradox through the lens of effort level in information processing. *Inform. Systems Res.* 34(4):1415–1436.

Alshehri A, Alotaibi F (2019) Predicting users mobile app privacy preferences. *J. Comput. Comm.* 7(10):147–156.

Balapour A, Nikkhah HR, Sabherwal R (2020) Mobile application security: Role of perceived privacy as the predictor of security perceptions. *Internat. J. Inform. Management* 52:102063.

Bansal G, Zahedi F, Gefen D (2010) The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. *Decision Support Systems* 49(2):138–150.

Barbosa NM, Park JS, Yao Y, Wang Y (2019) "What if?" Predicting individual users' smart home privacy preferences and their changes. *Proc. Privacy Enhancing Tech. Sympos.* 2019(4):211–231.

Bartel Sheehan K (1999) An investigation of gender differences in on-line privacy concerns and resultant behaviors. *J. Interactive Marketing* 13(4):24–38.

Barth S, De Jong MD (2017) The privacy paradox–investigating discrepancies between expressed privacy concerns and actual online behavior–A systematic literature review. *Telematics Informatics* 34(7):1038–1058.

Belanger F, Hiller JS, Smith WJ (2002) Trustworthiness in electronic commerce: The role of privacy, security, and site attributes. *J. Strategic Inform. Systems* 11(3–4):245–270.

Berendt B, Günther O, Spiekermann S (2005) Privacy in e-commerce: Stated preferences vs. actual behavior. *Comm. ACM* 48(4):101–106.

Bigwood G, Abdesslem FB, Henderson T (2012) Predicting location-sharing privacy preferences in social network applications. *Proc. First Workshop Recent Adv. Behav. Prediction Pro-active Pervasive Comput. (AwareCast)*, 1–12.

Bonde JPE (2008) Psychosocial factors at work and risk of depression: A systematic review of the epidemiological evidence. *Occupational Environ. Medicine* 65(7):438–445.

Bongers PM, de Winter CR, Kompier MA, Hildebrandt VH (1993) Psychosocial factors at work and musculoskeletal disease. *Scandinavian J. Work Environ. Health* 19(5):297–312.

Brandão A, Mendes R, Vilela JP (2022) Prediction of mobile app privacy preferences with user profiles via federated learning. *Proc. Twelfth ACM Conf. Data Appl. Security Privacy* (ACM, New York), 89–100.

Chen K, Rea AI Jr (2004) Protecting personal information online: A survey of user privacy concerns and control techniques. *J. Comput. Inform. Systems* 44(4):85–92.

Culnan MJ (1995) Consumer awareness of name removal procedures: Implications for direct marketing. *J. Direct Marketing* 9(2): 10–19.

Culnan MJ, Armstrong PK (1999) Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organ. Sci.* 10(1):104–115.

Dong C, Jin H, Knijnenburg B (2015) Predicting privacy behavior on online social networks. *Proc. Internat. AAAI Conf. Web Social Media*, vol. 9 (AAAI Press, Palo Alto, CA), 91–100.

Easwara Moorthy A, Vu KPL (2015) Privacy concerns for use of voice activated personal assistant in the public space. *Internat. J. Human Comput. Interaction* 31(4):307–335.

Everson-Rose SA, Lewis TT (2005) Psychosocial factors and cardiovascular diseases. *Annual Rev. Public Health* 26:469–500.

Ghose A, Ipeirotis PG, Li B (2012) Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Sci.* 31(3):493–520.

Goodwin C (1991) Privacy: Recognition of a consumer right. *J. Public Policy Marketing* 10(1):149–166.

Hartmann J (2022) Emotion English DistilRoBERTa-base. https://huggingface.co/j-hartmann/emotion-english-distilroberta-base/.

Hochheiser H (2002) The platform for privacy preference as a social protocol: An examination within the US policy context. *ACM Trans. Internet Tech.* 2(4):276–306.

Hong W, Chan FK, Thong JY (2021) Drivers and inhibitors of Internet privacy concern: A multidimensional development theory perspective. *J. Bus. Ethics* 168(3):539–564.

Hoogendoorn WE, Van Poppel MN, Bongers PM, Koes BW, Bouter LM (2000) Systematic review of psychosocial factors at work and private life as risk factors for back pain. *Spine* 25(16): 2114–2125.

IBISworld (2020) Educational services in the US market size 2005–2026. Retrieved August 24, https://www.ibisworld.com/industry-statistics/market-size/educational-services-united-states/.

Jai TMC, King NJ (2016) Privacy vs. reward: Do loyalty programs increase consumers' willingness to share personal information with third-party advertisers and data brokers? *J. Retailing Consumer Services* 28:296–303.

Jiang Z, Heng CS, Choi BC (2013) Research note—Privacy concerns and privacy-protective behavior in synchronous online social interactions. *Inform. Systems Res.* 24(3):579–595.

Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY (2017) LightGBM: A highly efficient gradient boosting decision tree. Guyon I, Von Luxburg U, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, eds. *Adv. Neural Inform. Processing Systems*, vol. 30 (Curran Associates Inc., Red Hook, NY), 3149–3157.

Klopfer PH, Rubenstein DI (1977) The concept privacy and its biological basis. *J. Soc. Issues* 33(3):52–65.

Kyle K, Crossley SA (2015) Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quart.* 49(4):757–786.

Lee HS (2008) Selected antecedents of customer loyalty within a restaurant loyalty program: Perceived control, privacy concern, perceived value of a loyalty program, and willingness to disclose information. Unpublished doctoral dissertation, The Pennsylvania State University, State College, University Park, PA.

Li R, Ghose A, Xu K, Li B (2023) Predicting consumer in-store purchase using real-time retail video analytics. Preprint, submitted July 18, http://dx.doi.org/10.2139/ssrn.4513385.

Li H, Luo XR, Zhang J, Xu H (2017) Resolving the privacy paradox: Toward a cognitive appraisal and emotion approach to online privacy behaviors. *Inform. Management* 54(8):1012–1022.

Liu X, Singh PV, Srinivasan K (2016) A structured analysis of unstructured big data by leveraging cloud computing. *Marketing Sci.* 35(3):363–388.

Lu Y, Tan B, Hui KL (2004) Inducing customers to disclose personal information to Internet businesses with social adjustment benefits. *Proc. Internat. Conf. Inform. Systems (ICIS)* (AIS eLibrary, Atlanta).

Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. Guyon I, Von Luxburg U, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, eds. *Adv. Neural Inform. Processing Systems*, vol. 30 (Curran Associates Inc., Red Hook, NY), 4768–4777.

Malhotra NK, Kim SS, Agarwal J (2004) Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Inform. Systems Res.* 15(4):336–355.

Maree JG (2021) The psychosocial development theory of Erik Erikson: Critical overview. *Early Child Development Care* 191(7–8):1107–1121.

Marreiros H, Tonin M, Vlassopoulos M, Schraefel M (2017) "Now that you mention it": A survey experiment on information, inattention and online privacy. *J. Econom. Behav. Organ.* 140:1–17.

Mendes R, Cunha M, Vilela JP, Beresford AR (2022) Enhancing user privacy in mobile devices through prediction of privacy preferences. *Comput. Security ESORICS 2022 27th Eur. Sympos. Res. Comput. Security Proc. Part I* (Springer, Berlin), 153–172.

Metzger MJ (2007) Communication privacy management in electronic commerce. *J. Comput. Mediated Comm.* 12(2):335–361.

Milberg SJ, Smith HJ, Burke SJ (2000) Information privacy: Corporate management and national regulation. *Organ. Sci.* 11(1):35–57.

Mullainathan S, Thaler RH (2000) Behavioral economics. NBER Working Paper No. 7948, National Bureau of Economic Research, Cambridge, MA.

Naini KD, Altingovde IS, Kawase R, Herder E, Niederée C (2015) Analyzing and predicting privacy settings in the social web. *User Model. Adaptation Personalization 23rd Internat. Conf. UMAP 2015. Proc.* (Springer, Berlin), 104–117.

Oh H, Goh KY, Phan TQ (2023) Are you what you tweet? The impact of sentiment on digital news consumption and social media sharing. *Inform. Systems Res.* 34(1):111–136.

Pavlou PA (2011) State of the information privacy literature: Where are we now and where should we go? *MIS Quart.* 35(4):977–988.

Phelps J, Nowak G, Ferrell E (2000) Privacy concerns and consumer willingness to provide personal information. *J. Public Policy Marketing* 19(1):27–41.

Quinn K (2016) Why we share: A uses and gratifications approach to privacy regulation in social media use. *J. Broadcasting Electronic Media* 60(1):61–86.

Reimers N, Gurevych I (2019) Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *Proc. 2019 Conf. Empirical Methods Natl. Language Processing* (Association for Computational Linguistics, Kerrville, TX), 3982–3992.

Schoenmueller V, Netzer O, Stahl F (2023) Frontiers: Polarized America: From political polarization to preference polarization. *Marketing Sci.* 42(1):48–60.

Serramia M, Seymour W, Criado N, Luck M (2023) Predicting privacy preferences for smart devices as norms. Preprint, submitted February 21, https://arxiv.org/abs/2302.10650.

Sheehan KB, Hoy MG (2000) Dimensions of privacy concern among online consumers. *J. Public Policy Marketing* 19(1):62–73.

Smith HJ (2001) Information privacy and marketing: What the US should (and shouldn't) learn from Europe. *Calif. Management Rev.* 43(2):8–33.

Smith HJ, Dinev T, Xu H (2011) Information privacy research: An interdisciplinary review. *MIS Quart.* 35(4):989–1015.

Steed R, Liu T, Wu ZS, Acquisti A (2022) Policy impacts of statistical uncertainty and privacy. *Science* 377(6609):928–931.

Vijayasarathy LR (2004) Predicting consumer intentions to use on-line shopping: The case for an augmented technology acceptance model. *Inform. Management* 41(6):747–762.

Villegas DS, Preoţiuc-Pietro D, Aletras N (2020) Point-of-interest type inference from social media text. Preprint, submitted September 30, https://arxiv.org/abs/2009.14734.

Wang W, Li B, Smith MD (2023) Divergent thinking and online videos: A study of TED talks via multi-modal video analytics. Preprint, submitted September 8, https://ssrn.com/abstract=4566394.

Wang Y, Qin MS, Luo X, Kou Y (2022) Frontiers: How support for Black Lives Matter impacts consumer responses on social media. *Marketing Sci.* 41(6):1029–1044.

Wang W, Zhou M, Li B, Zhuang H (2024) Predicting instructor performance in online education: An interpretable hierarchical transformer with contextual attention. *Inform. Systems Res.* Forthcoming.

Xu H (2007) The effects of self-construal and perceived control on privacy concerns. *Twenty-Eighth Internat. Conf. Inform. Systems (Montreal, Quebec)*.

Xu H, Dinev T (2022) Reflections on the 2021 Impact Award: Why privacy still matters. *MIS Quart.* 46(4):xx–xxxii.

Xu H, Teo HH (2005) Consumers' privacy concerns toward using location-based services: An exploratory framework and research proposal. *Proc. 13th Eur. Conf. Inform. Systems Inform. Systems Rapidly Changing Econom. ECIS 2005* (AIS eLibrary, Atlanta).

Xu H, Zhang N (2022) From contextualizing to context theorizing: Assessing context effects in privacy research. *Management Sci.* 68(10):7383–7401.

Xu H, Zhang N, Zhou L (2020) Validity concerns in research using organic data. *J. Management* 46(7):1257–1274.

Xu H, Teo HH, Tan BC, Agarwal R (2012) Research note—Effects of individual self-protection, industry self-regulation, and government regulation on privacy concerns: A study of location-based services. *Inform. Systems Res.* 23(4):1342–1363.

Zhang K, Moe W (2021) Measuring brand favorability using large-scale social media data. *Inform. Systems Res.* 32(4):1128–1139.

Zhang B, Xu H (2016) Privacy nudges for mobile applications: Effects on the creepiness emotion and privacy attitudes. *Proc. 19th ACM Conf. Comput. Supported Cooperative Work Soc. Comput.* (ACM, New York), 1676–1690.