Dec 12th, 12:00 AM

Solving the Social Dilemma with Equilibrium Data Harvesting Strategies: A Game-Theoretic Approach

Hyeonsik Shin
Temple University, hshin@temple.edu

Leila Hosseini
Temple University, leila.hosseini@temple.edu

Subodha Kumar
Temple University, subodha@temple.edu

Solving the Social Dilemma with Equilibrium Data Harvesting Strategies: A Game-Theoretic Approach

Completed Research Paper

Hyeonsik Shin
Temple University
1801 N Broad St,
Philadelphia, PA 19122
hshin@temple.edu

Leila Hosseini
Temple University
1801 N Broad St,
Philadelphia, PA 19122
leila.hosseini@temple.edu

Subodha Kumar
Temple University
1801 N Broad St,
Philadelphia, PA 19122
subodha@temple.edu

Abstract

Social media platforms generate huge profits from targeted advertising by collecting massive amounts of data from their users, usually referred to as data harvesting. However, practitioners from the social media industry suggest that data harvesting hurts users by promoting social media addiction and the spread of misinformation. Therefore, policymakers have recently been considering regulating social media platforms. This paper investigates how imposing the regulation on data harvesting impacts social media platforms and users by developing a game-theoretic model. Our main finding shows that while the objective of the regulation on data harvesting is to discourage platforms from collecting a massive amount of data from the users, imposing the regulation may sometimes increase the data harvesting levels and profits of social media platforms. We contribute to the Information Systems literature by broadening the knowledge of the impact of the government's regulation on social media platforms and users.

Keywords: Social Dilemma, Data Harvesting, Game-Theoretic Model, Social Media, Regulation.

Introduction

Social media platforms have been growing dramatically in recent years. For example, the worldwide total number of social media users has reached 4.62 billion as of January 2022, with a rapid growth of 10.1% over the past year (Kemp 2022). Along with the growth, the average daily time spent on social media has steadily grown over the last decade, resulting in 2.27 hours per day spent on social media in 2022 (Statista 2022). As social media has become a meaningful part of our lives, there have been increasing discussions about the influences of social media on users. In spite of some benefits being discussed (Berger 2022), a lot of recent debates focus on the downsides of social media platforms (Reinberg 2022).

One of the major downsides being widely discussed is the decrease in the subjective well-being of social media users (Allcott et al. 2022; Lambert et al. 2022). Moreover, practitioners featured in a popular documentary on Netflix, "The Social Dilemma," have confessed that the business model of social media platforms could have exacerbated the downside of these platforms (Orlowski 2020). In the documentary, former executives and developers of leading social media companies (e.g., Facebook, Twitter) argue that these platforms generate profits by exploiting the data of their users (e.g., browsing and communication histories) with a lack of ethical considerations. For example, they express concerns about social media platforms being able to manipulate how users think and act their lives (Edelman 2021) or promoting addictive behaviors (Deighton 2022). Thus, these practitioners have been calling for a regulation on social media platforms.

In response to the serious concerns about the negative consequences of social media use raised by the practitioners, the government has expressed concerns about the business model of social media platforms and proposed new regulations to implement (Conger et al. 2021). However, there is a lack of research on how the business model of social media platforms may harm their users and how social media platforms would react to new regulations. This paper attempts to address this gap, which is critical in lessening the negative impacts of social media platforms on their users.

Motivation

Social media platforms, such as Facebook and Twitter, offer online social networking services that enable users to share content (e.g., texts, images, videos) and interact with others (e.g., likes, comments). More importantly, social media platforms provide personalized content (e.g., personalized ads, updates of friends, recent news) that is more relevant to users' preferences to keep users engaging in the platforms (Satterfield 2020). This is possible for social media platforms because all available data of users' activities (e.g., posts, usage hours, locations, and dialogues) are collected by the platforms (Orlowski 2020), which is usually referred to as *data harvesting* (Dutton 2022). This paper is motivated by the problem that data harvesting by social media platforms may be harmful to social media users but beneficial for the platforms.

Social media platforms generate a significant portion of revenue from advertising, known as an ad-supported revenue model (Sherman 2018). In 2021, Facebook and Twitter reported that 99% and 89% of their revenue came from advertising, respectively (Meta 2022; Twitter 2022). This business model allows users to join the platforms for free but charges advertisers to access targeted users who have the highest preference for a given advertiser's brand. This targeted advertising has been highly effective for advertisers because social media platforms can precisely predict users' preferences and actions (e.g., clicks, likes, shares), using the data collected from their users. In other words, the more data a social media platform collects from its users, the more the platform knows about its users.

While social media platforms could better serve users by leveraging users' data, however, they have been associated with the negative mental health outcomes of users (Allcott et al. 2022; Lambert et al. 2022). For example, social media could be related to the gigantic rises in depression, anxiety, and the suicide rate for American teenagers that have started around 2012 in which social media platforms became available on mobile devices (Allen 2020; Haidt and Allen 2020). In a recent discussion on the negative consequences of social media platforms, practitioners in the social media industry suggest that data harvesting by social media platforms could be the starting point of the negative impacts of social media use on the users. For instance, industry expert features in the documentary mentioned earlier describes that collecting massive amount of users' data has enabled social media platforms to exploit a vulnerability in human psychology because the platforms' algorithms can even predict when a user is bored, lonely, or depressed (Orlowski 2020).

The government has recently started to investigate data harvesting practices by social media platforms. For example, Federal Trade Commission (FTC) asked nine popular social media platforms to provide information about how they collect and use users' data and how their data harvesting practices influence children and teens (Federal Trade Commission 2020). While FTC's investigation has been still in-progress, there have been increasing discussions about regulating social media platforms. In a recent testimony of the CEOs of top social media companies in March 2021, policymakers show consensuses that the business model of social media platforms has problems and that the industry itself is not likely to be trusted to solve the issue (Conger et al. 2021). Therefore, in this paper, we consider that the government requires social media platforms to make effort to address issues arising from collecting massive amount of data from social media users.

Research Questions and Contributions

As both the government and practitioners have been highly concerned about social media platforms that collect a massive amount of data from users (Federal Trade Commission 2020; Orlowski 2020), there have been ongoing discussions on regulating social media platforms. For example, policymakers have been recently considering imposing a regulation on social media platforms generating negative impacts on users (Deighton 2022). Even though the question of how imposing a regulation on data harvesting would impact platforms and users is becoming more important, to the best of our knowledge, there has been a limited understanding to answer this question. Thus, our first research question is as follows: *Does a regulation on data harvesting really reduce data harvesting levels and profits but increase demands of social media platforms? Furthermore, does a regulation on data harvesting really increase consumer welfare?* We find that even after the government imposes the regulation on data harvesting, social media platforms can sometimes set higher data harvesting levels, have smaller demands, and generate greater profits. Moreover, we find that imposing the regulation on data harvesting always increases consumer welfare.

Even if the government imposes a regulation on data harvesting, users' concerns about data harvesting may further increase. For example, after the General Data Protection Regulation (GDPR) has been imposed in the European Union since 2016, several platforms have failed to comply the regulation and received hefty fines (e.g., $60 million for Facebook in 2022 and $547k for Twitter in 2020) (Collins 2022; Lomas 2020). With such violations, users have been more concerned about the potential privacy risks in data harvesting practices (Goswami 2020). Moreover, in the rising concerns about children users, policymakers have been interested in whether social media platforms could limit data harvesting practices under the regulation (Deighton 2022). Therefore, it is important to understand how the increase in the net negative effect of data harvesting on users impacts social media platforms under the regulation. Thus, our second research question is as follows: *Does the increase in users' concerns about data harvesting really reduce data harvesting levels, demands, and profits of social media platforms under the regulation?* One may expect that platforms would be worse off when users are more concerned about sharing their personal data with platforms. However, we find that when users are more concerned about data harvesting, a social media platform with a weaker network effect can sometimes set a higher data harvesting level and generate greater profits under the regulation.

This paper contributes to the information systems (IS) literature by providing insights into two aspects. First, we provide valuable insight into how implementing a new regulation on data harvesting would affect social media platforms. Our results show that, while the objective of the new regulation is to discourage data harvesting practices, sometimes platforms increase their data harvesting levels and generate greater profits under the regulation on data harvesting. Second, this paper provides insights on how the increase in the users' concerns about data harvesting impacts social medial platforms under the regulation. While the IS literature has paid little attention to the platform decision that hurts users, a stream of literature has focused on the harm from advertisements (Aseri et al. 2020; Hann et al. 2008). Specifically, Aseri et al. (2020) find that, when ads generate a larger disutility to users, the intensity of advertisements decreases. However, we find that social media platforms are not necessarily to decrease data harvesting levels, which are highly related to target advertising. Therefore, by considering data harvesting practices by social media platforms, our results expand the body of knowledge on the ad-supported business model fueled by data harvesting. This paper provides important implications for both social media platforms and policymakers on how social media platforms could react to a new regulation when the core business decision is regulated by the government.

The rest of the paper is organized as follows. We briefly review the related literature in the following subsection. Next, we introduce our analytical model and present our main results. We then conclude and highlight our contributions and implications.

Literature Review

This paper mainly contributes to two streams of literature: (i) social media and its negative impacts on users, (ii) the impact of the government's policy on the market. In this section, we briefly review these streams of literature and highlight our contributions.

Negative Impacts of Social Media on Users

The first stream of literature studies the negative impacts of social media on users. In IS literature, previous studies have focused on the role of envy (Krasnova et al. 2015) and cyberbullying in social media (Wong et al. 2021). Other studies investigate that advertisements cause harm to users and, therefore, users make effort to avoid ads, known as *ad-avoidance* behaviors (Aseri et al. 2020; Hann et al. 2008). In the economics literature, (Allcott et al. 2020) conduct a large-scale field experiment on 1,637 Facebook users and find that deactivating Facebook accounts significantly increased subjective well-being. In the psychology literature, a stream of research reports that the use of social media can be harmful to mental health outcomes (Haidt and Allen 2020).

Much of the literature under this topic has focused on various unfavorable consequences of social media use. However, there is a lack of research on how the platform's business decision contributes to the negative impacts of social media use. In this study, we investigate the ad-supported revenue model of social media platforms, which is fueled by the amount of data gathered from social media users. This data harvesting is related to both the negative impacts of social media use (e.g., depression, comparisons) and the increase in the advertising revenue. Similar to our study, Aseri et al. (2020) consider ad-intensity that increases the advertising revenue of a website but decreases the net utility of a potential user. In their work, the website is not penalized by increasing its ad-intensity, unlike our model in which platforms face a higher cost as they collect more data from users. Therefore, our model is fundamentally different from that in previous studies and better captures the phenomenon of unfavorable consequences of social media use on users.

The Impact of the Government's Policy on the Market

The second stream of literature related to our paper is the literature on the impact of government's policy on the market. A stream of literature on piracy has studied how the government's policy to reduce illegal content affects users' and publishers' strategies in the context of platform markets. An earlier work by (Chen and Png 2003) study the government's penalty rate for piracy that impacts only the supply side. (Jaisingh 2009) studies a policy that increases the perceived cost to the users of pirated software and finds that stricter policy may reduce innovation. In the context of social media platforms, (Jain et al. 2020) investigate how the platform's monitoring efforts to block the illegal content impact the profit of the platform. In their work, the platform's effort increases the user's cost of accessing the illegal content, thereby decreasing the utility of the users, unlike our model that the platform's effort increases the utility of the users. Another stream of literature on privacy concerns has documented how the government's privacy regulation impacts the welfare of firms in the market (Goldfarb and Tucker 2011).

Different from these studies focusing on the impact of government's regulation as a fixed cost or a condition limiting the access to users' data, our study examines that the government requires social media platforms to put effort to address issues arising from the data harvesting practices. As industry practitioners call for regulations on social media platforms and policymakers seek for new regulations, our model provides useful insights for policymakers on regulating the data harvesting decision by social media platforms.

Model

Social media market is with intense competition among social media platforms (Global Data 2022). In this market, previous IS studies have considered Facebook and Twitter as major social media platforms (Wang et al. 2021). Therefore, we consider a social media market with two platforms competing for a heterogeneous set of users. In the following subsections, we present our model in detail listing the different characteristics of social media users, platforms, and the game structure. After we introduce our model, we discuss two scenarios we consider that are prevalent in practice. Main notations are summarized in Table 1.

Social Media Users

A survey conducted by Pew Research Center shows that 72 percentage of the U.S. adults use any social media platform as of February 2021 (Pew Research Center 2021). Moreover, another recent report suggests that this share will continue to grow in the next few years (Statista 2021). Therefore, we consider that the market

Social media platform's decision variable	
h_i	Data harvesting level of platform $i, i \in \{1,2\}$
Other variables	
$\mathbb{E}(n_i)$	Expected number of social media users joining platform $i, i \in \{1,2\}$
u_i	Net utility for a social media user from joining platform $i, i \in \{1,2\}$
π_i	Profit for platform $i, i \in \{1,2\}$
Parameters	
θ	User's preference for a social media platform, and $\theta \sim$ Uniform $[0,1]$
p	Effort cost for a unit level of data harvesting
v	Value a user receives from basic features of a social media platform
α_i	Coefficient for platform i's network effect
β	Net negative effect of harvesting one unit of data on a user's utility without the government's regulation on data harvesting
γ	Net negative effect of harvesting one unit of data on a user's utility under the presence of the government's regulation on data harvesting
n	Total number of potential social media users
κ	Social media platform misfit cost

Table 1. Main Notation

is fully covered; that is, all users will join a platform.[1] These users derive different types of utility from joining a platform. The utility of a user joining a social media platform consists of the value a user derives from social communication features, the value derives from the network effect, the negative impact of data harvesting and the misfit in social media platforms.

The first component of social media user utility is the value the user derives from social communication features. Social media platforms provide features for promoting communication among the users and for sharing content (e.g., wall posts, likes, and comments) (Cao et al. 2018). These features are generally common across the platforms and, therefore, the value that a user earns from these features is likely similar across the platforms (VanDyke 2022). Therefore, we denote v as the value a user gains by using communicating features of a social media platform.

The second component of social media user utility is the value that the user gains by communicating with other users. The more the users join the platform, the more the content is generated and shared. Thus, the value that a user earns from the platform increases with the number of users, a phenomenon known as the network effect (Katz and Shapiro 1985). Following previous literature (Demirezen et al. 2016), we also consider that the user utility increases linearly in the size of the network. We denote $\mathbb{E}(n_i)$ as the expected network size in social media platform $i, i \in \{1,2\}$.

A survey conducted by Pew Research Center in 2021 shows that 71% of Facebook users visit the platform at least once a day, but 46% of Twitter users do so (Auxier and Anderson 2021). To capture such difference, we consider that the network effect can differ across platforms. We denote $\alpha_i, i \in \{1,2\}$ as the coefficient that captures platform i's network effect. If a user expects $\mathbb{E}(n_i)$ users to join platform i, this user gains value $\alpha_i\mathbb{E}(n_i)$ from joining platform i.

The third component is the net negative effect effect of data harvesting on social media users. In the documentary, "The Social Dilemma", discussed earlier, industry experts point out that data harvesting enables social media platforms to precisely predict users' preferences and provide personalized content (Orlowski 2020). While they admit that offering personalized content could be helpful to users, they emphasize that

[1] We also consider a case in which social media users may join both platforms at the same time, calling it as "Multihoming" scenario. In addition, we also consider a case in which some users do not join any platform (i.e., the market is partially covered). We will present these results during the conference presentation.

social media platforms' ability to accurately predict users' actions is causing more harms (e.g., social media addiction, the wide spread of fake news, and mental health issues for teenagers). Other discussions also suggest that the negative impact of data harvesting on users outweigh the positive impact. For example, policymakers have been highly concerned about the business model of social media platforms that hurts users (Deighton 2022; Wells et al. 2021) and social media users have been greatly concerned about privacy when they share personal information with the platforms (Auxier et al. 2019). In order to capture these concerns, we consider that the net impact of data harvesting creates disutility on social media users. We denote the net negative effect of harvesting one unit of data on a users' utility when there is no regulation by the government by β and $\beta > 0$. Therefore, we denote $-\beta h_i$ as the net negative impact of data harvesting by platform $i, i \in \{1, 2\}$ on a user's utility, where h_i is the amount of a user's data collected by platform i. Note that social media platforms have been collecting users' data regardless of users' willingness to share their data with these platforms (Orlowski 2020). Thus, we consider the additivity of the network effect and the disutility from data harvesting in our model.

The last component is the disutility from the misfit between the social media platform and the user's taste. A recent report shows that Twitter is the most popular among the age group of 18- to 29-years-old in urban areas, but Facebook is popular for all adults regardless of locations (Auxier and Anderson 2021). Therefore, we consider that social media users are heterogeneous in terms of their preferences for social media platforms. We denote θ as the user's preference for a social media platform, which follows a uniform distribution on $[0, 1]$. Then, the degree of misfit between the social media platform and the taste of a user is measured by the distance between the platform's and a user's location on the line θ. We denote $\kappa\theta$ and $\kappa(1 - \theta)$ as the disutility caused due to the misfit between user type θ and platform 1 and 2, respectively, where κ is the misfit cost which is greater than zero.

We denote the social media user's utility derived from joining to social media platform i as $u_i, i \in \{1, 2\}$. Based on the above discussion, when no regulation is imposed on data harvesting by government (i.e., the *no regulation* scenario), the utility for a social media user located at θ from each social media platform is formulated as follows.

$$u_1(\theta) = v + \alpha_1 \mathbb{E}(n_1) - \beta h_1 - \kappa\theta \tag{1}$$

$$u_2(\theta) = v + \alpha_2 \mathbb{E}(n_2) - \beta h_2 - \kappa\,(1 - \theta) \tag{2}$$

When the government implements a regulation on data harvesting, such regulation is likely to affect users' privacy concerns. For instance, a recent survey shows that, after the GDPR in the European Union is implemented, the majority of users (65%) feel they have more control over the information they provide online (European Commission 2019). Another survey suggests that users have more trust and satisfaction toward the GDPR-compliant firms because the GDPR increases cybersecurity and users' awareness about what data is being collected and how it is used (Jiang et al. 2019). To capture this reality, we consider that the disutility caused by data harvesting on a user becomes smaller under the presence of the government's regulation. We denote the net negative effect of harvesting one unit of data on a users' utility when the government regulates data harvesting practices (i.e., the *with regulation* scenario) by γ and $\beta \geq \gamma > 0$. Therefore, the utility for a social media user located at θ from each social media platform in the *with regulation* scenario can be formulated as follows.

$$u_1(\theta) = v + \alpha_1 \mathbb{E}(n_1) - \gamma h_1 - \kappa\theta \tag{3}$$

$$u_2(\theta) = v + \alpha_2 \mathbb{E}(n_2) - \gamma h_2 - \kappa\,(1 - \theta) \tag{4}$$

In both scenarios, we define platform 1 as a platform whose network effect parameter is greater without loss of generality. Therefore, we consider that $\alpha_1 = 1$ and $\alpha_2 = \alpha$, where $0 < \alpha < 1$.

Social Media Platforms

According to reports released by Facebook and Twitter, 99% and 89%, respectively, of their revenues in 2021 were driven from advertising (Meta 2022; Twitter 2022). Therefore, we consider an ad-supported revenue model of social media platforms in this study. The revenue from advertising depends on the number of users because an increase in the number of users attracts advertisers into the platform and boosts the consumption of ads by users (Zhang and Sarvary 2015). The revenue from advertising also depends on how much data a social media platform collects from its users because the platform can precisely target user segments with predicted preferences on products and services (Orlowski 2020; Satterfield 2020). Based on the above discussion, we consider that when social media platform i expects $\mathbb{E}(n_i)$ users and collects h_i amount of users' data, the platform generates revenue of $\mathbb{E}(n_i)h_i$.

When Facebook failed to comply with privacy rules in the U.S. in 2019, FTC imposed three types of requirements on Facebook for deceiving its users about privacy (Federal Trade Commission 2019). First, FTC required Facebook to put more IT effort to ensure user privacy by establishing and maintaining an independent privacy committee and data security programs. Second, FTC also required Facebook to make more effort to inform users about potential privacy risks by providing clear notice and obtaining user consent. Third, FTC required Facebook to put effort to report to the government about its compliance by submitting quarterly compliance certifications and delivering data breach reports to the Commission.

In order to capture these approaches, we consider that the government regulates social media platforms by requiring them to make effort to address issues arising from data harvesting practices. First, social media platforms should make extra IT effort to safeguard the data collected from the users by preventing data breaches and maintaining comprehensive data security programs. Second, the platforms should put effort to inform the users about data harvesting practices by describing what data is collected, how data is used, and what controls the users have. Informing users about data harvesting practices has also been highlighted in the Dashboard act in the U.S. and the GDPR in the European Union (Gordon 2019). Third, the platforms should make effort to report to the government about their compliance with the regulation and any incident against safeguarding users' data.

Therefore, in our model, we denote p as the effort cost for a unit level of data harvesting. When a social media platform collects massive amount of data from its users, the data could be beyond platform's control and the potential risks arising from securing massive amount of data increase at a faster rate. Therefore, the amount of efforts imposed by the government increases non-linearly with the data harvesting level. In order to capture this reality in our model, when social media platform i collects h_i units of a user's data, the platform must make the effort by $ph_i^2, i \in \{1, 2\}$. We also solve a linear model where the amount of efforts imposed by the government increases linearly with the data harvesting level as an extension and our results are qualitatively the same.

Based on the above discussions, social media platform i's net profit in the *no regulation* scenario is

$$\pi_i = \mathbb{E}(n_i)h_i, \tag{5}$$

and social media platform i's net profit in the *with regulation* scenario is as follows.

$$\pi_i = \mathbb{E}(n_i)h_i - ph_i^2 \tag{6}$$

Timing of the Model

For social media platforms, the decision of how much data to collect from their users is a part of their long-term privacy policy. Such long-term policy helps users make decisions such as which platform to join. In order to capture this reality, we consider a game with two stages as follows. In stage 1, social media platforms, 1 and 2, first decide how much data to be harvested from users, given the effort imposed by the government to address issues arising from data harvesting practices. In stage 2, each social media user decides which social media platform to join. Given the two-stage game setting, we solve the game in the backward order.

Equilibrium Analysis

In this section, we study our two-stage model to understand social media platforms' equilibrium data harvesting strategies. Our goal is to investigate how implementing the regulation on data harvesting affects social media platforms and user surplus. Therefore, we consider the following two scenarios: (i) the *no regulation* scenario in which there is no regulation on data harvesting; and (ii) the *with regulation* scenario in which the government implements the regulation on data harvesting by requiring social media platforms to make effort to address issues arising from the data harvesting practices.

No Regulation Scenario

As a benchmark scenario, we first analyze the scenario where the government does not implement any regulation on the data harvesting level. In stage 2, each user joins a social media platform, given the data harvesting levels set by the platforms in stage 1. For a potential user, the decisions of other users on joining a platform affect her net utility because her utility depends on the expected number of other users joining the platform. Therefore, we provide details on calculating the expected number of users joining platform i as follows.

The utilities that a social media user with type θ for a social media platform obtains from joining platform 1 or 2 are shown as follows, respectively: (i) join platform 1: $v + \mathbb{E}(n_1) - \beta h_1 - \kappa\theta$, and (ii) join platform 2: $v+\alpha\mathbb{E}(n_2)-\beta h_2-\kappa(1-\theta)$. If a user with type θ decides to join platform 1, then she should obtain the highest non-negative utility from joining platform 1, that is, $v+\mathbb{E}(n_1)-\beta h_1-\kappa\theta \geq v+\alpha\mathbb{E}(n_2)-\beta h_2-\kappa(1-\theta)$ and $v+\mathbb{E}(n_1)-\beta h_1-\kappa\theta \geq 0$. Let us denote $\tilde{\theta}$ as a user who is indifferent between joining platform 1 and 2. By solving $v+\mathbb{E}(n_1)-\beta h_1-\kappa\theta = v+\alpha\mathbb{E}(n_2)-\beta h_2-\kappa(1-\theta)$, we get $\tilde{\theta} = (\mathbb{E}(n_1)-\alpha\mathbb{E}(n_2)-\beta h_1+\beta h_2+\kappa)/2\kappa$. Let us denote $\hat{\theta_1}$ as a user who obtains zero utility from joining platform 1, which is located at $\hat{\theta_1} = (v + \mathbb{E}(n_1) - \beta h_1)/\kappa$. Solving the two inequalities, we obtain that users with type $0 \leq \theta \leq \min\{\tilde{\theta}, \hat{\theta_1}\}$ prefer joining platform 1 to joining platform 2. If a user with type θ decides to join platform 2, then there should be $v+\alpha\mathbb{E}(n_2)-\beta h_2-\kappa(1-\theta) \geq v+\mathbb{E}(n_1)-\beta h_1-\kappa\theta$ and $v+\alpha\mathbb{E}(n_2)-\beta h_2-\kappa(1-\theta) \geq 0$. Let us denote $\hat{\theta_2}$ as a user who obtains zero utility from joining platform 2, which is located at $\hat{\theta_2} = (-v-\alpha\mathbb{E}(n_2)+\beta h_2+\kappa)/\kappa$. Solving the two inequalities, we obtain that users with type $\max\{\tilde{\theta}, \hat{\theta_2}\} \leq \theta \leq 1$ prefer joining platform 2 to joining platform 1. Given $\tilde{\theta} \leq \hat{\theta_1}$, there is $v \geq (-\mathbb{E}(n_1)-\alpha\mathbb{E}(n_2)+(h_1+h_2)\beta+\kappa)/2$. Given $\tilde{\theta} \geq \hat{\theta_2}$, there is also $v \geq (-\mathbb{E}(n_1)-\alpha\mathbb{E}(n_2)+(h_1+h_2)\beta+\kappa)/2$.

As discussed in the Model Section, we consider that all users in the market join a platform. When $v \geq (-\mathbb{E}(n_1)-\alpha\mathbb{E}(n_2)+(h_1+h_2)\beta+\kappa)/2$ holds, users with $0 \leq \theta \leq \tilde{\theta}$ prefer to join platform 1 and users with $\tilde{\theta} < \theta \leq 1$ join platform 2. Thus, all users in the market join a platform, resulting in the full market coverage.

The demand for each social media platform is

$$\mathbb{E}(n_1) = \tilde{\theta}n = \frac{n(n\alpha+(h_1-h_2)\beta-\kappa)}{n(1+\alpha)-2\kappa}, \tag{7}$$

$$\mathbb{E}(n_2) = (1-\tilde{\theta})n = \frac{n(n-(h_1-h_2)\beta-\kappa)}{n(1+\alpha)-2\kappa}. \tag{8}$$

As discussed in the Motivation Section, both industry experts and policymakers express concerns about data harvesting which negatively affects social media users (Conger et al. 2021; Orlowski 2020). Therefore, to ensure that the utility of a user and the demand of each platform decrease in the data harvesting level, we consider that the misfit cost factor, κ, is large enough (i.e., $\kappa \geq n(1+\alpha)/2$), satisfying $\frac{\partial\mathbb{E}(n_i)}{\partial h_i}<0$ and $\frac{\partial u_i(\theta)}{\partial h_i}<0$ for $i \in \{1,2\}$.

In stage 1, the social media platforms decide the data harvesting level to maximize their profits. The profit maximization problem of social media platform $i, i \in \{1,2\}$ is $\max\{\pi_i\} = \max\{\mathbb{E}(n_i)h_i\}$ for $i \in \{1,2\}$. Social media platform *i*'s objective function is $\pi_1 = (h_1 n(n\alpha+(h_1-h_2)\beta-\kappa))/(n(1+\alpha)-2\kappa)$ and $\pi_2 = (h_2 n(n-(h_1-h_2)\beta-\kappa))/(n(1+\alpha)-2\kappa)$. The following lemma shows the results under the *no regulation* scenario, with the superscript NR denoting the equilibrium outcomes. Proofs of Lemmas and Propositions are omitted for the space limit. Proofs are available upon request.

Lemma 1. *Under the no regulation scenario,*

1. *The equilibrium data harvesting level of social media platform* $i, i \in \{1, 2\}$*, is*

$$h_1^{NR} = \frac{-n(1+2\alpha)+3\kappa}{3\beta}, \tag{9}$$

$$h_2^{NR} = \frac{-n(2+\alpha)+3\kappa}{3\beta}. \tag{10}$$

2. *The equilibrium demand of social media platform* $i, i \in \{1, 2\}$*, is*

$$\mathbb{E}(n_1^{NR}) = \frac{n(n(1+2\alpha)-3\kappa)}{3(n(1+\alpha)-2\kappa)}, \tag{11}$$

$$\mathbb{E}(n_2^{NR}) = \frac{n(n(2+\alpha)-3\kappa)}{3(n(1+\alpha)-2\kappa)}. \tag{12}$$

3. *The equilibrium profit of social media platform* $i, i \in \{1, 2\}$*, is*

$$\pi_1^{NR} = \frac{-n(n(1+2\alpha)-3\kappa)^2}{9\beta(n(1+\alpha)-2\kappa)}, \tag{13}$$

$$\pi_2^{NR} = \frac{-n(n(2+\alpha)-3\kappa)^2}{9\beta(n(1+\alpha)-2\kappa)}. \tag{14}$$

A number of observations from the equilibrium are worth highlighting. First, note that $h_1^{NR} \geq h_2^{NR}, \mathbb{E}(n_1^{NR}) \geq \mathbb{E}(n_2^{NR})$, and $\pi_1^{NR} \geq \pi_2^{NR}$. This result is expected because platform 1 has an advantage of the network effect over platform 2 (i.e., $\alpha < 1$). Second, equilibrium demands are unrelated to the net negative effect of data harvesting on a user's utility, β, even though social media user's utilities in Eq(1) and Eq(2) depend on it. This is because, even though social media users are generally concerned with disclosing personal data in social media platforms, they continue to use these platforms and provide their personal information, which is known as the *privacy paradox* (Acquisti and Gross 2006).

With Regulation Scenario

In this subsection, we analyze the scenario where the government requires social media platforms to make effort to address issues arising from the data harvesting practices. In stage 2, each user joins a social media platform, given the data harvesting levels set by the platforms in stage 1. The utilities that a social media user with type θ for a social media platform obtains from joining platform 1 or 2 are shown as follows, respectively: (i) join platform 1: $v + \mathbb{E}(n_1) - \gamma h_1 - \kappa\theta$ and (ii) join platform 2: $v + \alpha\mathbb{E}(n_2) - \gamma h_2 - \kappa(1-\theta)$.

Following the same steps described in the analysis of the *no regulation* scenario, we locate a user who is indifferent between joining platform 1 and 2, $\tilde{\theta}$. By solving $v+\mathbb{E}(n_1)-\gamma h_1-\kappa\theta = v+\alpha\mathbb{E}(n_2)-\gamma h_2-\kappa(1-\theta)$, we get $\tilde{\theta} = (\mathbb{E}(n_1) - \alpha\mathbb{E}(n_2) - (h_1 - h_2)\gamma + \kappa)/2\kappa$. Let us denote $\hat{\theta_1}$ as a user who obtains zero utility from joining platform 1, which is located at $\hat{\theta_1} = (v + \mathbb{E}(n_1) - \gamma h_1)/\kappa$. Similarly, let us denote $\hat{\theta_2}$ as a user who obtains zero utility from joining platform 2, which is located at $\hat{\theta_2} = (-v - \alpha\mathbb{E}(n_2) + \gamma h_2 + \kappa)/\kappa$. As discussed in the Model Section, we consider that every user in the market joins a platform. When $v \geq (-\mathbb{E}(n_1) - \alpha\mathbb{E}(n_2) + (h_1 + h_2)\gamma + \kappa)/2$ holds, max $\{\tilde{\theta}, \hat{\theta_2}\}$ = min $\{\tilde{\theta}, \hat{\theta_2}\} = \tilde{\theta}$. Therefore, users with $0 \leq \theta \leq \tilde{\theta}$ prefer to join platform 1 and users with $\tilde{\theta} < \theta \leq 1$ join platform 2. Thus, all users in the market join a platform, resulting in the full market coverage. We get the demand for each social media platform as follows.

$$\mathbb{E}(n_1) = \tilde{\theta}n = \frac{n(n\alpha + (h_1 - h_2)\gamma - \kappa)}{n(1+\alpha)-2\kappa}, \tag{15}$$

$$\mathbb{E}(n_2) = (1-\tilde{\theta})n = \frac{n(n - (h_1 - h_2)\gamma - \kappa)}{n(1+\alpha)-2\kappa}. \tag{16}$$

To ensure that the utility of a user and the demand of each platform decrease in the data harvesting level, we consider that the misfit cost factor, κ, is large enough (i.e., $\kappa \geq n(1+\alpha)/2$), satisfying $\frac{\partial \mathbb{E}(n_i)}{\partial h_i} < 0$ and $\frac{\partial u_i(\theta)}{\partial h_i} < 0$ for $i \in \{1,2\}$.

In stage 1, the social media platforms observe the government requiring the effort to address issues from data harvesting and decide how much data to be harvested from their users to maximize the profits. Based on the discussion in the Model Section, the profit maximization problem of social media platform $i, i \in \{1,2\}$ is $\max\{\pi_i\} = \max\{\mathbb{E}(n_i)h_i - ph_i^2\}$ for $i \in \{1,2\}$. Social media platform i's objective function is $\pi_1 = h_1(-h_1 p + (n(n\alpha + (h_1 - h_2)\gamma - \kappa))/(n(1+\alpha) - 2\kappa))$, and $\pi_2 = h_2(-h_2 p + (n(n - (h_1 - h_2)\gamma - \kappa))/(n(1+\alpha) - 2\kappa))$. The following lemma shows the results under the *with regulation* scenario, with the superscript R denoting the equilibrium outcomes.

Lemma 2. *Under the with regulation scenario,*

1. *The equilibrium data harvesting level of social media platform $i, i \in \{1,2\}$, is*

$$h_1^R = \frac{n}{4}\left(\frac{2n(-1+\alpha)}{2np(1+\alpha) - 3n\gamma - 4p\kappa} + \frac{1}{p}\left(1 + \frac{n\gamma}{2np(1+\alpha) - n\gamma - 4p\kappa}\right)\right), \tag{17}$$

$$h_2^R = \frac{n}{4}\left(-\frac{2n(-1+\alpha)}{2np(1+\alpha) - 3n\gamma - 4p\kappa} + \frac{1}{p}\left(1 + \frac{n\gamma}{2np(1+\alpha) - n\gamma - 4p\kappa}\right)\right). \tag{18}$$

2. *The equilibrium demand of social media platform $i, i \in \{1,2\}$, is*

$$\mathbb{E}(n_1^R) = \frac{n}{6}\left(3 + n(-1+\alpha)\left(\frac{1}{n(1+\alpha) - 2\kappa} + \frac{4p}{2np(1+\alpha) - 3n\gamma - 4p\kappa}\right)\right), \tag{19}$$

$$\mathbb{E}(n_2^R) = \frac{n}{6}\left(3 + n(-1+\alpha)\left(\frac{1}{-n(1+\alpha) + 2\kappa} - \frac{4p}{2np(1+\alpha) - 3n\gamma - 4p\kappa}\right)\right). \tag{20}$$

3. *The equilibrium profit of social media platform $i, i \in \{1,2\}$, is*

$$\pi_1^R = \frac{n^2(n(p(1+\alpha) - \gamma) - 2p\kappa)(n^2(-2p\alpha(1+\alpha) + \gamma + 2\alpha\gamma) + n(p(2+6\alpha) - 3\gamma)\kappa - 4p\kappa^2)^2}{(n(1+\alpha) - 2\kappa)(-2np(1+\alpha) + 3n\gamma + 4p\kappa)^2(n(-2p(1+\alpha) + \gamma) + 4p\kappa)^2}, \tag{21}$$

$$\pi_2^R = \frac{n^2(n(p(1+\alpha) - \gamma) - 2p\kappa)(n^2(2p(1+\alpha) - (2+\alpha)\gamma) + n(-2p(3+\alpha) + 3\gamma)\kappa + 4p\kappa^2)^2}{(n(1+\alpha) - 2\kappa)(-2np(1+\alpha) + 3n\gamma + 4p\kappa)^2(n(-2p(1+\alpha) + \gamma) + 4p\kappa)^2} \tag{22}$$

A number of observations from the equilibrium are worth highlighting. First, similar to the *no regulation* scenario, $h_1^R \geq h_2^R$, $\mathbb{E}(n_1^R) \geq \mathbb{E}(n_2^R)$, and $\pi_1^R \geq \pi_2^R$. This is expected because platform 1 has an advantage of the network effect over platform 2 (i.e., $\alpha < 1$). Second, different from the *no regulation* scenario, equilibrium demands now depend on the net negative effect of data harvesting, γ. It is because the effort cost of data harvesting imposed by the regulation becomes salient for social media users' decisions to join a platform under the regulation on data harvesting. For example, when the U.S. government started to investigate Facebook for its data breach incident in 2018, users began #DeleteFacebook campaign in response to the FTC's interrogation of Facebook (Pachelli 2018). Thus, equilibrium demands depend on the net negative effect of data harvesting, which is different from the *no regulation* scenario.

Results

Now we have established the equilibrium outcomes under both scenarios. Thus, we are ready to investigate how implementing the regulation on data harvesting affects social media platforms and consumer welfare. In this section, we present our main results by addressing our key research questions in detail.

Impacts of the Government's Regulation

As discussed in the Motivation Section, industry experts have been calling for some form of regulation on social media platforms (Orlowski 2020). In response to the call, policymakers have recently started considering new regulations on the business model of social media platforms (Conger et al. 2021). Although policymakers have especially focused on data harvesting practices (Federal Trade Commission 2020), to the best of our knowledge, there has been a limited understanding of how the government's regulation on data harvesting would affect social media platforms. Therefore, we consider the government's regulation that has two impacts, as discussed in the Model Section: (i) the regulation requires social media platforms to make efforts to address issues arising from data harvesting practices and (ii) the efforts required by the regulation positively impact the users by reducing the net negative effect of data harvesting on a user's utility from β to γ, as shown in Equations (1)-(4). In the following proposition, we address the question of how implementing the government's regulation on data harvesting impacts the data harvesting levels, demands, and profits of social media platforms.

Proposition 1. *When the government imposes the regulation on data harvesting,*

(a) *platform i can actually sometimes set a higher data harvesting level (i.e., $h_i^R \geq h_i^{NR}$). Interestingly, this happens when the positive impact of the data harvesting regulation on users' utility is higher than a given threshold (i.e., $\delta \geq \overline{\xi_i}$, where $\delta = \beta - \gamma$);*

(b) *the demand for a platform with a weaker network effect (i.e., platform 2), in fact, always becomes smaller (i.e., $\mathbb{E}(n_2^R) \leq \mathbb{E}(n_2^{NR})$), in spite of the fact that the demand for the other platform always becomes greater (i.e., $\mathbb{E}(n_1^R) \geq \mathbb{E}(n_1^{NR})$);*

(c) *the profit for platform i can actually sometimes become greater (i.e., $\pi_i^R \geq \pi_i^{NR}$). Interestingly, this happens when the positive impact of the data harvesting regulation on users' utility is higher than a given threshold (i.e., $\delta \geq \hat{\xi}_i$).*

Proofs Propositions are omitted for the space limit. Proofs are available upon request.

Under the regulation that restricts collecting data from users (e.g., GDPR), the platforms should identify the minimum amount of personal information to be collected from users (Burgess 2020). Therefore, one may expect that the platforms would set lower data harvesting levels under the regulation that restricts data harvesting practices. Similarly, recent surveys show that such a regulation decreases users' privacy concerns and increases trust and satisfaction in using platforms that collect their personal information (European Commission 2019; Jiang et al. 2019). Therefore, one may expect that the demand for the platforms would increase under the regulation. Lastly, it has been suggested that online platforms have been making less profits because regulations such as GDPR restrict collecting users' data that directly related to the revenue of these platforms (Chen 2019). Thus, one may expect that the profits would decrease when the government imposes the regulation. However, Proposition 1 demonstrates that none of these always holds. This result is explained as follows.

The intuition of Proposition 1(a) is as follows. Whether platforms set higher data harvesting levels under the *with regulation* scenario depends on the *utility effect* and the *cost effect* defined below. On the one hand, the regulation reduces the net negative effect of data harvesting on a user's utility from β to γ, as discussed in the Model Section. With this reduction, users take less amount of disutility from data harvesting under the regulation (i.e., $\gamma h_i \leq \beta h_i$), which makes it easier for the platforms to collect more data from the users. Therefore, the regulation softens the intensity of the competition between the two platforms, driving the data harvesting levels up under the regulation. We call this the *utility effect*. While the utility effect is determined by the sum of two effects: (i) the reduction in the disutility from data harvesting and (ii) the changes in demands, we show that the sum of these two effects can be captured by δ, where $\delta = \beta - \gamma$. Thus, the utility effect, δ, indicates the amount of reduction in the net negative effect of data harvesting on a user's utility. On the other hand, recall that the regulation introduces the effort cost, p, for a unit level of data harvesting, and the total effort cost increases non-linearly in the data harvesting level (i.e., ph_i^2), as discussed in the Model Section. The effort cost imposes a constraint on the platforms in setting a higher data harvesting level. Therefore, the regulation increases the intensity of the competition between two platforms, lowering the data harvesting levels under the regulation. We call this the *cost effect* that is impacted by p.

The left-hand side of the condition in Proposition 1(a) represents the utility effect and the right-hand side that is impacted by p is the cost effect. When the utility effect dominates the cost effect, the regulation softens the competition between two platforms, and thus the platforms set higher data harvesting levels under the *with regulation* scenario than under the *no regulation* scenario.

The result in Proposition 1(b) occurs because platform 2 has a disadvantage of the network effect over platform 1 (i.e., $\alpha < 1$). Although the regulation reduces the net negative effect of data harvesting on the users for both platforms (i.e., from β to γ), the platform with a stronger network effect leverages its advantage in the network effect to obtain even more market share under the regulation. This has been consistently shown that platform 1 with a stronger network effect collects more data from users, takes more market share, and generate higher profits than platform 2 under both scenarios, as shown in Lemma 1 and Lemma 2.

The result in Proposition 1(c) occurs when the utility effect dominates the cost effect. The left-hand side of the condition in Proposition 1(c) represents the utility effect and the right-hand side is the cost effect. Again, under the perfect regulation case (i.e., $\gamma = 0$), the condition reduces to $\beta \geq p\hat{\omega}_i$. We can see the right-hand side represents the cost effect that depends on p. Thus, when the utility effect dominates the cost effect, the regulation decreases the competition between two platforms, driving the data harvesting levels up and resulting in higher profits. Note that the cost effect for higher profits is greater than that for higher data harvesting levels (i.e., $\hat{\omega}_i > \overline{\omega_i}$). Therefore, when the utility effects dominates the higher cost effect, platforms not only collect more data from the users but also become better off.

Next, we discuss how imposing the regulation on data harvesting impacts users. As the usage of the social media could be highly related to the negative mental health outcomes of teenagers (Allen 2020; Haidt and Allen 2020) and increasing privacy concerns from data harvesting practices (Auxier et al. 2019), policymakers need to understand whether imposing the regulation on data harvesting would increase the consumer welfare in order to resolve these harms on the users. Following the literature (Jain et al. 2020), we define consumer welfare as the total consumer utility realized from joining a social media platform. We calculate the consumer welfare in each scenario that can be formulated as $CW = \int_0^{\tilde{\theta}} u_1(\theta)\, d\theta + \int_{\tilde{\theta}}^1 u_2(\theta)\, d\theta$. We denote the consumer welfare under the *with regulation* and *no regulation* scenario by CW^R and CW^{NR}, respectively. In the following corollary, we address the question of how implementing the government's regulation on data harvesting impacts the consumer welfare.

Corollary 1. *When the government imposes the regulation on data harvesting, consumer welfare always becomes greater (i.e., $CW^R \geq CW^{NR}$).*

The intuition behind this result is as follows. The regulation impacts a user's utility in two-ways, through the net negative effect of data harvesting and demand. First, the regulation reduces the net negative effect of data harvesting (i.e., from β to γ, where $\beta \geq \gamma$) and, therefore, increases the utility that a user gains from using a platform. Second, the regulation impacts the demand of each platform that is also a part of a user's utility. We know that the demand of a platform with a stronger network effect (platform 1) always increases under the regulation (i.e., $\mathbb{E}(n_1^R) \geq \mathbb{E}(n_1^{NR})$) but the demand of the other platform decreases (i.e., $\mathbb{E}(n_2^R) \leq \mathbb{E}(n_2^{NR})$), as shown in Proposition 1(b). We find that even though the consumer welfare of platform 2 can decrease (i.e., $CW_2^R \leq CW_2^{NR}$) when its demand decreases (i.e., $\mathbb{E}(n_2^R) \leq \mathbb{E}(n_2^{NR})$) under the regulation, the overall consumer welfare always increases (i.e., $CW^R \geq CW^{NR}$). It is because the increment in the consumer welfare of platform 1 whose network effect is stronger overcomes the decrement in the consumer welfare of platform 2. In sum, the regulation always increases the overall consumer welfare.

Impacts of the Net Negative Effect of Data Harvesting on Equilibrium Outcomes Under the Regulation

When platforms are under the regulation on data harvesting, users have remaining concerns about the net negative effect of data harvesting (i.e., γ). In recent years, these concerns have increased even under the regulations (e.g., GDPR) for the following two reasons. First, some platforms have failed to comply the regulation and, consequently, increase the users' concerns about data harvesting. For example, popular platforms have faced hefty fines for misusing users' data and, thus, violating the GDPR, including Facebook ($60 million in 2022) and Twitter ($547k in 2020) (Collins 2022; Lomas 2020). Second, such violations by popular platforms have informed users about the potential privacy risks in data harvesting practices (Goswami 2020). For example, a recent survey shows that 69% of users are concerned about the way their data is collected (Internet Society 2019). All the above examples show that users' concerns about data harvesting have increased under the government's regulation. Thus, it is important to understand how the increase in the net negative effect of data harvesting on the users impacts the platforms under the regulation. In the following proposition, we attempt to answer this question.

Proposition 2. *Under the with regulation scenario, when the net negative effect of data harvesting on a user's utility, γ, is higher,*

(a) *a platform with a weaker network effect (platform 2) can actually sometimes set a higher data harvesting level* $(i.e., \frac{\partial h_2^R}{\partial \gamma} \geq 0)$*, whereas the other platform always sets a lower data harvesting level (i.e.,* $\frac{\partial h_1^R}{\partial \gamma} \leq 0$*). Interestingly,* $\frac{\partial h_2^R}{\partial \gamma} \geq 0$ *when the effort cost for platforms is higher than a given threshold (i.e.,* $p \geq \tilde{p}$*);*

(b) *the demand of a platform with a weaker network effect (platform 2) becomes greater* $(i.e., \frac{\partial \mathbb{E}(n_2^R)}{\partial \gamma} \geq 0)$*, in spite of the fact that the demand for the other platform becomes smaller (i.e.,* $\frac{\partial \mathbb{E}(n_1^R)}{\partial \gamma} \leq 0$*);*

(c) *the profit of a platform with a weaker network effect (platform 2) can actually sometimes become greater (i.e.,* $\frac{\partial \pi_2^R}{\partial \gamma} \geq 0$*), whereas the profit of the other platform always becomes smaller* $(i.e., \frac{\partial \pi_1^R}{\partial \gamma} \leq 0)$*. Interestingly,* $\frac{\partial \pi_2^R}{\partial \gamma} \geq 0$ *when the effort cost for platforms is higher than a given threshold (i.e.,* $p \geq \hat{p}$*).*

With increasing concerns about data harvesting by users, it is hard for the platforms to maintain the same level of data harvesting. For example, users may even quit from the platforms, as shown in the #DeleteFacebook campaign following the data breach incident in 2018 (Pachelli 2018). Therefore, one may expect that the platforms would set lower data harvesting levels and have smaller demands when the users are more concerned about data harvesting practices. Similarly, when users are highly concerned about data breach incident of Facebook in 2018, Facebook's revenue of social media platforms is directly related to the amount of data collected from the users (Rodriguez 2018). Therefore, one may expect that the profits would decrease when the users are more concerned about data harvesting practices. However, Proposition 2 demonstrates that none of these always holds. This result is explained as follows.

Before we proceed, it is worth to recall that a platform with a stronger network effect (platform 1) sets a higher data harvesting level than the other platform in equilibrium under the *with regulation* scenario, as shown in the discussions of Lemma 2 (i.e., $h_1^R \geq h_2^R$). This means platform 2 generates less amount of disutility from data harvesting on a user's utility than platform 1 in equilibrium (i.e., $\gamma h_1^R \geq \gamma h_2^R$). Therefore, while platform 1 has the advantage of the network effect over platform 2, platform 2 has the advantage of the disutility from data harvesting over platform 1 in equilibrium. Under the *with regulation* scenario, from Eq(3) and Eq(4), the advantage of the disutility of platform 2 can be represented as $\gamma(h_1^R - h_2^R) \geq 0$. We use this finding to explain the results in Proposition 2 below.

The intuition of Proposition 2(a) is as follows. When the effort cost of data harvesting, p, is higher than a given threshold (i.e., $p \geq \tilde{p}$), it becomes much harder for the platforms to collect data from the users. As such, with a high level of p, both platforms involve in an intense competition and set low data harvesting levels (i.e., $\frac{\partial h_1^R}{\partial p} \leq 0$ and $\frac{\partial h_2^R}{\partial p} \leq 0$). Under this case, when the net negative effect of data harvesting on a user's utility, γ, increases, platform 2's advantage of the disutility on a user's utility (i.e., $\gamma(h_1^R - h_2^R)$) increases. Therefore, platform 2 becomes more attractive to the users and has an incentive to set a higher data harvesting level to increase its revenue, leveraging its advantage (i.e., $\frac{\partial h_2^R}{\partial \gamma} \geq 0$). Since the data harvesting level of platform 2 is low, if platform 2 increases its data harvesting level, the increment in the total effort cost is less than the increment in the revenue, making platform 2 better off. In response, as γ increases, platform 1 sets a lower data harvesting level to compete against platform 2 by reducing the amount of disutility from data harvesting (i.e., $\frac{\partial h_1^R}{\partial \gamma} \leq 0$).

The result in Proposition 2(b) occurs because the increase in γ strengthens platform 2's advantage of the disutility from data harvesting (i.e., $\gamma(h_1^R - h_2^R)$ increases). In other words, the gap between the disutilities of platform 1 and 2 gets amplified as γ increases, strengthening the advantage of platform 2. Even though the demand of platform 2 increases in γ, platform 2 cannot take a larger market share than platform 1 because platform 2's advantage is always dominated by the platform 1's advantage of the network effect in equilibrium. From Eq(15) and Eq(16), we can see $\mathbb{E}(n_1^R) \geq \mathbb{E}(n_2^R)$ always holds because $\frac{n}{2}(1-\alpha) \geq \gamma(h_1^R - h_2^R)$ holds. The left-hand side represents platform 1's advantage of the network effect (i.e., $\alpha_1 = 1$ and $\alpha_2 = \alpha < 1$). The right-hand side represents platform 2's advantage of the disutility that increases in γ. This means that, while platform 1 always takes a greater market share than platform 2, some users find platform 2 more attractive and join platform 2 as γ increases. Therefore, as γ increases, the demand of platform 2 increases (i.e., $\frac{\partial \mathbb{E}(n_2^R)}{\partial \gamma} \geq 0$), whereas that of platform 1 decreases (i.e., $\frac{\partial \mathbb{E}(n_1^R)}{\partial \gamma} \leq 0$).

The result in Proposition 2(c) occurs because, as γ increases, platform 2 can leverage its increase in demand to maximize its profit in the following two cases. First, when the effort cost is higher than a given threshold but lower than $\tilde{p}$ (i.e., $\hat{p} \leq p \leq \tilde{p}$), the competition between the platforms is moderately intensive. Therefore, as γ increases, platform 2 sets a lower data harvesting level (i.e., $\frac{\partial h_2^R}{\partial \gamma} \leq 0$ when $p \leq \tilde{p}$, as shown in Proposition 2(a)) and platform 2's total effort cost decreases (i.e., $p(h_2^R)^2$ decreases). In this case, platform 2's total effort cost decreases faster than its revenue changes in γ (i.e., $\frac{\partial p(h_2^R)^2}{\partial \gamma} \leq \left|\frac{\partial \mathbb{E}(n_2^R) h_2^R}{\partial \gamma}\right|$) because its demand always increases in γ (i.e., $\frac{\partial \mathbb{E}(n_2^R)}{\partial \gamma} \geq 0$). Thus, as γ increases, platform 2 sets a lower data harvesting level to increase its overall profit, leveraging its demand (i.e., $\frac{\partial h_2^R}{\partial \gamma} \leq 0$ and $\frac{\partial \pi_2^R}{\partial \gamma} \geq 0$). Second, when the effort cost is very high (i.e., $p \geq \tilde{p}$), platform 2 sets a higher data harvesting level because of the highly intensive competition, as shown in Proposition 2(a) (i.e., $\frac{\partial h_2^R}{\partial \gamma} \geq 0$). In this case, platform 2's revenue increases faster than its total effort cost increases in γ (i.e., $\frac{\partial \mathbb{E}(n_2^R) h_2^R}{\partial \gamma} \geq \frac{\partial p(h_2^R)^2}{\partial \gamma} \geq 0$) because its demand always increases in γ (i.e., $\frac{\partial \mathbb{E}(n_2^R)}{\partial \gamma} \geq 0$). Therefore, as γ increases, platform 2 sets a higher data harvesting level to increase its overall profit, leveraging its demand (i.e., $\frac{\partial h_2^R}{\partial \gamma} \geq 0$ and $\frac{\partial \pi_2^R}{\partial \gamma} \geq 0$). For platform 1, however, its demand always decreases in γ (i.e., $\frac{\partial \mathbb{E}(n_1^R)}{\partial \gamma} \leq 0$), as shown in Proposition 2(b). This means platform 1 cannot leverage its demand as γ increases and must set a lower data harvesting level to compete against platform 2. As a result, platform 1's revenue decreases faster than its total effort cost decreases in γ (i.e., $\frac{\partial \mathbb{E}(n_1^R) h_1^R}{\partial \gamma} \leq \frac{\partial p(h_1^R)^2}{\partial \gamma} \leq 0$), resulting in an overall decrease in its profit $(i.e., \frac{\partial \pi_1^R}{\partial \gamma} \leq 0)$.

Conclusions and Managerial Insights

In the increasing concerns about the negative influences of social media on users (Lambert et al. 2022), policymakers have been considering imposing a regulation on these platforms (Deighton 2022). However, the broad literature on social media provides a limited understanding of how implementing a regulation would impact platforms and users as well as how platforms should react to the regulation. Motivated from these critical gaps, in this study, we conduct an economic analysis on the ad-supported business model of social media platforms, which is fueled by data harvesting practices. We build an analytical model to understand how imposing a regulation on data harvesting would impact social media platforms and users and how these platforms should respond to the regulation.

Our study differs from the literature on the negative impacts of social media on users in which social media platforms hurt users through advertising (Aseri et al. 2020; Hann et al. 2008). However, industry experts have argued that the fundamental reason for the negative impacts of social media on users is that platforms collect too much data from their users (Orlowski 2020). Thus, we believe that our focus on data harvesting practices adds an important business decision that has been overlooked in the broad literature on social media and users.

Our findings offer the following implications for social media platforms and policymakers. For social media platforms, the implications are two-folds. First, our results in Proposition 1(a) and 1(c) suggest that platforms should do their best in making effort to reduce the net negative effect of data harvesting on the users in order to maximize the *utility effect*. Platforms should consider how to effectively safeguard the data collected from the users, inform the users about data harvesting practices, and report to the government about data harvesting practices. Second, when a platform finds its demand decreases under the regulation, the reason could be its weaker network effect compared to the competing platform, not the regulation. In this case, the platform should consider improving its network effect by boosting social interactions among the users to better serve the users.

For policymakers, we provide four implications as follows. First, policymakers should consider how to construct the details of the effort that lead to the reduction in the net negative effect of data harvesting on the users. For example, the amount of effort required by the regulation should not be overwhelming for social media platforms in order to increase the effectiveness of the regulation. Second, when users are more concerned about potential risks related to data harvesting, policymakers may rush to impose a high effort cost to mitigate the net negative effect of data harvesting. The expected result would be that platforms set lower data harvesting levels in response to the heavy effort cost. However, our results in Proposition 2 show that this is not always the case for a platform with a weaker network effect (platform 2). Therefore, policymakers should assess the impact of the regulation on data harvesting before imposing the heavy effort cost on the platforms. Third, policymakers can be more confident on that imposing the regulation could be beneficial for both the platforms and the users. We hope our results in Proposition 1(c) and Corollary 1 together would be helpful in advancing the ongoing discussion about regulating data harvesting practices by social media platforms. Fourth, from the result in Corollary 1, policymakers should know that the overall increase in consumer welfare mostly comes from the users joining a platform that has a stronger network effect. If policymakers expect that the consumer welfare would increase in both the platform with a stronger network effect and the other with a weaker network effect, our result suggests that that is not likely the case.

References

Acquisti, A. and Gross, R. 2006. "Imagined communities: Awareness, information sharing, and privacy on the Facebook," in *International workshop on privacy enhancing technologies,* Springer, pp. 36–58.

Allcott, H., Braghieri, L., Eichmeyer, S., and Gentzkow, M. 2020. "The Welfare Effects of Social Media," *American Economic Review* (110:3), pp. 629–76.

Allcott, H., Gentzkow, M., and Song, L. 2022. "Digital addiction," *American Economic Review* (112:7), pp. 2424–63.

Allen, F. 2020. "Dying For Likes Dark Truth of Social Media as US Pre-teen Girl Suicides Soar 150% & Self-Harm Triples, Netflix's Social Dilemma Reveals," (available online at https://www.the-sun.com/news/1487147/social-media-suicides-self-harm-netflix-social-dilemma/; accessed Sept. 1, 2022).

Aseri, M., Dawande, M., Janakiraman, G., and S.. Mookerjee, V. 2020. "Ad-Blockers: a blessing or a curse?," *Information Systems Research* (31:2), pp. 627–646.

Auxier, B. and Anderson, M. 2021. "Social Media Use in 2021," (available online at https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/; accessed Sept. 1, 2022).

Auxier, B., Rainie, L., Anderson, M., Perrin, A., Kumar, M., and Turner, E. 2019. "Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information," (available online at https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/; accessed Sept. 1, 2022).

Berger, R. 2022. "Influencing The Education Of Authentic Social Media," (available online at https://www.forbes.com/sites/rodberger/2022/08/02/influencing-the-education-of-authentic-social-media; accessed Sept. 1, 2022).

Burgess, M. 2020. "What is GDPR? The summary guide to GDPR compliance in the UK," (available online at https://www.wired.co.uk/article/what-is-gdpr-uk-eu-legislation-compliance-summary-fines-2018; accessed Sept. 1, 2022).

Cao, Z., Hui, K.-L., and Xu, H. 2018. "An Economic Analysis of Peer Disclosure in Online Social Communities," *Information Systems Research* (29:3), pp. 546–566.

Chen, A. 2019. "Websites are (probably) Making Less Money Because of GDPR," (available online at https://www.technologyreview.com/2019/07/24/134067/gdpr-privacy-revenue-economics-online-business-legislation/; accessed Sept. 1, 2022).

Chen, Y.-n. and Png, I. 2003. "Information goods pricing and copyright enforcement: Welfare analysis," *Information Systems Research* (14:1), pp. 107–123.

Collins, K. 2022. "GDPR Fines: The Biggest Privacy Sanctions Handed Out So Far," (available online at https://www.cnet.com/tech/gdpr-fines-the-biggest-privacy-sanctions-handed-out-so-far/; accessed Sept. 1, 2022).

Conger, K., Wakabayashi, D., McCabe, D., and Kang, C. 2021. "Big Tech C.E.O.s Face Lawmakers on Disinformation," (available online at https://www.nytimes.com/live/2021/03/25/business/social-media-disinformation; accessed Sept. 1, 2022).

Deighton, K. 2022. "Silicon Valley Faces Pressure to Redesign Children's Online Experience," (available online at https://www.wsj.com/articles/silicon-valley-faces-pressure-to-redesign-childrens-online-experience-11645227819; accessed Sept. 1, 2022).

Demirezen, E. M., Kumar, S., and Sen, A. 2016. "Sustainability of Healthcare Information Exchanges: A Game-Theoretic Approach," *Information Systems Research* (27:2), pp. 240–258.

Dutton, J. 2022. "Facebook Sued For Billions Over Harvesting People's Data," (available online at https://www.newsweek.com/facebook-sued-billions-over-harvesting-peoples-data-1669352; accessed Sept. 1, 2022).

Edelman, G. 2021. "Facebook Quietly Makes a Big Admission," (available online at https://www.wired.com/story/facebook-quietly-makes-big-admission-political-content/; accessed Sept. 1, 2022).

European Commission. 2019. "Special Eurobarometer 487a "The General Data Protection Regulation" Report," (available online at https://europa.eu/eurobarometer/api/deliverable/download/file?deliverableId=69701; accessed Sept. 1, 2022).

Federal Trade Commission. 2019. "FTC Imposes $5 Billion Penalty and Sweeping New Privacy Restrictions on Facebook," (available online at https://www.ftc.gov/news-events/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions; accessed Sept. 1, 2022).

Federal Trade Commission. 2020. "FTC Issues Orders to Nine Social Media and Video Streaming Services Seeking Data About How They Collect, Use, and Present Information," (available online at https://www.ftc.gov/news-events/press-releases/2020/12/ftc-issues-orders-nine-social-media-video-streaming-services; accessed Sept. 1, 2022).

Global Data. 2022. "Growing Competition Among Social Media Platforms Create a Challenging 2021," (available online at https://www.verdict.co.uk/growing-competition-social-media/; accessed Sept. 1, 2022).

Goldfarb, A. and Tucker, C. E. 2011. "Privacy regulation and online advertising," *Management science* (57:1), pp. 57–71.

Gordon, M. 2019. "Dashboard Act would Force Tech Giants to Tell Users What Data They Collect, How Much It's Worth," (available online at https://www.chicagotribune.com/business/ct-biz-personal-data-value-legislation-20190624-64wyj62jajfchckcugmeurxu3q-story.html; accessed Sept. 1, 2022).

Goswami, S. 2020. "The Rising Concern Around Consumer Data And Privacy," (available online at https://www.forbes.com/sites/forbestechcouncil/2020/12/14/the-rising-concern-around-consumer-data-and-privacy/?sh=3a09696f487e; accessed Sept. 1, 2022).

Haidt, J. and Allen, N. 2020. "Digital Technology under Scrutiny," *Nature* (578:7794), pp. 226–227.

Hann, I.-H., Hui, K.-L., Lee, S.-Y. T., and Png, I. P. 2008. "Consumer privacy and marketing avoidance: A static model," *Management science* (54:6), pp. 1094–1103.

Internet Society. 2019. "The Trust Opportunity: Exploring Consumers' Attitudes to the Internet of Things," (available online at https://www.internetsociety.org/wp-content/uploads/2019/05/CI_IS_Joint_Report-EN.pdf; accessed Sept. 1, 2022).

Jain, T., Hazra, J., and Cheng, T. E. 2020. "Illegal content monitoring on social platforms," *Production and Operations Management* (29:8), pp. 1837–1857.

Jaisingh, J. 2009. "Impact of piracy on innovation at software firms and implications for piracy policy," *Decision Support Systems* (46:4), pp. 763–773.

Jiang, Z., Jones, S., Tolido, R., Hunt, G., Budor, I., Bartoli, E., Linden, P. van der, Buvat, J., Theisler, J., Wortmann, A., et al. 2019. "Championing data protection and privacy–a source of competitive advantage in the digital century," *Capgemini. Luettavissa:* (15), p. 2020.

Katz, M. L. and Shapiro, C. 1985. "Network Externalities, Competition, And Compatibility," *The American economic review* (75:3), pp. 424–440.

Kemp, S. 2022. "DIGITAL 2022: Global Overview Report," (available online at https://datareportal.com/reports/digital-2022-global-overview-report; accessed Sept. 1, 2022).

Krasnova, H., Widjaja, T., Buxmann, P., Wenninger, H., and Benbasat, I. 2015. "Why following friends can hurt you: an exploratory investigation of the effects of envy on social networking sites among college-age users," *Information Systems Research* (26:3), pp. 585–605.

Lambert, J., Barnstable, G., Minter, E., Cooper, J., and McEwan, D. 2022. "Taking a one-week break from social media improves well-being, depression, and anxiety: a randomized controlled trial," *Cyberpsychology, Behavior, and Social Networking* (25:5), pp. 287–293.

Lomas, N. 2020. "Twitter fined ∼$550K over a data breach in Ireland's first major GDPR decision," (available online at https://techcrunch.com/2020/12/15/twitter-fined-550k-over-a-data-breach-in-irelands-first-major-gdpr-decision/; accessed Sept. 1, 2022).

Meta. 2022. "Meta Reports Fourth Quarter and Full Year 2021 Results," (available online at https://investor.fb.com/investor-news/press-release-details/2022/Meta-Reports-Fourth-Quarter-and-Full-Year-2021-Results/default.aspx; accessed Sept. 1, 2022).

Orlowski, J. 2020. "The Social Dilemma," (available online at https://www.netflix.com/title/81254224; accessed Sept. 1, 2022).

Pachelli, N. 2018. "The FTC is Investigating Facebook. Here's What That Means." (Available online at https://www.esquire.com/news-politics/a19596648/facebook-federal-trade-commission-investigation/; accessed Sept. 1, 2022).

Pew Research Center. 2021. "Social Media Fact Sheet," (available online at https://www.pewresearch.org/internet/fact-sheet/social-media/; accessed Sept. 1, 2022).

Reinberg, S. 2022. "Week Off Social Media Boosts Mental Health: Study," (available online at https://www.usnews.com/news/health-news/articles/2022-05-11/week-off-social-media-boosts-mental-health-study; accessed Sept. 1, 2022).

Rodriguez, S. 2018. "Here are the Scandals and Other Incidents that Have Sent Facebook's Share Price Tanking in 2018," (available online at https://www.cnbc.com/2018/11/20/facebooks-scandals-in-2018-effect-on-stock.html; accessed Sept. 1, 2022).

Satterfield, S. 2020. "Personalized Advertising and Privacy Are Not at Odds," (available online at https://about.fb.com/news/2020/12/personalized-advertising-and-privacy-are-not-at-odds/; accessed Sept. 1, 2022).

Sherman, L. 2018. "Why Facebook Will Never Change Its Business Model," (available online at https://www.forbes.com/sites/lensherman/2018/04/16/why-facebook-will-never-change-its-business-model/?sh=2aeafc9364a7; accessed Sept. 1, 2022).

Statista. 2021. "Number of Social Network Users in the United States From 2017 to 2026," (available online at https://www.statista.com/statistics/278409/number-of-social-network-users-in-theunited-states/; accessed Sept. 1, 2022).

Statista. 2022. "Daily Time Spent on Social Networking by Internet Users Worldwide from 2012 to 2022," (available online at https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/; accessed Sept. 1, 2022).

Twitter. 2022. "Q4 and Fiscal Year 2021 Letter to Shareholders," (available online at https://s22.q4cdn.com/826641620/files/doc_financials/2021/q4/Final-Q4'21-Shareholder-letter.pdf; accessed Sept. 1, 2022).

VanDyke, E. 2022. "The Growing Similarities Between Social Media Platforms," (available online at https://globaledge.msu.edu/blog/post/56940/the-growing-similarities-between-social; accessed Sept. 1, 2022).

Wang, Y.-Y., Guo, C., Susarla, A., and Sambamurthy, V. 2021. "Online to offline: the impact of social media on offline sales in the automobile industry," *Information Systems Research* (32:2), pp. 582–604.

Wells, G., Horwitz, J., and Seetharaman, D. 2021. "Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show," (available online at https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739; accessed Sept. 1, 2022).

Wong, R. Y. M., Cheung, C. M., Xiao, B., and Thatcher, J. B. 2021. "Standing up or standing by: Understanding bystanders' proactive reporting responses to social media harassment," *Information Systems Research* (32:2), pp. 561–581.

Zhang, K. and Sarvary, M. 2015. "Differentiation with user-generated content," *Management Science* (61:4), pp. 898–914.