An Economic Analysis of Peer Disclosure in Online Social Communities

Zike Cao[†], Kai-Lung Hui[‡], and Hong Xu[‡]

July 21, 2017

Forthcoming in *Information Systems Research*

Abstract

We study a novel privacy concern: peer disclosure of sensitive personal information in online social communities. We model peer disclosure as the imposition of a negative externality on other people. Our model encompasses the benefits of posting information, positive externalities in the form of recognition and entertainment benefits due to others' sharing of information, and heterogeneous privacy preferences. We find that regulation of peer disclosure is necessary. We consider two candidate regulations – nudging and quotas. Nudging reduces user participation and privacy harm and sometimes improves social welfare. By contrast, imposing a quota often improves user participation, privacy protection, and social welfare. Adding a nudge on top of a quota does not bring additional benefits. We show that any regulation that uniformly controls the disclosure of sensitive and nonsensitive information will not serve the triple objectives of reducing privacy harm, increasing social welfare, and increasing information contribution. We derive a necessary condition for solutions that can fulfill these three objectives. We also compare the incentives of the platform owner and social planner and draw related managerial and policy implications.

Keywords: peer disclosure; privacy; regulation; online social communities; nudging; quota

†Department of Technology and Operations Management, Rotterdam School of Management, Erasmus University, 3062 PA Rotterdam, The Netherlands. ‡Department of Information Systems, Business Statistics and Operations Management, School of Business and Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. Cao: cao@rsm.nl, Hui: klhui@ust.hk, Xu: hxu@ust.hk.

1. Introduction

User-generated content is ubiquitous on social-networking websites such as Facebook and YouTube. Such content sometimes contains information (generally referring to any text, voice recording, image, and video) about other people (*peers*), which can bring unintended fame or consequences. For example, a high school teacher was fired in 2013 after her student posted a photo to Instagram of her providing alcohol and condoms in a prom after-party (*New York Daily News* 2013). A 15-year-old boy in Quebec was abused after a video of him playing a character in Star Wars went viral on the Internet. The boy made the video for a school project, but the video was shared by his friends on the Internet without his consent (*New York Times* 2003). Many children today are upset because their parents share their personal pictures or videos online (*New York Times* 2016). Generally, a person may inflict privacy harm on other people by disclosing their personal information to the public (DiMicco and Millen 2007; Tufekci 2008; Henne and Smith 2013; Choi et al. 2015).

From a social-welfare point of view, the privacy harm from peer disclosure should be balanced against the disclosure benefits. People enjoy social interaction and sharing interesting moments with friends. Friendly disclosure such as birthday greetings or achievement recognitions can bring joy to a social community. Practically, it is difficult for a social community to avoid mentioning related people in its conversations or exchanges. Hence, the pressing issue is to help users interact effectively without excessively infringing other peoples' privacy in online social communities.

The current privacy practice in online social communities mainly targets users' voluntary disclosure of their *own* information. For instance, Facebook offers users an option to restrict access to their information, including posts, profiles, and photos by other users. Users can also select whose information to view in their own timelines and remove tags about themselves from related posts and photos.[1] However, these controls do not address peer disclosure, where a user's privacy is infringed by friends' posts. Therefore, we face a novel challenge: *How does the privacy externality arising from peer disclosure of personal information affect the development of online social communities? Should we impose new policies to reduce the harm due to peer disclosure? If so, how should we*

[1] For a comprehensive review of Facebook's privacy options, see `https://www.facebook.com/help/325807937506242/` (accessed May 2017).

design the policies? Would community owners prefer such regulation?

To address these questions, we develop a stylized model that captures users' strategic decisions when they share information in an online social community. In our model, a fraction of users always participates in the community due to psychological commitment, membership, or other altruistic motivations (Hosanagar et al. 2010; Bateman et al. 2011). The other users are not committed and will strategically decide whether to join the community, taking into account the expected benefits from posting information, positive externalities from viewing posts containing others' personal information, and privacy harm resulting from the disclosure of their personal information by other people. An uncommitted user would join the community if and only if she receives a higher utility from participation than by staying out. The users differ in their privacy sensitivity.

With this model, we characterize the impacts of peer disclosure: how it affects users' decisions to join the community and post information about other people. In particular, we seek economic policies that motivate users to internalize the privacy harm caused by their posts. Two broad solutions prevail in the literature of negative externalities: Indirect and direct control. Indirect control, dating back to Pigou (1920), uses an appropriate pricing scheme that charges agents for the externalities that they impose on others. Such externality pricing has been shown to be effective in areas such as environmental protection and traffic control (Cropper and Oates 1992). A common implementation of externality pricing is to impose a tax, where the tax rate is set such that the agents would choose the efficient levels of externalities (Vickrey 1963; Sandholm 2002, 2005). In online social communities, we propose "nudging" as an alternative form of externality pricing. A nudge is a soft paternalistic measure that operates as a cue to remind users of the potential privacy damage that their posts could bring to others, or as an extra time delay in the form of a "cooling-off" period for users to consider withdrawing their posts. The purpose is to "nudge" users to think carefully about the privacy consequences of their posts (Acquisti 2009; Wang et al. 2013; Almuhimedi et al. 2015). A nudge and a Pigouvian tax have essentially similar effects, where the "tax" here is nonmonetary but exhibited in the form of additional time or effort in posting each piece of information.[2]

[2] Unlike a monetary tax, the cost due to a nonmonetary tax, such as a nudge, cannot be recovered and hence becomes a deadweight loss to the society.

To directly control negative externalities, the classical approach is to use command-and-control regulations that directly restrict agents' actions (Fullerton and Metcalf 2001). A typical implementation is to impose a quota or provide an allowance to each agent with the objective that the agent will generate the efficient level of externalities (Copes 1986; Calthrop and Proost 1998). Imposing a quota in online social communities is straightforward. We simply need to set a limit on the number or length of posts allowed for each user within a given time period.

The nudging and quota policies, corresponding to externality pricing and command-and-control regulation of negative externalities, are aligned with the practices adopted in many industries, including cigarette and alcohol taxes, pollution permits, and road-space rationing. Besides externality pricing and command-and-control regulation, the prior literature has advanced other solutions to address negative externalities, including subsidies for abatement, marketable permits, and deposit-refund systems (Stavins 2011). These solutions, however, may not be applicable in online social communities, which mostly feature large numbers of users, making the exchange of user permits practically infeasible. It is also difficult to provide subsidies or request deposits as most online social communities charge no fees to users. Another approach to address the peer disclosure externality is to deploy new technologies. For instance, using text and image processing and advanced data analytics, online social community owners may attempt to distinguish sensitive from nonsensitive personal information and directly regulate a user's disclosure of sensitive information about other people. We analyze these technical solutions in Section 4.

We find several unique results on nudging and quotas. A nudge decreases user participation and information contributions, but it also decreases the total privacy harm and sometimes increases social welfare by driving some users out of the community. By contrast, a quota preserves users' incentive to join the community and always increases social welfare, but it cannot encourage information contribution either. These findings exemplify the conflicting goals of enhancing social welfare and privacy protection vis-à-vis promoting community development in terms of increasing participation and information contribution. Our model provides a novel theoretical framework for analyzing the optimal policy designs in regulating online information contribution and peer disclosure.

We also find that quotas dominate nudging in increasing user participation and social welfare. Contrary to the prior literature, which suggests that a composite measure is more effective

in addressing externalities (Roberts and Spence 1976; Christiansen and Smith 2012), we find that nudging users on top of a quota does not bring additional benefits. Furthermore, although the social planner and community owner may variously benefit from imposing a quota, they mostly prefer different quotas because of misaligned objectives. They may prefer the same quota only when the community owner wants to grow the number of participating users. If it wants to maximize information contribution by participating users, then it will never prefer a nudge or a quota. Following this result, we derive a general necessary condition for any economic policy to reduce privacy harm and increase social welfare while increasing overall information contribution.

Our contributions are three-fold. First, we show that regulation is necessary when users can freely post information about other people in an online social community. To our knowledge, this is the first analysis addressing the privacy harm caused by peer disclosure on the Internet. Second, we illustrate the nuanced impacts of imposing a nudge and a quota, particularly their implications on user participation, which has not been formally analyzed in prior studies (Schulze and d'Arge 1974; Weitzman 1974; Collinge and Oates 1982). Third, we uncover a novel dilemma: Welfare maximization and privacy protection are not aligned with community development. We suggest some directions to resolve this dilemma, such as tailoring the nudge and quota for privacy-infringing posts or enabling users to prune sensitive information related to themsevles.

The rest of this paper is organized as follows. Section 2 reviews the related literature. Section 3 presents the model and analyzes the impacts of imposing a nudge and a quota. Section 4 derives a necessary condition for solutions that reduce privacy harm without sacrificing information contribution. Section 5 illustrates the ideas in this paper using a numerical example. Section 6 analyzes three extensions. Section 7 discusses the implications of this research and concludes the paper.

2. Related Literature

This study is closely related to the emerging stream of research studying how peer disclosure affects consumers and possible remedial actions. Choi et al. (2015) study how embarrassing posts by friends in online social networks affect individuals' perceptions of social relationships and their subsequent behavioral responses. Several studies have proposed measures to help consumers remove information shared by others without their consent (Besmer and Lipford 2010; Henne and Smith 2013). In general, the solutions involve identifying the shared information (e.g., by facial recognition

technologies) and helping affected users negotiate with the parties posting the information (e.g., by requesting removal of infringing photos). Such solutions apply *ex post*, after the information has already been posted. Hence, they are inadequate because the damage is inflicted once the information is available to the public. An ideal solution should apply *ex ante*: It should encourage people to not haphazardly post information about others. This is the focus here.

A large body of research has studied voluntary disclosure of personal information (see, e.g., Gross and Acquisti 2005; Dwyer et al. 2007; Acquisti and Gross 2009) and its regulation (Hermalin and Katz 2006; Hui and Png 2006). This literature has variously advocated the use of "privacy nudges" (Acquisti 2009; Wang et al. 2013; Almuhimedi et al. 2015), which can be visual cues about the potential audience of a post, a time delay before the post is published, or feedback on the potential sentiment and sensitivity of the post. The essential idea is to nudge users so that they will think twice about the privacy consequences of their posts. The focus of this literature lies in protecting consumer privacy in an online environment and the economic efficiency of information disclosure. It does not address the externalities due to information disclosure.

Prior research on privacy externalities focus mostly on marketing activities (Anderson and de Palma 2009; Anderson and Gans 2011; Johnson 2013). Seller marketing imposes a direct externality on consumers by either congesting consumers' attention span to process marketing promotions or increasing their costs of reading or processing the marketing. This literature has proposed solutions to help consumers address the externality. For example, Van Zandt (2004) shows that increasing senders' transmission costs using a tax or technical measures can help increase the welfare of receivers (consumers) and benefit all senders. By considering consumers' privacy harm due to seller solicitations, Hann et al. (2008) find that it is optimal to impose a charge on seller solicitations. Motivated by these suggestions, we analyze nudging as one candidate economic policy to regulate third-party externalities from peer disclosure (cf., second-party externalities from sellers).

More broadly, the economics literature has extensively analyzed the impacts of imposing a Pigouvian tax and a limit on the externality-generating activities in various contexts featuring production or consumption externalities, such as air and water pollution, smoking, and alcohol consumption (Weitzman 1974; Baumol and Oates 1988; Cropper and Oates 1992; Pizer 2002). An important consideration in this literature is entry and exit. A tax penalizes a firm and hence may force the firm to leave the industry in the long run, which can contract the industry and dampen

social welfare (Schulze and d'Arge 1974; Collinge and Oates 1982; Cropper and Oates 1992). In our setting, the privacy nudge resembles a Pigouvian tax. Hence, it is important to endogenize users' participation decisions in studying the regulation of peer disclosure in online social communities.

By contrast, limiting the externality-generating activities may have a smaller impact on participation. We consider the use of a quota as an alternative economic policy to cap or limit externalities due to peer disclosure. Prior research has also shown that one single policy, such as imposing a tax alone, may not differentiate activities that generate different degrees of externalities. Hence, adding a direct control of the externality-generating activities may further enhance social welfare (Roberts and Spence 1976; Bennear and Stavins 2007; Christiansen and Smith 2012). For example, to address the externality due to smoking, we can apply a cigarette tax and concurrently restrict the number of outlets or limit the opening hours of outlets that sell cigarettes. We adopt a similar idea and analyze the merit of combining a nudge and a quota in this paper.

Finally, our work is related to studies of negative network externalities (Liebowitz and Margolis 1994), such as the congestion externality due to free-riding in peer-to-peer file-sharing networks (Asvanund et al. 2004). The peer disclosure externality differs from congestion externalities in that it is directly imposed at the individual-user level instead of the community level. Hence, we must account for the size of the user community in analyzing its impact and regulation.

3. The Model

Consider a unit mass of users who can participate and post information in an online social community. $1-\alpha$, $0<\alpha<1$, of these users are *committed* and always participate in the community. The other α users are *uncommitted* and will participate if and only if they receive a higher utility from participation than from staying out.[3] Among all committed and uncommitted users, β, $0<\beta<1$, have high privacy sensitivity (*high types*) and $1-\beta$ have low privacy sensitivity (*low types*).

Each user is connected to some *peers* in the social community. A connection can be interpreted as a friendship link. We refer to a user's connected peers as *friends* and unconnected peers as *nonfriends*. As is the case with popular social networking websites such as Facebook or LinkedIn,

[3] Individuals may participate in online social communities because of undisclosed self-interests, psychological commitments, or other altruistic motivations (Bateman et al. 2011). The way we model heterogeneity in user participation resembles the distinction between "altruistic" and "strategic" nodes in peer-to-peer (P2P) media-distribution networks in Hosanagar et al. (2010).

the connections are undirected, i.e., two users accept each other as a friend once a connection is established. Each user, i, has a probability of $n_i \in [0, 1]$ of establishing a connection with any other user. Note that n_i can also be interpreted as the number of friends of user i because we normalize the total mass of users to 1. We use N_i to denote the set of user i's friends and $\overline{N}_i$ as the set of user i's nonfriends. For a large population of users, n_i is very small. This is consistent with the case of Facebook, which has more than 700 million users but more than 95% of them have fewer than 1,000 friends (Backstrom 2011). We start with a simple setup where every user has the same number of friends: $n_i = n$ for all i. We relax this assumption in an extension later.

We assume the user types are evenly distributed in each user's friend and nonfriend networks. Hence, for any user i, both N_i and $\overline{N}_i$ contain a proportion α of uncommitted users and a proportion β of high-type users. Figure 1 depicts the composition of the population and the connection of user i in a network with 18 users, where $\alpha = 1/2$, $\beta = 1/3$, and $n_i = 1/3$.

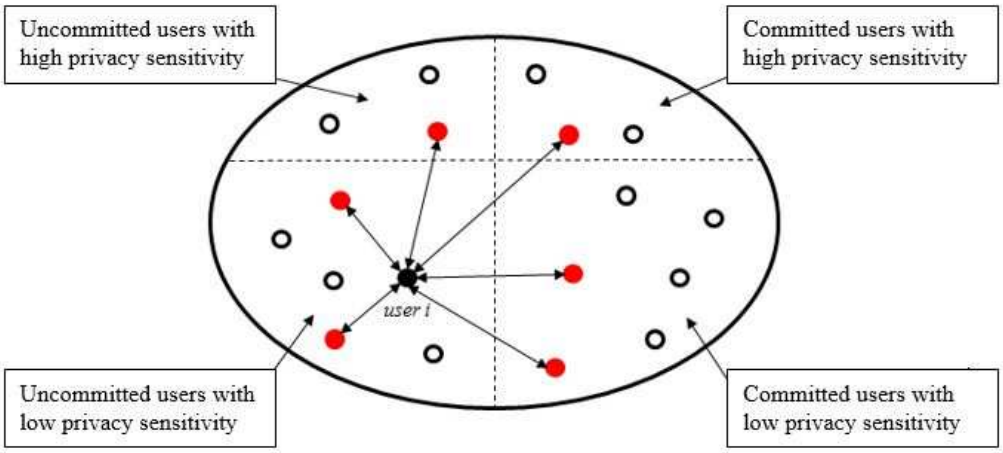

Figure 1: An example of user population and connection: Total population = 18, $\alpha = 1/2$, $\beta = 1/3$, and $n_i = 1/3$.

Each participating user can post sensitive and nonsensitive information about other people. Let x_{it} and y_{it} be the amount of nonsensitive and sensitive information that user i posts about user t, $i \neq t$. The posting of sensitive information imposes a negative externality (*privacy harm*) on user t. Evidently, each piece of sensitive information could cause different degrees of privacy harm, which will likely follow some statistical distribution. Without loss of generality, We use ρ_H (ρ_L) to denote the expected privacy harm that a high- (low-) type user suffers from the release of each piece

of sensitive information about her. We assume a participating user can post information about any other users, including nonfriends and nonparticipating users. In practice, Facebook users can share posts or photos about anyone, including celebrities who do not have a Facebook account.

A user receives a unit benefit, v, from posting each piece of information about other people. The benefit can come from the gratification of being perceived as knowledgeable, or tangible gains from advertising if the information attracts high viewership (e.g., garnering a large number of "likes" on Facebook). The cost for posting a piece of information, including the time and effort to acquire, edit, and upload it, varies by the type of information and connection. We use $C_x(x_{ij}) = \frac{1}{2}c_x x_{ij}^2$, $C_y(y_{ij}) = \frac{1}{2}c_y y_{ij}^2$, $\frac{1}{\delta}C_x(x_{ik})$, and $\frac{1}{\delta}C_y(y_{ik})$ to denote the cost functions for posting nonsensitive and sensitive information about friends, $j \in N_i$, and nonfriends, $k \in \overline{N}_i$.[4] We use j to index friends and k to index nonfriends. The convex cost functions capture the increasing difficulty of collecting and posting information as the posting volume increases.

We assume it is more costly to post sensitive information than nonsensitive information, i.e., $c_x = c$ and $c_y = \frac{c}{\psi}$, $c > 0$ and $0 < \psi < 1$. Intuitively, people guard their sensitive information such as medical history or salary more carefully. People may also feel more uncomfortable in divulging embarrassing posts about others when their own identities are observable in the community. We assume that it is more difficult to post information about nonfriends than friends, i.e., $0 < \delta \ll 1$, because of increased social distance and decreased level of trust toward nonfriends. Realistically, people post more about their online friends who are likely to be friends, relatives, classmates, or colleagues in their offline social circles (DiMicco and Millen 2007).

Besides the direct benefit from posting, a participating user also benefits from information posted by others. We use e to denote the *entertainment benefit* that a user enjoys from reading a piece of information unrelated to her posted by others (e.g., many people enjoy gossip about celebrities shared by others on Facebook). Similarly, we use w to denote the *recognition benefit* that a user enjoys when a piece of her nonsensitive information is posted by others (e.g., a person may enjoy pride when other people share the news that he/she has won an award).

---

[4] The linear benefit and quadratic cost functions give rise to diminishing marginal utility, which is a common feature in the literature because it is mathematically tractable and often guarantees an interior solution. It also fits real online social networks well. For example, no Facebook user would post all information about every other user, perhaps because doing so is prohibitively costly.

Let s be the set of participating users. User i's expected utility from participation,

$$u_{i|s}^{in} = \int_{j\in N_i} [v(x_{ij}+y_{ij}) - C_x(x_{ij}) - C_y(y_{ij})]\mathrm{d}j + \int_{k\in\overline{N}_i} \left[v(x_{ik}+y_{ik}) - \frac{1}{\delta}C_x(x_{ik}) - \frac{1}{\delta}C_y(y_{ik})\right]\mathrm{d}k$$
$$+ e\int_{m\in s, m\neq i}\left[\int_{t\neq i}(x_{mt}+y_{mt})\mathrm{d}t\right]\mathrm{d}m + w\int_{m\in s,m\neq i} x_{mi}\mathrm{d}m - \rho_i\int_{m\in s,m\neq i} y_{mi}\mathrm{d}m. \tag{1}$$

The first two integrals are user i's expected benefits from posting about her friends and non-friends. The remaining three terms capture the externalities inflicted by other users. Specifically, the third term is the entertainment benefit, the fourth term is the recognition benefit, and the last term is the privacy harm.

Let $X_{\cdot i} \equiv \int_{m\in s,m\neq i} x_{mi}dm$ be the total quantity of nonsensitive information related to user i, $Y_{\cdot i} \equiv \int_{m\in s,m\neq i} y_{mi}dm$ be the total quantity of sensitive information related to user i, and $Q_{-i} \equiv \int_{m\in s,m\neq i}\left[\int_t (x_{mt}+y_{mt})dt\right]dm$ be the total quantity of information posted by all participating users except user i.[5] With these notations, $\int_{m\in s,m\neq i}\left[\int_{t\neq i}(x_{mt}+y_{mt})dt\right]dm \equiv Q_{-i} - X_{\cdot i} - Y_{\cdot i}$. Equation (1) can be rearranged as

$$u_{i|s}^{in} = n_i\left[vx_{ij} - \frac{cx_{ij}^2}{2} + vy_{ij} - \frac{cy_{ij}^2}{2\psi}\right] + (1-n_i)\left[vx_{ik} - \frac{cx_{ik}^2}{2\delta} + vy_{ik} - \frac{cy_{ik}^2}{2\delta\psi}\right]$$
$$+ eQ_{-i} + \omega X_{\cdot i} - \theta_i Y_{\cdot i}, \tag{2}$$

where $\omega = w - e$ and $\theta_i = \rho_i + e$ represent the recognition benefit and privacy harm *net of the entertainment value* for each piece of information. Without loss of generality, we assume $\omega > 0$, $\theta_H = 1$, and $0 < \theta_L < 1$. We refer to $eQ_{-i} + \omega X_{\cdot i}$ as the "positive externalities" that user i receives from *all* information posted by other users, and $\theta_i Y_{\cdot i}$ as the privacy harm that user i suffers from information posted about her by other users.

Note that participating users can post information about nonparticipating users. We assume that a nonparticipating user is affected by the externalities caused by the information shared in the community. Realistically, people may get exposed to information that goes viral in other media,

[5] Because the size of the user population is a continuous measure, including user i's contribution (which is just a point in the integral) does not affect the aggregate sum, Q_{-i}. Continuous measures of user population are quite common in the literature (see, e.g., Daughety and Reinganum 2010; Conitzer et al. 2012). This reflects the realistic assumption that one single agent's decision will not affect the collective outcome for the population.

Table 1: Notations

α:	The fraction of uncommitted users
β:	The fraction of users with high privacy sensitivity
n_i:	The fraction of population connected with user i, with $n_i = n$ in the main model
v:	Benefit from posting each unit of information
c:	Cost coefficient of posting nonsensitive information
ψ:	Cost ratio between posting sensitive and nonsensitive information
δ:	Cost ratio between posting information about friends and nonfriends
γ:	Weighted cost ratio between posting about friends and nonfriends, $\gamma = n + (1-n)\delta$
e:	Entertainment benefit from each unit of information posted by others
w	Recognition benefit from each unit of nonsensitive information posted by others
ω:	Net recognition benefit from each unit of nonsensitive information posted by others, $\omega = w - e$
ρ_i:	User i's privacy harm from each unit of sensitive information posted by others, $i \in \{L, H\}$
θ_i:	User i's net privacy harm from each unit of sensitive information posted by others, $i \in \{L, H\}$, $\theta_i = \rho_i + e$
ϵ:	The degree of information posted within the community being exposed to outsiders
λ_i:	User i's average net privacy harm from each unit of information posted in the community, $i \in \{L, H\}$
τ:	Unit cost due to nudging
Λ:	Posting limit or quota

and celebrities can be defamed by information posted in an online community even if they are not its members. In our model, conditional on s, user i's utility from staying out,

$$u_{i|s}^{out} = \epsilon \left(eQ_{-i} + \omega X_{\cdot i} - \theta_i Y_{\cdot i} \right), \tag{3}$$

where $\epsilon \in (0, 1)$ captures how easy it is for an outsider to be exposed to (or become aware of) information posted within the community. The larger ϵ is, the easier the information posted in the community spreads outside the community. We can interpret $1 - \epsilon$ as the difference between the objective and subjective (perceived) information externalities, including privacy harm, faced by a nonparticipating user i due to the information posted in the community. Hereafter, we refer to measures scaled by ϵ as *perceived* measures and unscaled ones as *objective* measures. For example, $\theta_i Y_{\cdot i}$ is objective privacy harm whereas $\epsilon \theta_i Y_{\cdot i}$ is perceived privacy harm. Table 1 summarizes the key notations.

We study a three-stage game as shown in Figure 2. Stage 1 defines the environment, particularly whether the posting of information is regulated. In Stage 2, uncommitted users decide whether to join the community. In Stage 3, participating users decide how much information to post.

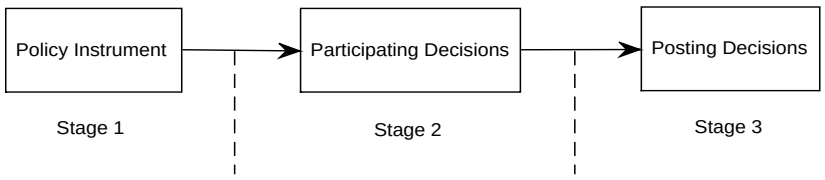

Figure 2: Timing of the Model

3.1. The Status Quo

We first consider a baseline setting where users can freely make participation and posting decisions without accounting for the privacy harm they inflict on others. We use backward induction to derive the subgame perfect equilibrium. Conditional on joining the community, user i's posting decisions, x_{ij}, x_{ik}, y_{ij} and y_{ik} can be obtained from the first-order conditions of equation (2):

$$x_{ij}^{sq} = \frac{v}{c},\ y_{ij}^{sq} = \frac{\psi v}{c},\ x_{ik}^{sq} = \frac{\delta v}{c},\ y_{ik}^{sq} = \frac{\delta \psi v}{c}. \tag{4}$$

The user would simply post information based on the ratios of posting benefits over posting costs when she omits the privacy harm caused by her. Her own privacy sensitivity will not affect her posting decisions.

However, user i's participation decision depends not only on her own posting, but also on the quantity of information others post about her. We consider a rational-expectations equilibrium in which users can anticipate other users' posting decisions. Substituting (4) into (2), user i's conditional utility from participation,

$$u_{i|s}^{sq,in} = \frac{\gamma(1+\psi)v^2}{2c} - \frac{\gamma(1+\psi)vs\lambda_i}{c},\ i \in \{L, H\}, \tag{5}$$

where $s \in [1-\alpha, 1]$ because committed users always participate in the community. To simplify the exposition, we let $\gamma \equiv n + (1-n)\delta$ (the weighted cost ratio of posting information about friends and nonfriends) and $\lambda_i \equiv \frac{\psi\theta_i - e(1+\psi) - \omega}{1+\psi}, i \in \{L, H\}$.

The first term in (5), $\frac{\gamma(1+\psi)v^2}{2c}$, is user i's total benefit from posting. The second term, $\frac{\gamma(1+\psi)vs\lambda_i}{c}$, is the net privacy harm (i.e., privacy harm net of entertainment and recognition benefits) that user i suffers from participating in the community. In the second term, $\frac{\gamma(1+\psi)v}{c} =$

$n(x_{ij}^{sq} + y_{ij}^{sq}) + (1-n)(x_{ik}^{sq} + y_{ik}^{sq})$ is the amount of information posted by each participating user, and so $\frac{\gamma(1+\psi)vs}{c}$ is the total quantity of information posted by the entire community. Therefore, we can interpret λ_i as user i's average net privacy harm caused by each piece of information posted in the community.[6]

User i's utility from staying out of the community follows equation (3):

$$u_{i|s}^{sq,out} = -\epsilon \cdot \frac{\gamma(1+\psi)vs\lambda_i}{c},\ i \in \{L, H\}. \tag{6}$$

We impose the following regularity assumption.

Assumption 1 $\theta_L > e(1+\frac{1}{\psi}) + \frac{\omega}{\psi}$.

$\theta_L \leq e(1+\frac{1}{\psi}) + \frac{\omega}{\psi}$ is equivalent to $\lambda_L \leq 0$, which means that all low-type users will participate in the status quo because their net privacy harm is negative (they benefit from other peoples' posting). Assumption 1 enables us to focus on a more realistic scenario where the privacy concern is sufficiently salient to hinder some users from participating in the community.

Given s, user i will participate if and only if $u_{i|s}^{sq,in} \geq u_{i|s}^{sq,out}$. Let s_L (s_H) be the participation rate for uncommitted low- (high-) type users. The following results characterize users' equilibrium participation as the posting benefit, v, varies.

Lemma 1 *In the status quo equilibrium, uncommitted users will participate according to the following schedule.*

v	s_L^{sq}	s_H^{sq}	s^{sq}
$(0, 2(1-\epsilon)(1-\alpha)\lambda_L)$	0	0	$1-\alpha$
$[2(1-\epsilon)(1-\alpha)\lambda_L, 2(1-\epsilon)(1-\alpha\beta)\lambda_L]$	$\frac{v/(2(1-\epsilon)\lambda_L)-(1-\alpha)}{\alpha(1-\beta)}$	0	$\frac{v}{2(1-\epsilon)\lambda_L}$
$(2(1-\epsilon)(1-\alpha\beta)\lambda_L, 2(1-\epsilon)(1-\alpha\beta)\lambda_H)$	1	0	$1-\alpha\beta$
$[2(1-\epsilon)(1-\alpha\beta)\lambda_H, 2(1-\epsilon)\lambda_H]$	1	$\frac{v/(2(1-\epsilon)\lambda_H)-(1-\alpha\beta)}{\alpha\beta}$	$\frac{v}{2(1-\epsilon)\lambda_H}$
$(2(1-\epsilon)\lambda_H, +\infty)$	1	1	1

Notes. $s^{sq} = (1-\alpha)+\alpha(1-\beta)s_L^{sq}+\alpha\beta s_H^{sq}$, i.e., the sum of all committed and uncommitted low- and high-type participating users.

In the status quo, participating users ignore the privacy harm inflicted on others. When the posting benefit $v < 2(1-\epsilon)(1-\alpha)\lambda_L$, both low- and high-type uncommitted users prefer to stay

[6] A random piece of information may or may not be privacy infringing to user i. λ_i is the expected privacy harm if a random piece of information is sensitive and related to user i minus the expected entertainment and recognition benefits otherwise.

out because the privacy harm outweighs the benefit from posting information. As v increases, the users will gradually participate by order of privacy sensitivity, and the participation rate increases with v. When v is sufficiently large, all users participate in the community.

Let Π be the social welfare, defined as the aggregate surplus of all users including the *perceived* privacy harm suffered by all nonparticipating users, let Q be the total quantity of information posted in the community, and let ξ be the total objective privacy harm including the harms inflicted on non-participating users.[7] The next lemma characterizes the outcomes in the status quo. Figure 3 illustrates how the outcomes in Lemmas 1 and 2 vary with v.

Lemma 2 *In the status quo, the total quantity of information posted, Q, and total privacy harm, ξ, increase with v. Social welfare is negative if and only if (i) $0 < v < 2[\bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H]$, or (ii) $2[\bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H] < v < 2\bar{\lambda}$ and $\beta > \frac{(1-\epsilon)\lambda_H - \lambda_L}{\lambda_H - \lambda_L}$.*

Social welfare can be negative for two reasons. Nonparticipating users carry negative utility since they are also (partially) affected by the privacy harm generated in the community. When the proportion of nonparticipating users increases as the posting benefit v decreases, the overall negative utility outweighs the positive utility from some of the participating users, therefore leading to negative social welfare. As posting benefit v increases, high-type users could still have a large negative utility even after they participate, because the positive utility from posting is lower than the privacy harm imposed on them. This could also lead to negative social welfare when there are many such highly privacy-sensitive users, i.e., β is sufficiently high.

Referring to Figure 3, one interesting observation in the status quo is that, when the social welfare is negative, a slight increase in v may further decrease social welfare. An increase in posting benefit v induces users to post more information and more uncommitted users to join the community and post information. Such additional information brings negative marginal benefit due to its privacy harm. The implication is that in some online social communities, having more users or encouraging users to post more information involving peers can be bad. Indeed, many people post unverified gossips on the Internet. Recent research has shown that online social communities or, more broadly, the Internet can help propagate wicked materials to support racial hatred, political

[7] We present the total perceived privacy harm in the Online Supplement. The analysis of objective and perceived privacy harms gives similar qualitative insights.

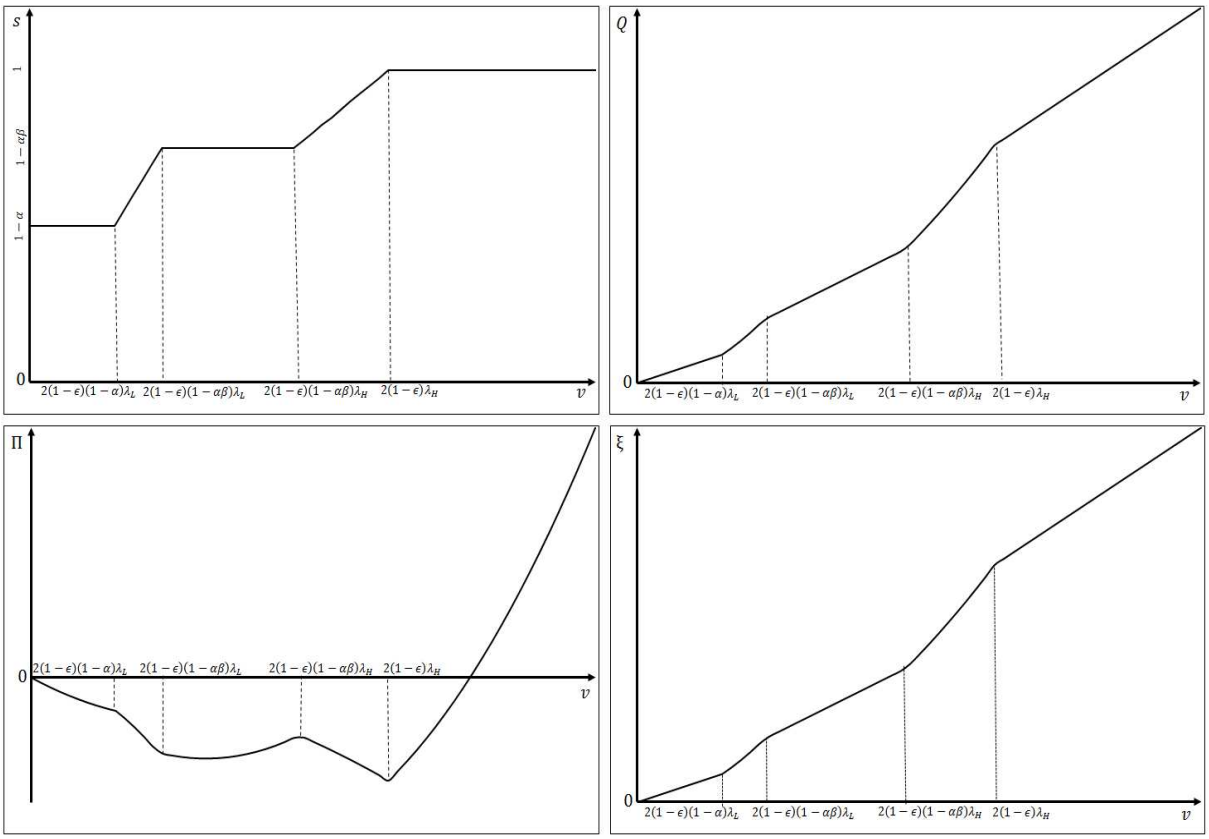

Figure 3: Equilibrium outcomes in the status quo.

flaming, and cyberbullying (see, e.g., Davis 1999; Keith and Martin 2005; Kowalski and Limber 2007; Bhuller et al. 2013; Chan et al. 2016).

As a benchmark, we next derive the first-best posting decisions, in which we assume the users account for the externalities from their posting behavior. Without loss of generality, we let $\epsilon = 1$ and derive the first-best posting strategies as:

$$x_{ij}^{fb} = \frac{v + e + \omega}{c},\ y_{ij}^{fb} = \max\left\{0, \frac{\psi(v - \theta_j)}{c}\right\},\ x_{ik}^{fb} = \frac{\delta(v + e + \omega)}{c},\ y_{ik}^{fb} = \max\left\{0, \frac{\delta\psi(v - \theta_k)}{c}\right\}. \tag{7}$$

A simple comparison of the above posting amount with those in the status quo, as shown in equations (4) shows that users in the first-best case post more nonsensitive information and less sensitive information than in the status quo, and the amount of sensitive information is contingent on the privacy sensitivity of the subjects being posted.

In the following sections, we analyze a few policies and compare their impacts on users' deci-

sions. We focus on two policies, nudge and quota, that are highly feasible because these policies uniformly apply to both sensitive and nonsensitive information. We then extend the analysis to other policies, including targeted nudge, targeted quota, and information pruning, that differentiate between the two types of information, and these policies may become feasible with future advancement in technology. We also provide implications for each policy from the perspectives of the social planner and the community owner.

3.2. Nudge

We first study the use of a privacy nudge as a regulation policy (Acquisti 2009; Wang et al. 2013; Almuhimedi et al. 2015). The online social community can remind users of the potential privacy harm that their posts may cause for other people through, e.g., visual cues or warning messages. The community can also introduce an extra time delay before a post is really publicized to allow users a "cooling-off" period during which they can revoke the post. Such privacy nudges increase users' mental efforts needed to post information and offer a chance for them to reconsider their decisions. We model the privacy nudge as imposing an additional linear posting cost on users, $\tau \in [0, v]$.[8] Intuitively, if the nudge exceeds users' posting benefit, v, then users would obtain negative benefit from posting and hence no user would post any information.

Evidently, because the nudge applies to all users and information, it increases the overall cost of posting information in the online social community. Highly drastic as it seems, mechanisms similar to a nondiscriminatory nudge are commonly proposed. For example, one renowned solution to combat music or movie piracy is to impose a tax on all blank storage media such as CD or DVD even though they are often used for legitimate purposes such as data storage. Bill Gates has famously suggested an email tax to curb spam, which inevitably affects all legitimate uses of email. We add the superscript n to all variables in the setting with a nudging policy.

With nudging, user i's utility from participation becomes:

$$\begin{aligned} u_{i|s}^{n,in} =& n_i \left[v x_{ij} - \frac{c x_{ij}^2}{2} + v y_{ij} - \frac{c y_{ij}^2}{2\psi} - \tau(x_{ij} + y_{ij}) \right] + (1 - n_i) \left[v x_{ik} - \frac{c x_{ik}^2}{2\delta} + v y_{ik} - \frac{c y_{ik}^2}{2\delta\psi} \right. \\ & \left. - \tau(x_{ik} + y_{ik}) \right] + e Q_{-i} + \omega X_{\cdot i} - \theta_i Y_{\cdot i}, \end{aligned} \tag{8}$$

[8] The findings are qualitatively similar if we use a quadratic nudging cost.

The first-order conditions of (8) give the following posting decisions:

$$x_{ij}^n = \frac{v-\tau}{c},\ y_{ij}^n = \frac{\psi(v-\tau)}{c},\ x_{ik}^n = \frac{\delta(v-\tau)}{c},\ y_{ik}^n = \frac{\psi\delta(v-\tau)}{c}. \tag{9}$$

Obviously, the nudge decreases the amount of information posted by participating users. However, it does not affect nonparticipating users. Hence, user i's utility from staying out, $u_{i|s}^{n,out}$, has the same form as equation (3). Her participation decision then depends on the comparison between $u_{i|s}^{n,in}$ and $u_{i|s}^{n,out}$. The following lemma summarizes the equilibrium participation rates.

Lemma 3 *When a nudge, $\tau \in (0, v]$, is imposed, uncommitted users will participate according to the following schedule.*

v	s_L^n	s_H^n	s^n
$(0, 2(1-\epsilon)(1-\alpha)\lambda_L + \tau)$	0	0	$1-\alpha$
$[2(1-\epsilon)(1-\alpha)\lambda_L + \tau, 2(1-\epsilon)(1-\alpha\beta)\lambda_L + \tau]$	$\frac{\frac{v-\tau}{2(1-\epsilon)\lambda_L} - (1-\alpha)}{\alpha(1-\beta)}$	0	$\frac{v-\tau}{2(1-\epsilon)\lambda_L}$
$(2(1-\epsilon)(1-\alpha\beta)\lambda_L + \tau, 2(1-\epsilon)(1-\alpha\beta)\lambda_H + \tau)$	1	0	$1-\alpha\beta$
$[2(1-\epsilon)(1-\alpha\beta)\lambda_H + \tau, 2(1-\epsilon)\lambda_H + \tau]$	1	$\frac{\frac{v-\tau}{2(1-\epsilon)\lambda_H} - (1-\alpha\beta)}{\alpha\beta}$	$\frac{v-\tau}{2(1-\epsilon)\lambda_H}$
$(2(1-\epsilon)\lambda_H + \tau, +\infty)$	1	1	1

Notes. $s^n = (1-\alpha) + \alpha(1-\beta)s_L^n + \alpha\beta s_H^n$.

By comparing Lemma 3 with Lemma 1, with nudging, a higher v is needed to encourage both types of users to participate and post information. Not surprisingly, the nudge also leads to less information posting. Hence, it effectively reduces the total privacy harm created by the community. The following proposition states these results formally.

Proposition 1 *Comparing with the status quo, a nudge reduces user participation, total quantity of information posted, and total privacy harm. Further, the participation rates, total quantity of information posted, and total privacy harm decrease in the level of nudge, τ.*

Intuitively, users may be annoyed by the privacy nudge (e.g., warning messages, time delay) and hence may post less information or even drop out from the community. Our result is consistent with previous research showing that many consumers are impatient and may drop out of online communities due to inconvenience (Galletta et al. 2006; Rajamma et al. 2009; Ding et al. 2015). It is worth noting that, although a nudge can effectively reduce the total privacy harm as it discourages

peer disclosure, it also leads to less activity in the community and lower posting, entertainment, and recognition benefits. Its overall impact on the community is determined by the tradeoffs between these benefits and costs. The following result characterizes when a nudge is socially preferred.

Proposition 2 *(i) A nudge can improve social welfare when social welfare is negative in the status quo. In this case, the socially optimal nudge is* $\tau^* = v$*, which gives social welfare of* 0*.*

(ii) A nudge always decreases social welfare when social welfare is positive in the status quo. In this case, the socially optimal nudge is $\tau^* = 0$*.*

As discussed in Lemma 2, allowing users to post information is socially undesirable when the social welfare is negative. Imposing a nudge can dissuade people from posting and hence improve social welfare. The optimal outcome is to nudge all users extensively so that they do not post any information. Social welfare will then increase from being negative to 0. A harsher nudge could be viewed as a less user-friendly interface. The results here show that when privacy externalities predominate, an easier-to-use interface may actually hurt social welfare. By contrast, when social welfare is positive, allowing users to post information brings more benefits than harm. Imposing a nudge will only increase the cost to the community. Hence, the optimal nudge is $\tau^* = 0$.

The implication of the nudging analysis is that, if a person is concerned about privacy, perhaps she should simply not join the community. Incidentally, this implication is consistent with the Chicago School's view of how privacy should be treated, although here the privacy harm arising from peer disclosure is a form of negative externality that calls for regulation (Posner 1978, 1979, 1981; Stigler 1980).

Since a nudge always decreases user participation and posting of information, it obviously is not in the interest of the community owner to impose a nudge. However, a social planner such as the government cares more about social welfare and privacy. So, the social planner prefers nudging for all v specified in Condition (i) of Proposition 2 because it helps achieve higher social welfare and lower total privacy harm. Such discrepancy in objectives helps explain why most online social communities today do not alert users about the potential adverse consequences of peer disclosure.

3.3. Quota

We next consider the merit of a quota, Λ, which is a limit on the total quantity of information that a user can post in the community. We assume $\Lambda \in \left(0, \frac{\gamma(1+\psi)v}{c}\right]$. Recall from equations (4) that

a user would post $\frac{\gamma(1+\psi)v}{c}$ units of information in the status quo. Hence, when $\Lambda > \frac{\gamma(1+\psi)v}{c}$, the quota is not binding. We refer to any $\Lambda \in (0, \frac{\gamma(1+\psi)v}{c}]$ as an "effective quota" as it will affect the user's equilibrium behavior. We say that the quota is "ineffective" otherwise.

We add the superscript q to all variables in the setting with a quota. Conditional on participation, the user's posting decisions are now subject to an additional constraint

$$n(x_{ij}^q + y_{ij}^q) + (1-n)(x_{ik}^q + y_{ik}^q) \leq \Lambda. \tag{10}$$

We compute x_{ij}^q, y_{ij}^q, x_{ik}^q and y_{ik}^q by solving equation (2) with constraint (10). Because the utility function in (2) is concave and Λ can not be greater than its unique interior solution, $\frac{\gamma(1+\psi)v}{c}$, a participating user will always use up the quota, meaning constraint (10) is binding, or $n(x_{ij}^q + y_{ij}^q) + (1-n)(x_{ik}^q + y_{ik}^q) = \Lambda$. Similar to the status quo, with an effective quota, the equilibrium quantities of information are determined by the corresponding posting costs:

$$x_{ij}^q = \frac{\Lambda}{\gamma(1+\psi)},\ y_{ij}^q = \frac{\psi\Lambda}{\gamma(1+\psi)},\ x_{ik}^q = \frac{\delta\Lambda}{\gamma(1+\psi)},\ y_{ik}^q = \frac{\delta\psi\Lambda}{\gamma(1+\psi)}. \tag{11}$$

In sharp contrast to nudging which affects community development in obvious ways (refer to Proposition 2, the platform owner prefers not to nudge any participating users), a quota does not impose any additional cost on users. Hence, it is not obvious whether the community owner's and social planner's interests are aligned. In the following analysis, we examine the optimal quota in terms of user participation, total quantity of information posted, and social welfare.

3.3.1. Participation-Maximizing Quota

We first consider user participation. For online communities, particularly those at an early stage of development, having a large user base is critical to triggering network effects among users and seeking financial support from venture capitals. We use the superscript $\star$ to denote the optimal outcomes when the objective is to maximize the number of participating users. Let ι be an infinitesimally small positive number. The following proposition characterizes the optimal quota.

Proposition 3 *Imposing a quota on the status quo will weakly increase the number of participating users. The participation-optimal schedule of quota is*

(i) When $(1-\epsilon)(1-\alpha)\lambda_L < v \leq (1-\epsilon)(1-\alpha\beta)\lambda_L$*, the optimal quota is* $\Lambda^\star = \iota$*, which gives*

participation rates $s_L^\star = \frac{(v-\iota)/((1-\epsilon)\lambda_L)-(1-\alpha)}{\alpha(1-\beta)}$, $s_H^\star = 0$, *and* $s^\star = \frac{v-\iota}{(1-\epsilon)\lambda_L}$.

(ii) When $(1-\epsilon)(1-\alpha\beta)\lambda_L < v \leq \min\{(1-\epsilon)(1-\alpha\beta)\lambda_H, 2(1-\epsilon)(1-\alpha\beta)\lambda_L\}$, *the optimal quota is* $\Lambda^\star \in \left(0, \frac{2\gamma(1+\psi)[v-(1-\epsilon)(1-\alpha\beta)\lambda_L]}{c}\right)$, *which gives participation rates* $s_L^\star = 1$, $s_H^\star = 0$, *and* $s^\star = 1-\alpha\beta$.

(iii) When $(1-\epsilon)(1-\alpha\beta)\lambda_H < v \leq (1-\epsilon)\lambda_H$, *the optimal quota is* $\Lambda^\star = \iota$, *which gives participation rates* $s_L^\star = 1$, $s_H^\star = \frac{(v-\iota)/((1-\epsilon)\lambda_H)-(1-\alpha\beta)}{\alpha\beta}$, *and* $s^\star = \frac{v-\iota}{(1-\epsilon)\lambda_H}$.

(iv) When $(1-\epsilon)\lambda_H < v < 2(1-\epsilon)\lambda_H$, *the optimal quota is* $\Lambda^\star \in \left(0, \frac{2\gamma(1+\psi)[v-(1-\epsilon)\lambda_H]}{c}\right)$, *which gives participation rates* $s_L^\star = 1$, *and* $s_H^\star = 1$, $s^\star = 1$.

(v) For all other v, *imposing a quota will not improve the participation rate relative to the status quo.*

We illustrate Proposition 3 in Panel (A) of Figure 4. Uncommitted users will participate only when they obtain higher utility from participation than from staying out, which requires the positive posting benefits and externalities to outweigh the (objective) privacy harm. An effective quota limits the amount of information that each user can post. This decrease in information causes the privacy harm to decrease faster than the positive benefits decrease, which tends to encourage users to participate in the community. As shown in Panel (A) of Figure 4, both types of user are now willing to participate with a lower posting benefit, v, when compared with the status quo.

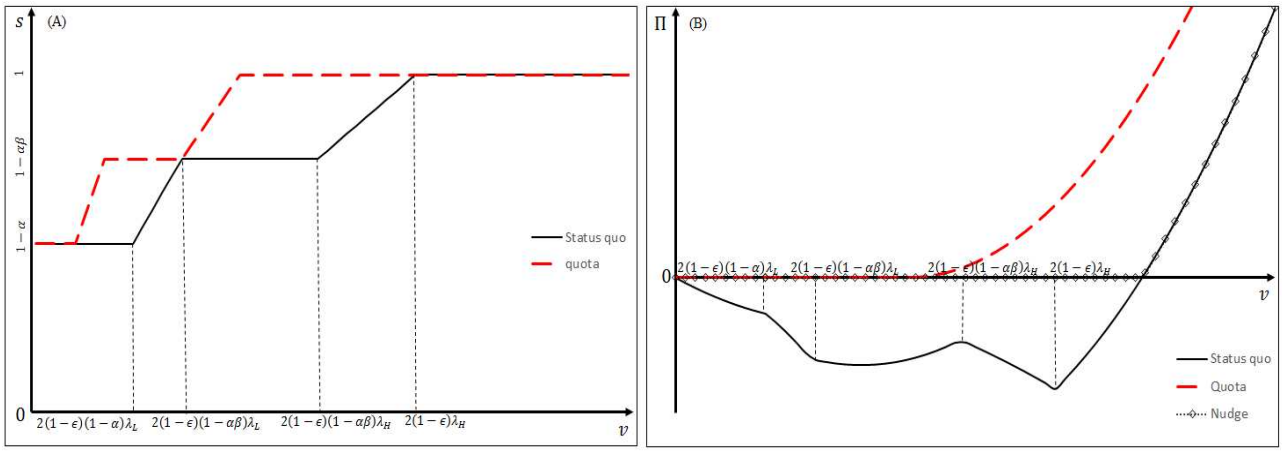

Figure 4: Participation and welfare comparison.

3.3.2. Information-Maximizing Quota

We next consider the total quantity of information. Some online communities rely on user activities to generate revenue. For example, Facebook places contextual advertisements related to user feeds. Some specialized communities such as cyberlockers obtain advertising revenues by capitalizing on users' sharing of digital materials of broad interest. The following proposition characterizes the impact of the quota when the objective is to maximize total information contribution.

Proposition 4 *Imposing any effective quota,* $\Lambda < \frac{\gamma(1+\psi)v}{c}$*, will decrease the total quantity of information posted in the community.*

A quota does not increase information contribution because it limits the amount of information that a participating user can post. Although it helps attract more users to participate, the incremental gain in information due to these marginal users does not outweigh the loss due to the reduction of contribution from *every* user. Overall, Proposition 4 suggests that imposing a quota will not help an online social community in terms of increasing information contribution.

3.3.3. Welfare-Maximizing Quota

We now consider social welfare. For ease of exposition, we present the optimal quota in Appendix A. The following proposition summarizes its impact.

Proposition 5 *Imposing a quota to the status quo will increase social welfare. The socially optimal quota presented in Appendix A weakly reduces the aggregate privacy harm and increases the number of participating users.*

Recall from Proposition 2 that a nudge can improve social welfare only by nudging users out of the community. Here, a quota encourages user participation but reduces the quantity of information posted by each user. More users will join the community and contribute information but, by Proposition 4, the total quantity of information will always decrease. This implies that both the net benefit from posting information and privacy harm will decrease for each user. Nevertheless, by the utility specification in (2), the net benefit of positing information will decrease at a slower rate than the privacy harm, leading to an overall improvement in social welfare.[9]

[9] Proposition 5 also holds if the net benefit from posting information increases linearly but the privacy harm increases

Furthermore, Lemma 2 shows that social welfare is negative when the posting benefit, v, is not high enough because of excessive posting from all participating users. Imposing a quota will also help because, by choosing a sufficiently small quota, the social planner can effectively contain the privacy harm generated. As shown in Panel (B) of Figure 4, the social planner can always achieve a positive social welfare by choosing an appropriate quota.

3.3.4. Quota vs. Nudge

By the above analysis, it is straightforward to see that the quota is better than the nudge in terms of enhancing social welfare.

Proposition 6 *The socially optimal quota weakly dominates the socially optimal nudge in improving user participation and social welfare.*

By Propositions 1 and 3, a quota increases user participation, but a nudge decreases user participation. More importantly, the quota preserves users' incentives to post information. Panel (B) of Figure 4 shows that the socially optimal quota achieves higher social welfare than the socially optimal nudge when v is sufficiently high. When v is small, allowing users to post information is not socially beneficial. Hence, both the quota and nudge apply restrictively to discourage user posting. It is worth noting that, as shown in Propositions 1 and 4, neither a nudge nor a quota can increase the total quantity of information posted in the community.

3.3.5. Quota Choice between the Community Owner and Social Planner

Interestingly, Propositions 3, 4 and 5 suggest that the community owner will never prefer a quota when it wants to maximize information contribution, but it may prefer a quota when it wants to grow the number of participating users. Will the community owner and social planner ever prefer the same quota? The answer is yes; a community owner who wants to maximize user participation may prefer the socially optimal quota. Figure 5 plots the optimal quotas that maximize the participation rate and social welfare. The optimal quotas overlap in some ranges of v, meaning they can serve both purposes. We present the formal conditions where both the community owner and social planner prefer the same quota in the Online Supplement.

---

exponentially with the amount of information posted by each user on another user. Realistically, the marginal privacy harm may increase when a user posts more information about her friends – e.g., the accumulated information may allow others to track a person with a higher precision which poses a bigger privacy threat.

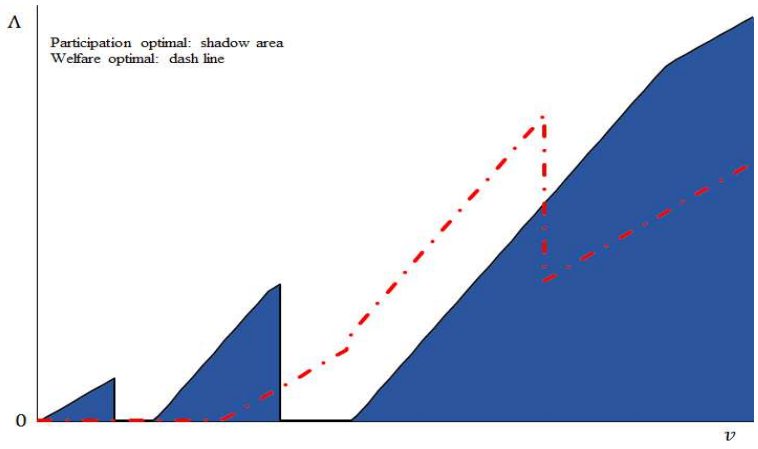

Figure 5: Participation-optimal vs. welfare-optimal quotas.

3.4. Combining Nudge and Quota

We next examine if combining a nudge and a quota gives any complementary benefit. Obviously, by Proposition 2, the optimal nudge can improve social welfare only by discouraging all users from posting information. Adding a quota is not meaningful and will not make any difference in such a scenario. Therefore, we only consider the effect of adding a nudge on top of a quota.

Let $\tilde{\Lambda} \in \left(0, \frac{\gamma(1+\psi)v}{c}\right]$ be the free quota given to each user, beyond which a nudge $\tilde{\tau} > 0$ will be applied to each additional piece of information posted by the user. Although the option of posting beyond the quota gives users more flexibility in posting information, the following results show that it does not bring any benefit to the community.

Proposition 7 *The optimal quota weakly dominates the composite policy of adding a nudge to the quota in terms of increasing user participation and social welfare.*

With an appropriate composite policy, $\tilde{\Lambda} < \frac{\gamma(1+\psi)(v-\tilde{\tau})}{c}$, the users can post beyond the quota if they are willing to be nudged. Such marginal posts generate more privacy harm on other people, which dissuades sensitive users from joining the community. With fewer participating users, the community as a whole would generate less information. Hence, social welfare will decrease because of both lower user participation and less surplus received by each participating user. Intuitively, the composite policy neutralizes the benefit of the quota. The purpose of having a quota is to reduce privacy harm so that the privacy-sensitive users will find joining the community attractive.

Allowing users to post beyond the quota, however, induces more privacy harm. This tends to drive the sensitive users out and hence counteracts the quota.

Note that such a composite policy will not increase the total quantity of information posted in the community when compared with the status quo. This is because, by Propositions 1 and 4, both quota and nudge always decrease information contribution. Hence, any combination of them will lead to further reduction of information. Together with Proposition 7, we conclude that a composite policy cannot serve the interests of either the social planner or the community owner.

4. Solving Discord between Community Owner and Social Planner

Many online social communities generate revenues by "dollarizing" their user bases. A common business model is to serve context-based advertisements. For example, Facebook feeds advertisements to users based on their activities such as "likes", shares, and comments. The success of such context-based advertising depends greatly on whether the users are active, meaning the community owner may prefer to maximize information contribution by its users. However, our results so far show that a nudge or a quota, or their combinations, will never serve this preference of the community owner despite the fact that the social planner may prefer a nudge or a quota.

In general, consider the user's utility from joining vis-à-vis staying out of the community. Subtracting equation (6) from (5),

$$u_{i|s}^{sq,in-out} = \frac{\gamma(1+\psi)v^2}{2c} - (1-\epsilon)\frac{\gamma(1+\psi)vs\lambda_i}{c} = \frac{\gamma(1+\psi)v^2}{2c} - (1-\epsilon)\cdot\lambda_i\cdot Q^{sq}(s),\ i\in\{L,H\},\ (12)$$

where $Q^{sq}(s)$ is the total quantity of information posted in the community given size of participation, s. User i will join the community if and only if $u_{i|s}^{in-out}$ is nonnegative. To address the privacy harm due to peer disclosure, we have to curb user posting of information about other people. This reduces the first term in equation (12). Because each user would post less now, to serve the community owner's interest in increasing the total quantity of information, we must increase the number of users participating in the community. However, given fixed $(1-\epsilon)\lambda_i$, such a requirement necessarily causes $Q^{sq}(s)$ in the second term of equation (12) to increase. Hence, taken together, $u_{i|s}^{sq,in-out}$ will decrease, meaning fewer users will want to participate in the community. This contradicts the

requirement of increasing user participation.

Accordingly, any feasible welfare-maximizing solutions that also serve the community owner's interest in increasing the total quantity of information must decrease $(1-\epsilon)\lambda_i$ in the second term of equation (12). The next result formalizes the necessary conditions for such solutions.

Proposition 8 *To reduce the privacy harm due to peer disclosure without decreasing the total quantity of information posted in the community, we must either reduce the average net privacy harm from posting each piece of information, λ_L or λ_H, or increase outsiders' exposure to the information posted within the community, ϵ.*

Equation (12) and Proposition 8 are important because they highlight the intricate dilemma in the peer disclosure problem: the contradictory objectives of decreasing privacy harm and maintaining information contribution. They crystallize the necessary characteristics of solutions that can address the peer disclosure problem. How can a community owner simultaneously curb peer disclosure, increase total posts, and improve social welfare? Increasing the community's visibility, ϵ, so that nonparticipating users are more aware of the information posted in the community is one way to entice users to join the community. However, increasing ϵ means that all nonparticipating users would suffer more (perceived) privacy harm, which is not desirable to the social planner. Arguably, making more nonparticipating people suffer is not a good way to boost participation and information contribution. Therefore, in the following discussion, we focus on policies that decrease the average net privacy harm from each piece of information, λ_L and λ_H.

The nudge and quota analyzed in Sections 3.2 and 3.3 do not change λ_i because they penalize sensitive and nonsensitive information uniformly, causing them to decrease by the same proportion. The community owner may be able to improve the distinction of sensitive from nonsensitive information with, for example, the latest photo-recognition technologies or text mining and natural-language processing techniques. Taking this possibility as given (i.e., we do not consider the community owner's cost of investing in technologies to detect sensitive information), we discuss some suggestive solutions to address the dilemma highlighted in Proposition 8.

4.1. Targeted Nudge and Quota

If distinguishing sensitive from nonsensitive information is possible, then the community owner can impose a nudge or a quota to target sensitive information. As a result, participating users will

post less sensitive information relative to nonsensitive information, reducing the overall privacy harm generated by the community. Formally, we add the superscripts "tn" and "tq" to all variables when a targeted nudge and a targeted quota are used. When sensitive information can be *perfectly* separated from nonsensitive information, equation (12) becomes

$$u_{i|s}^{tn,in-out} = \underbrace{\frac{\gamma v^2}{2c} + \frac{\gamma\psi(v-\tau)^2}{2c}}_{*} - \underbrace{(1-\epsilon)\cdot Q^{tn}(s)\cdot(\lambda_i - \zeta^{tn})}_{**},\ i \in \{L,H\}, \\ \text{where } \zeta^{tn} = \frac{\psi\tau(\theta_i+\omega)}{(1+\psi)[v+\psi(v-\tau)]}, \tag{13}$$

and

$$u_{i|s}^{tq,in-out} = \underbrace{\frac{\gamma v^2}{2c} + v\Lambda - \frac{c\Lambda^2}{2\gamma\psi}}_{\#} - \underbrace{(1-\epsilon)\cdot Q^{tq}(s)\cdot(\lambda_i - \zeta^{tq})}_{\#\#},\ i \in \{L,H\}, \\ \text{where } \zeta^{tq} = \frac{(\theta_i+\omega)(\gamma\psi v/c - \Lambda)}{(1+\psi)(\gamma v/c + \Lambda)}. \tag{14}$$

Note that $0 < \tau \leq v$ and $0 < \Lambda < \frac{\gamma\psi v}{c}$ because a user would post only $\frac{\gamma\psi v}{c}$ pieces of sensitive information in the status quo.

Recall $Q^{tn}(s)$ and $Q^{tq}(s)$ are the total quantities of information posted in the community given participation size, s. Comparing equations (13) and (12), $(*)$ is smaller than $\frac{\gamma(1+\psi)v^2}{2c}$ because the nudge makes posting information more costly. However, if the targeting is sufficiently accurate (i.e., ζ^{tn} is sufficiently large), then even if user participation increases leading to $Q^{tn}(s) > Q^{sq}(s)$, the second term in equation (13), $(**)$, can still be smaller than the second term in equation (12), $(1-\epsilon)Q^{sq}(s)\lambda_i$. Hence, the targeted nudge reduces the average net privacy harm due to peer disclosure, meaning the contradiction highlighted by equation (12) need not exist. A similar analysis applies to a targeted quota. When ζ^{tq} is sufficiently large, a targeted quota may increase overall information contribution but suppress privacy harm and increase social welfare.

A real-life implementation of targeted nudge is the smartphone app "ReThink" (*ABC News* 2015). It alerts users when they try to post offensive words or phrases in social media. The core component of the app is a database of offensive trigger words and phrases. In practice, such targeted information controls or nudges cannot completely accurately distinguish sensitive from nonsensitive information. They may generate false positives – nonsensitive or inoffensive information could be

wrongly detected as sensitive or offensive information, or false negatives – passing sensitive or offensive information that causes harm to others. However, as long as the targeting is sufficiently accurate, such approaches to regulating information contribution could help curb privacy harm and increase social welfare. Their deployment is also aligned with the interest of community owners because, by reducing privacy harm, more people may be willing to join online social communities, which can lead to an overall increase in information contribution.

4.2. Information Perturbation and Pruning

In the same spirit as the series of data perturbation techniques developed to protect privacy (e.g., Li and Sarkar 2006; Menon and Sarkar 2007), another possible solution to addressing the peer disclosure problem is to help users identify and prune sensitive information (e.g., automatically blurring faces or replacing faces with emoji in photos or videos). This can directly reduce the privacy harm caused by each piece of sensitive information and thus decrease λ_i. Such information pruning may cause users to obtain less pleasure in posting information, decreasing v. Referring to equation (12), this will decrease the user's posting benefit and the net privacy harm suffered from others' posting. However, as long as the pruning of sensitive information is sufficiently accurate to the extent that it reduces the privacy harm (by reducing λ_i) more than the posting benefit (due to a decrease in v), then it can be an effective solution. Arguably, blurring or substituting the faces in a picture by unobtrusive measures could eliminate most privacy harms inflicted on the involved people without significantly hurting the poster's pleasure.

5. Numerical Example

We use a numerical example to demonstrate two results from the above discussion. First, we show that a uniform quota and uniform nudge can increase social welfare and reduce privacy harm under different levels of v as shown in Lemma 1 and Propositions 1 to 5. Second, we show that properly constructed targeted nudge, targeted quota, and information pruning can resolve the conflict between social planner and community owner as characterized in Section 4, by effectively regulating privacy harm while increasing the number of participating users and social welfare.

Let $\alpha = 0.7$, $\beta = 0.5$, $n = 0.1$, $\psi = 0.5$, $\delta = 0.5$, $\epsilon = 0.1$, $e = 0.1$, $\omega = 0.01$, $c = 1$, $\theta_L = 0.5$, and $\theta_H = 1$. These values describe an online social community with many strategic users facing high costs of posting sensitive information and posting about strangers in a small friendship network.

Users in the community enjoy some recognition benefits. Users not in the community face a small chance of being affected by the privacy externality. As specified in Lemma 1, the equilibrium outcomes in the status quo differ in five ranges of v. We choose the average v in each of these five ranges and construct five sets of outcomes in Table 2.[10] In each panel of Table 2, we compare the outcomes in the status quo with the outcomes under different regulations, including uniform nudge, uniform quota, targeted nudge, targeted quota, and information pruning. For illustrative purposes, we set the uniform nudge at 30% of the posting benefit, v, and the uniform quota at 70% of the posting volume in the status quo. We set the targeted nudge at 60% of v and targeted quota at 50% of the posting volume of sensitive information in the status quo.[11] Figure 6 plots the results under the different levels of v as shown in Table 2.

Consistent with Proposition 1, a uniform nudge (UN) reduces user participation because it makes information contribution less beneficial. The total quantity of information in the community declines, leading to less privacy harm. It may increase social welfare by driving some marginal users out as Proposition 2 suggests. However, such welfare improvement is possible only when social welfare is negative in the status quo, i.e., when $v = 0.016$ or 0.051. When social welfare is positive in the status quo, i.e., when $v = 0.168$, 0.337, or 0.458, nudging decreases social welfare.

By Proposition 3, a uniform quota (UQ) helps some users obtain a higher surplus and encourages them to join the community. In our example, when $v = 0.051$ or 0.337, imposing a quota can increase user participation. It also enhances social welfare in all the scenarios as suggested in Proposition 5. However, consistent with Proposition 4, it always decreases the total quantity of information posted in the community.

One dilemma highlighted in Section 4, particularly equation (12), is that a nondiscriminatory nudge or quota cannot simultaneously increase social welfare, decrease privacy harm, and increase information contribution. We simulate the impacts of imposing a targeted nudge (TN) and a targeted quota (TQ) as analyzed in Section 4.1. We allow the targeting technologies to be imperfect:

---

[10] For the last equilibrium in Lemma 1 where both types of users participate in the community, the range of v can extend to infinity. We choose the last v in Table 2 as $2(1-\epsilon)\lambda_H + 0.5$.

[11] Note that these are not optimal nudges and quotas. The optimal nudges and quotas differ in settings with different v's. We choose these values just to illustrate that the uniform and targeted nudges and quotas indeed carry the properties and functions as analyzed in the propositions.

Table 2: Numerical example

(i) $v = 0.016$:	s	Q	Π ($\times 10^{-2}$)	ξ ($\times 10^{-2}$)
Status Quo:	0.300	0.004	-0.018	0.100
Uniform Nudge:	0.300	0.003	-0.013	0.070
Uniform Quota:	0.300	0.003	-0.012	0.070
Targeted Nudge:	0.650	0.007	-0.003	0.100
Targeted Quota:	0.650	0.007	0.004	0.088
Information Pruning:	0.300	0.004	-0.007	0.064
(ii) $v = 0.051$:	s	Q	Π ($\times 10^{-2}$)	ξ ($\times 10^{-2}$)
Status Quo:	0.475	0.020	-0.074	0.503
Uniform Nudge:	0.333	0.010	-0.036	0.246
Uniform Quota:	0.618	0.018	-0.067	0.457
Targeted Nudge:	0.650	0.021	0.022	0.316
Targeted Quota:	0.650	0.021	0.055	0.278
Information Pruning:	0.650	0.025	-0.032	0.442
(iii) $v = 0.168$:	s	Q	Π ($\times 10^{-2}$)	ξ ($\times 10^{-2}$)
Status Quo:	0.650	0.090	0.107	2.248
Uniform Nudge:	0.650	0.063	-0.083	1.574
Uniform Quota:	0.650	0.063	0.233	1.574
Targeted Nudge:	0.890	0.096	0.387	1.416
Targeted Quota:	1	0.105	0.757	1.396
Information Pruning:	0.740	0.092	0.279	1.643
(iv) $v = 0.337$:	s	Q	Π ($\times 10^{-2}$)	ξ ($\times 10^{-2}$)
Status Quo:	0.825	0.229	1.390	5.727
Uniform Nudge:	0.650	0.126	0.580	3.159
Uniform Quota:	1	0.194	1.467	4.860
Targeted Nudge:	1	0.217	2.229	3.193
Targeted Quota:	1	0.210	3.579	2.802
Information Pruning:	1	0.250	1.994	4.457
(v) $v = 0.458$:	s	Q	Π ($\times 10^{-2}$)	ξ ($\times 10^{-2}$)
Status Quo:	1	0.378	3.237	9.446
Uniform Nudge:	0.786	0.208	1.261	5.196
Uniform Quota:	1	0.264	4.083	6.612
Targeted Nudge:	1	0.295	4.546	4.345
Targeted Quota:	1	0.286	6.885	3.813
Information Pruning:	1	0.340	4.572	6.064

Notes. s: participation size; Q: total posts; Π: social welfare; ξ: total privacy harm.

Nonsensitive information can be misclassified as sensitive ("false positive") and wrongly suppressed, and sensitive information can be misclassified as nonsensitive ("false negative"). We set the probability of such mistargeting at 10%. Please refer to the Online Supplement for how we derive numerical results for targeted nudge and quota. Furthermore, to illustrate the effect of information pruning (IP) as analyzed in Section 4.2, we multiply v and λ_i by discount factors of 90% and 50%.

As shown in Table 2 and Figure 6, all three targeting and pruning measures, TN, TQ, and IP, can improve social welfare and reduce the privacy harm. They also increase total information contribution in many scenarios. Specifically, TN and TQ increase total posts when $v = 0.016$,

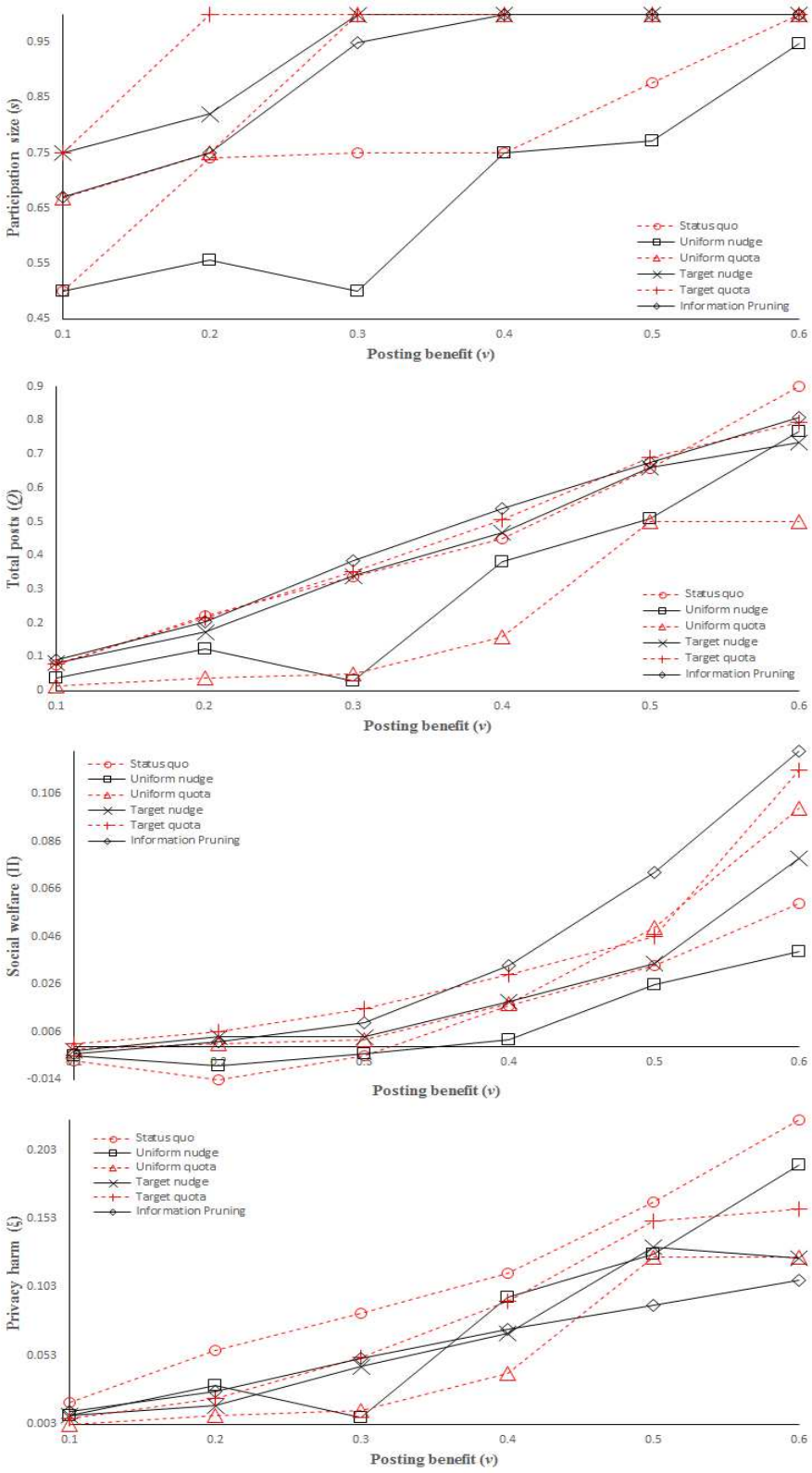

Figure 6: Numerical Example

0.051, and 0.168, and IP increases total posts when $v = 0.051$, 0.168, and 0.337. This is achieved by attracting more users to join the community. These numerical results are consistent with our analysis in Section 4, that targeted information control (TN or TQ) and information pruning (IP) can help align the interests of the social planner and the community owner. They can lead to a win–win situation – increase social welfare, decrease privacy harm, and increase total information contribution.

6. Extensions

We assess the robustness of the above analysis by relaxing several key assumptions.

(i) Heterogeneity in friendship. In this extension, we relax the assumption that all users must have the same expected number of friends. We allow for heterogeneity in users' number of friends and assume that n_i is uniformly distributed between 0 and 1, i.e., $n_i \in U[0,1]$. To ensure tractability, we let $\theta_i = 1 - n_i \in [0,1]$, meaning that users who are less sensitive about privacy have more friends. The justification is that users who are more privacy sensitive tend to minimize exposure of their personal information, and restricting their friendship network is one effective means to reduce such exposure. Previous research has shown that reciprocity is salient in social networking websites (e.g., Cha et al. 2009; Kumar et al. 2010; Weng et al. 2010), suggesting that people who share more information with others tend to have more friends, which is consistent with our assumption. To simplify the analysis and focus on the impact of heterogeneity in friendship, we assume that all users are uncommitted strategic users in this extension, i.e., $\alpha = 1$. All other setups remain the same as in the main model. The objective function for any user i with n_i friends is specified in equation (2). We characterize the equilibrium in the status quo in Lemma 4.

Lemma 4 *In the status quo, (1) when* $0 < v < (1-\epsilon)(1+\frac{1}{\delta})\frac{\psi - e(1+\psi)-\omega}{1+\psi}$, *users with* $\theta < \theta^o$ *participate in the community and users with* $\theta > \theta^o$ *stay out of the community, where* θ^o *is the solution to* $v = \frac{(1-\epsilon)[\psi\theta^o - e(1+\psi)-\omega]\theta^o[2-(1-\delta)\theta^o]}{(1+\psi)[1-(1-\delta)\theta^o]}$; *(2) when* $v \geq (1-\epsilon)(1+\frac{1}{\delta})\frac{\psi - e(1+\psi)-\omega}{1+\psi}$, *all users participate in the community.*

It is easy to verify that θ^o increases in v, meaning that the participation rate is increasing in the posting benefit. This is consistent with our findings from the main model.

We next consider the impact of a nudge. Imposing a nudge reduces a user's benefit from posting each piece of information. Hence, the impact of a nudge is similar to that of reducing the posting

benefit, v. As the participation rate is increasing in v, we expect that a nudge would decrease the participation rate. Furthermore, using a numerical example reported in the Online Supplement, we show that a nudge has a similar impact on the community's participation rate, total information contribution, total privacy harm, and aggregate user surplus as in the main model.

The analysis of a quota is less straightforward because a uniform quota is no longer appropriate when users adopt different posting strategies based on their numbers of friends. We consider a quota of the following form: $\Lambda_i = f \cdot \frac{\gamma_i(1+\psi)v}{c}$, where $f \in [0,1]$ and $\frac{\gamma_i(1+\psi)v}{c}$ is the total amount of information that user i would post in the status quo. In other words, users face different quotas contingent on their numbers of friends. We show in the Online Supplement that there exists a quota that (weakly) increases the participation rate and decreases total information contribution and total privacy harm caused by the community, and hence exhibits similar effects as the quota in the main model. We further use numerical analysis (reported in the Online Supplement) to show that a properly designed quota improves the social welfare under different parameters.

(ii) Nonlinear externality. We consider a scenario when the externalities, e, ω, and θ increase with the number of participating users. Realistically, the impact of disclosing sensitive information may increase with audience size. A larger audience increases the chance that the information resonates with interested friends or acquaintances. Let $e(s) = es$, $\omega(s) = \omega s$, $\theta_L(s) \equiv \theta_L s$, and $\theta_H(s) \equiv \theta_H s = s$, where $s \in (0,1)$ is the fraction of participating users. With these changes, the *total* privacy harm that a user suffers becomes a convex function.

We report the detailed equilibrium outcomes in the Online Supplement. Lemma 5 characterizes users' participation incentives.

Lemma 5 *When the externalities increase with the number of participating users, the uncommitted users are more likely to participate in the status quo.*

Recall uncommitted low-type users join the community at a lower v than uncommitted high-type users. When low-type users deliberate their participation decisions, they enjoy less externality because the size of participating users is small, $s < 1$. Hence, low-type users will suffer less privacy harm because the community is still small and so will join when v is lower. When v increases, some uncommitted high-type users will gradually join. These early high-type users also suffer less privacy harm and thus have more incentive to join. Here again, all of our earlier results regarding

the merits of the regulations continue to apply.

(iii) Unintended Disclosure. In our model, users make posting decisions about sensitive and nonsensitive information, meaning they intentionally divulge others' sensitive information. In the Online Supplement, we analyze a variant in which users make posting decisions about only one set of information containing both sensitive and nonsensitive information. In other words, they disclose others' sensitive information unintentionally. We assume an exogenous fraction (unknown to the user) of the posted information causes privacy harm to others. All results in the main model continue to apply in this new setting.

7. Discussion and Implications

Using a stylized model, we show that regulation is necessary to control peer disclosure in an online social community. Depending on the benefit from posting information, the community may have too many participants and the participants may post too much information about other people. Although many countries legislate explicit privacy laws to protect consumer privacy, most of these regulations focus on the merchant–consumer relationship involving direct privacy infringement, not third-party privacy harm such as peer disclosure.[12] In privacy disputes resulting from peer disclosure, individuals may pursue a defamation or libel lawsuit against the insulting party. However, such cases are rare because not every sensitive statement on social media can be considered as the basis for a defamation or libel lawsuit. For example, the mere disclosure of the whereabouts of a person may cause the person to lose her job because of dereliction of duties, but such disclosure does not constitute any defamation or libel. Furthermore, suing people for offensive or disturbing messages may have a negative impact on free speech (*The Telegraph* 2012). Practically, it is not feasible to stipulate what a person can say about her friends and peers. Defining and sanctioning peer disclosure could be immensely difficult or costly too. Hence, explicit legislation is not likely to be a practicable solution.

We propose two implementable policies: nudge and quota. A carefully selected nudge or quota can help enhance user welfare, but they work differently. A nudge helps by driving users who are concerned about privacy out of the community and suppressing those who participate from posting

[12] For example, the European Parliament and Council Directive 95/46/EC states that it "shall not apply to the processing of personal data...by a natural person in the course of a purely personal or household activity."

information. A quota helps by reducing the amount of information posted by each user and hence reducing the privacy harm and preserving participation incentives. We show that the community owner will never prefer nudging and mostly does not prefer a quota either except to grow its user base. We present an important necessary condition, Proposition 8, for any regulation to achieve the triple objectives of enhancing social welfare, reducing privacy harm, and increasing information contribution. Based on this condition, we propose three solutions that selectively target different kinds of information.

One immediate insight from our analysis in Sections 3.1 and 3.2 is that, lacking any regulation, some users should simply not join online social communities with mediocre benefits from information sharing (cf., the magnitude of privacy harm from peer disclosure). These users should be excluded not because they are "harmful" to other people. Instead, it is because they are more vulnerable to the privacy harm from peer disclosure. The practical implication is that if a person is sensitive about privacy, she should not join online social communities with particularly intimate themes, such as those promoting extramarital affairs or socially improper behaviors such as drug consumption. Similarly, users who are not ready for politically charged discussion or abusive comments with real personal identities should stay out from communities predominated by users who like to post information about, confront, or abuse others.

Notwithstanding this insight, regulation is necessary to enhance the collective welfare of users in online social communities. Our nuanced consideration of user participation and information contribution helps explain why almost no online social community is eager to nudge users despite the fact that nudging has been repeatedly advocated (Acquisti 2009; Acquisti et al. 2013; Wang et al. 2013). It also highlights the disadvantage of imposing a quota, that it preserves users' incentives to join the community but decreases overall information contribution.

Ideally, we want to limit the privacy harm due to peer disclosure but encourage users to participate in the community and contribute more information. Section 4 discusses three solutions that either impose a nudge or a quota selectively on sensitive information, or perturb or prune the sensitive information while retaining its informational benefits. Section 5 shows that these solutions can increase social welfare and are aligned with the community owner's interest too. However, they require sophisticated technologies that can identify and target sensitive personal information reasonably accurately. Such technologies may be impracticable or too expensive and

hence not all owners of online social communities are willing to develop and deploy them.

Lacking an extrinsic motivation to address privacy harm, how can the community owner be motivated to adopt these regulatory policies (either impose a uniform nudge or quota at the cost of less information contribution, or invest in relevant technologies to target sensitive information)? A promising direction is to attribute part of the damage from the privacy harm to the community owner so that it has an incentive to address the harm suffered by privacy-sensitive users. For example, the regulator can help victims take legal actions and seek compensation from the owner of an online social community when the privacy damage from peer disclosure is excessive. Such legal sanctions against platform owners who do not directly impose the damage is not without precedent. For example, the U.S. government shut down MegaUpload.com, a cyberlocker helping users download movies or music shared by other users, because it "contributed" to the infringement of the copyright of affected intellectual property owners. It is common for plaintiffs in defamation cases to sue also the media for contributing to the damage caused by other people (a situation highly similar to "peer disclosure" in our setting). If the platform owner can be held liable for the information disclosed by its users about their peers, then it should have a stronger, vested incentive to regulate user behaviors.[13]

Although our analysis is framed on peer *disclosure*, our insights extend to other settings where users impose negative externalities on peers. One example is game invitations on online social communities. Increasingly, games designed for mobile devices encourage players to send invitations to friends before granting additional game credits to the players. This promotional tactic has caused many players to invite friends to try the games, which arguably creates annoyance and inconvenience to peers. Although this practice does not involve disclosure, it intrudes on the peers' private space and so threatens the seclusion aspect of privacy (Stigler 1980). As such, our analysis directly applies. Nudging or imposing a limit on such "invitations" would help enhance the aggregate user welfare. In fact, we contend that platform owners may have a higher incentive to nudge or cap such game invitations because they are less vested in these promotions.

[13] Because we have not developed a utility function for the platform owner, we cannot explicitly analyze how this allocation of damage would affect the extent of regulation and the equilibrium outcomes. We do not model the platform owner's utility because its decision may not be driven purely by economic considerations. We defer the study of the platform owner's decisions and utility to future research.

Numerous instances of externality-curbing policies exist in other peer-to-peer (P2P) applications. For example, BitTorrent, a popular P2P file sharing protocol, reduces network traffic congestion by reducing the download speed for free riders (Hosanagar et al. 2010). Online music streaming services such as Spotify and Pandora have been pressed by music labels to limit free streaming access in order to alleviate the negative externalities imposed on paying users and musicians (Gigaom.com 2013; *The Independent* 2015). Facebook has experimented with charging fees to users who bombard celebrities with unwanted messages to limit the negative externalities generated from such harassment (*The Independent* 2013). Our framework provides a basis for analyzing the optimal choices of policy instruments in these applications.

7.1. Implementation

Given our conclusion that regulation is necessary and the identification of the properties of a good regulatory policy, this paper has made an important first step towards improving the privacy and welfare for users participating in online social communities. The next step is to determine how to implement the right nudges or quotas. Lacking an accurate account of privacy externality, we offer the following guidance.

The first step is to measure the size and privacy sensitivity of users. It may not be feasible to directly poll users about their privacy preferences. Users may not respond to such polls and, even if they do, their responses are likely biased because people tend to exaggerate their privacy needs (Harper and Singleton 2001; Hui et al. 2007; Vasalou et al. 2011). An alternative is to infer users' preferences from their activities. For example, Facebook can track the frequency of users untagging their names from photos shared by their friends. Google can track delisting requests related to identity removal. Such data can help construct users' privacy profiles. A potential challenge with this approach is accounting for selection bias as we can only observe the behavior of participating users. This selection bias poses a smaller threat when the participating group is large relative to the nonparticipating group, which is likely the case for large communities such as Facebook or Google. Nonetheless, measuring the privacy preferences of people not taking part in online social communities is a good topic for future research.

To set the nudge or quota levels, the community owner can gauge users' posting benefit, v, using standard marketing techniques such as conjoint analysis (e.g., Hann et al. 2007; Krasnova et al. 2009) or field experiments (e.g., Hui et al. 2007; Beresford et al. 2012). Users' posting cost,

c, consisting of the cost to collect and post peers' information, can be calibrated based on the nature of the community and the technological sophistication of the users. For example, users in a community of indecent affairs or paparazzi are likely to incur a higher posting cost as the underlying content is more privacy sensitive and difficult to obtain. By contrast, for a community targeting the mass market such as Facebook, users tend to bear much lower information collection and posting costs.

The implementation of a nudge or a quota also requires quantification of personal information because different types of personal information vary in privacy sensitivity. Previous research has attempted to quantify the value and sensitivity of personal data (e.g., Hui et al. 2007; Hann et al. 2007). Similar methodologies can be extended to other personal data such as photos or videos. A nudge or a quota can then be applied directly to each piece of "unitized" information.

Nudging can take different forms in practice, such as warning messages or visual cues. As a user gains experience, they may become unresponsive to privacy nudges. To ensure the salience of the privacy nudges, the community owner can include a time delay whenever a nudge is applied, for example, by forcing users to read the warning messages. It can also regularly change how and when the warning messages or cues are displayed, or the content of the messages or cues itself. By doing so, users will less likely skip the privacy cues.

7.2. Limitations and Future Research

Our analysis has several limitations. First, for ease of tractability, we assume two types of users, which allows us to show the responses of users with differing privacy sensitivity to the regulations. Future research should extend the analysis to more heterogeneous users. It may also model homophily which could affect how users form friendships and engage in peer disclosure.

Second, we consider the community owner's interest in maximizing user participation or posting of information, but we have not developed its objective function. Constructing such a function may help us gain a holistic view of social welfare and the equilibrium behaviors. The challenge lies in how to reasonably capture the differing objectives of online social communities.

Third, this analysis is confined to one online social community and does not consider "multi-homing" (Koh and Fichman 2014). An interesting extension is to allow users to choose between communities and study how their privacy preferences and peer disclosure interact.

Finally, because of complexity with interplays of many factors such as user commitment, pri-

vacy concerns, information sensitivity, costs, and the modeling of friendship-network structure, we cannot derive an unambiguous comparative static analysis. Hence, we cannot conclude if, for example, having more committed or privacy-sensitive users will favor a nudge or a quota, or whether regulation is more or less important when the user demographic changes in a particular direction. Developing a more parsimonious model may help overcome this difficulty. However, balancing parsimony and richness of insights is an obvious challenge.

8. Concluding Remarks

In 2017, the number of active users on Facebook, WeChat, Instagram, Twitter, and Pinterest were 1.97 billion, 889 million, 600 million, 319 million, and 150 million, respectively (Statistica 2017). There are around 3.2 billion Internet users. This mean that Facebook alone has more than 60% penetration. Evidently, these online platforms present people with novel avenues for social interaction. New research is necessary to uncover the implications of such interaction with unprecedented reach and scale.

This paper analyzes one novel behavior in online social communities, peer disclosure of personal information. The sharing of information among friends is mostly unregulated, but its consequence is starting to surface. For example, there have been cases when people were sacked from work because of friends' posting of their improper behavior in online social communities, and crime syndicates have used data harvested from online social networks to track targeted victims. Studying the implications of peer disclosure and its regulation is an important first step towards shaping a healthy online environment for social interaction. This study serves that purpose.

We find that regulation is necessary. The choice of regulation depends on how privacy is treated in the jurisdiction. A nudge is helpful if privacy is absolutely preferred, whereas a quota is better if we focus on economic utility and are willing to trade privacy for the pleasure of sharing information. Most importantly, having more users need not be good for the society. Any thoughtful analysis of the benefits of online social communities should consider the pros and cons of user participation and exit and the related benefits and damages in a holistic framework.

References

ABC News. 2015. 15-year-old's "rethink" app aims to prevent cyberbullying. URL `http://abcnews.go.com/Lifestyle/15-year-olds-rethink-app-aims-prevent-cyberbullying/story?id=33329748`.

Acquisti, Alessandro. 2009. Nudging privacy: The behavioral economics of personal information. *IEEE Security & Privacy*

7(6) 82–85.

Acquisti, Alessandro, Laura Brandimarte, Idris Adjerid. 2013. Gone in 15 seconds: The limit of privacy transparency and control. *IEEE Security & Privacy* **11**(4) 72–74.

Acquisti, Alessandro, Ralph Gross. 2009. Predicting Social Security Numbers from public data. *Proceedings of the National Academy of Sciences* **106**(27) 10975–10980.

Almuhimedi, Hazim, Florian Schaub, Norman Sadeh, Idris Adjerid, Alessandro Acquisti, Joshua Gluck, Lorrie Faith Cranor, Yuvraj Agarwal. 2015. Your location has been shared 5,398 times!: A field study on mobile app privacy nudging. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15, ACM, New York, 787–796.

Anderson, Simon P., Andre de Palma. 2009. Information congestion. *The RAND Journal of Economics* **40**(4) 688–709.

Anderson, Simon P., Joshua Gans. 2011. Platform siphoning: Ad-avoidance and media content. *American Economic Journal: Microeconomics* **3**(4) 1–34.

Asvanund, Atip, Karen Clay, Ramayya Krishnan, Michael D. Smith. 2004. An empirical analysis of network externalities in peer-to-peer music-sharing networks. *Information Systems Research* **15**(2) 155–174.

Backstrom, Lars. 2011. Anatomy of Facebook. *Facebook Data Team.* URL `https://www.facebook.com/notes/facebook-data-team/anatomy-of-facebook/10150388519243859`.

Bateman, Patrick J., Peter H. Gray, Brian S. Butler. 2011. Research note – The impact of community commitment on participation in online communities. *Information Systems Research* **22**(4) 841–854.

Baumol, William J., Wallace E. Oates. 1988. *The Theory of Environmental Policy (2nd Edition)*. Cambridge University Press, Cambridge.

Bennear, Lori Snyder, Robert N. Stavins. 2007. Second-best theory and the use of multiple policy instruments. *Environmental and Resource Economics* **37**(1) 111–129.

Beresford, Alastair R., Dorothea Kubler, Soren Preibusch. 2012. Unwillingness to pay for privacy: A field experiment. *Economics Letters* **117**(1) 25–27.

Besmer, Andrew, Heather R. Lipford. 2010. Moving beyond untagging: Photo privacy in a tagged world. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 1563–1572.

Bhuller, Manudeep, Tarjei Havnes, Edwin Leuven, Magne Mogstad. 2013. Broadband Internet: An information superhighway to sex crime? *Review of Economic Studies* **80**(4) 1237–1266.

Calthrop, Edward, Stef Proost. 1998. Road transport externalities. *Environmental and Resource Economics* **11**(3) 335–348.

Cha, Meeyoung, Alan Mislove, Krishna P. Gummadi. 2009. A measurement-driven analysis of information propagation in the Flickr social network. *Proceedings of the 18th International Conference on World Wide Web, ACM* 721–730.

Chan, Jason, Anindya Ghose, Robert Seamans. 2016. The Internet and racial hate crime: Offline spillovers from online access. *MIS Quarterly* **40**(2) 381–403.

Choi, Ben C. F., Zhenhui (Jack) Jiang, Bo Xiao, Sung S. Kim. 2015. Embarrassing exposures in online social networks: An integrated perspective of privacy invasion and relationship bonding. *Information Systems Research* **26**(4) 675–694.

Christiansen, Vidar, Stephen Smith. 2012. Externality-correcting taxes and regulation. *The Scandinavian Journal of Economics* **114**(2) 358–383.

Collinge, Robert A., Wallace E. Oates. 1982. Efficiency in pollution control in the short and long runs: A system of rental emission permits. *The Canadian Journal of Economics* **15**(2) 346–354.

Conitzer, Vincent, Curtis R. Taylor, Liad Wagman. 2012. Hide and seek: Costly consumer privacy in a market with repeat purchases. *Marketing Science* **31**(2) 277–292.

Copes, Parzival. 1986. A critical review of the individual quota as a device in fisheries management. *Land Economics* **62**(3) 278–291.

Cropper, Maureen L., Wallace E. Oates. 1992. Environmental economics: A survey. *Journal of Economic Literature* **30**(2) 675–740.

Daughety, Andrew F., Jennifer F. Reinganum. 2010. Public goods, social pressure, and the choice between privacy and publicity. *American Economic Journal: Microeconomics* **2**(2) 191–221.

Davis, Richard. 1999. *The Web of Politics: The Internet's Impact on the American Political System*. Oxford University Press, New York.

DiMicco, Joan Morris, David R. Millen. 2007. Identity management: Multiple presentations of self in Facebook. *Proceedings of International ACM Conference on Supporting Group Work* 383–386.

Ding, Amy Wenxuan, Shibo Li, Patrali Chatterjee. 2015. Learning user real-time intent for optimal dynamic web page transformation. *Information Systems Research* **26**(2) 339–359.

Dwyer, Catherine, Starr Roxanne Hiltz, Katia Passerini. 2007. Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. *Proceedings of the American Conference on Information Systems* 339.

Fullerton, Don, Gilbert E. Metcalf. 2001. Environmental controls, scarcity rents, and pre-existing distortions. *Journal of Public Economics* **80**(2) 249–267.

Galletta, Dennis F., Raymond M. Henry, Scott McCoy, Peter Polak. 2006. When the wait isn't so bad: The interacting effects of website delay, familiarity, and breadth. *Information Systems Research* **17**(1) 20–37.

Gigaom.com. 2013. Pandora caps monthly free tunes on mobiles to 40 hours. URL `https://gigaom.com/2013/02/28/pandora-caps-monthly-free-tunes-on-mobiles-to-40-hours/`.

Gross, Ralph, Alessandro Acquisti. 2005. Information revelation and privacy in online social networks: The Facebook case. *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society* 71–80.

Hann, Il-Horn, Kai-Lung Hui, Sang-Yong Tom Lee, Ivan P. L. Png. 2007. Overcoming online information privacy concerns: An information-processing theory approach. *Journal of Management Information Systems* **24**(2) 13–42.

Hann, Il-Horn, Kai-Lung Hui, Sang-Yong Tom Lee, Ivan P. L. Png. 2008. Consumer privacy and marketing avoidance: A static model. *Management Science* **54**(6) 1094–1103.

Harper, Jim, Solveig Singleton. 2001. With a grain of salt: What consumer privacy surveys don't tell us. *Competitive Enterprise Institute* 1–18.

Henne, Benjamin, Matthew Smith. 2013. Awareness about photos on the web and how privacy-privacy-tradeoffs could help. *International Conference on Financial Cryptography and Data Security*. Springer, 131–148.

Hermalin, Benjamin E., Michael L. Katz. 2006. Privacy, property rights and efficiency: The economics of privacy as secrecy. *Quantitative Marketing and Economics* **4**(2) 209–239.

Hosanagar, Kartik, Peng Han, Yong Tan. 2010. Diffusion models for peer-to-peer (P2P) media distribution: On the impact of decentralized, constrained supply. *Information Systems Research* **21**(2) 271–287.

Hui, Kai-Lung, Ivan P. L. Png. 2006. The economics of privacy. *Handbook of Economics and Information Systems* 471–497.

Hui, Kai-Lung, Hock Hai Teo, Sang-Yong Tom Lee. 2007. The value of privacy assurance: An exploratory field experiment. *MIS Quarterly* **31**(1) 19–33.

Johnson, Justin P. 2013. Targeted advertising and advertising avoidance. *The RAND Journal of Economics* **44**(1) 128–144.

Keith, Susan, Michelle E. Martin. 2005. Cyber-bullying: Creating a culture of respect in a cyber world. *Reclaiming Children and Youth* **13**(4) 224–228.

Koh, Tat-Koon, Mark Fichman. 2014. Multi-homing users' preferences for two-sided exchange networks. *MIS Quarterly* **38**(4) 977–996.

Kowalski, Robin M., Susan P. Limber. 2007. Electronic bullying among middle school students. *Journal of Adolescent Health* **41**(6) S22–S30.

Krasnova, Hanna, Thomas Hildebrand, Oliver Guenther. 2009. Investigating the value of privacy in online social networks: Conjoint analysis. *International Conference on Information Systems*.

Kumar, Ravi, Jasmine Novak, Andrew Tomkins. 2010. Structure and evolution of online social networks. *Link Mining: Models, Algorithms, and Applications*. Springer, 337–357.

Li, Xiao-Bai, Sumit Sarkar. 2006. Privacy protection in data mining: A perturbation approach for categorical data. *Information Systems Research* **17**(3) 254–270.

Liebowitz, S. J., Stephen E. Margolis. 1994. Network externality: An uncommon tragedy. *The Journal of Economic Perspectives* **8**(2) 133–150.

Menon, Syam, Sumit Sarkar. 2007. Minimizing information loss and preserving privacy. *Management Science* **53**(1) 101–116.

Pigou, Arthur Cecil. 1920. *The Economics of Welfare*. Palgrave Macmillan, London.

Pizer, William A. 2002. Combining price and quantity controls to mitigate global climate change. *Journal of Public Economics* **85**(3) 409–434.

Posner, Richard A. 1978. An economic theory of privacy. *Regulation* **9**(3) 19–26.

Posner, Richard A. 1979. Privacy, secrecy, and reputation. *Buffalo Law Review* **28** 1–55.

Posner, Richard A. 1981. The economics of privacy. *American Economic Review* **71**(2) 405–409.

Rajamma, Rajasree K., Audhesh K. Paswan, Muhammad M. Hossain. 2009. Why do shoppers abandon shopping cart? Perceived waiting time, risk, and transaction inconvenience. *Journal of Product & Brand Management* **18**(3) 188–197.

Roberts, Marc J., Michael Spence. 1976. Effluent charges and licenses under uncertainty. *Journal of Public Economics* **5**(3-4) 193–208.

Sandholm, William H. 2002. Evolutionary implementation and congestion pricing. *The Review of Economic Studies* **69** 667–689.

Sandholm, William H. 2005. Negative externalities and evolutionary implementation. *The Review of Economic Studies* **72**(3) 885–915.

Schulze, William, Ralph C. d'Arge. 1974. The Coase proposition, information constraints, and long-run equilibrium. *The American Economic Review* **64**(4) 763–772.

Statistica. 2017. Most famous social network sites worldwide as of April 2017, ranked by number of active users (in millions). URL `https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/`.

Stavins, Robert N. 2011. The problem of the commons: Still unsettled after 100 years. *The American Economic Review* **101**(1) 81–108.

Stigler, George J. 1980. An introduction to privacy in economics and politics. *Journal of Legal Studies* **9**(4) 623–644.

New York Daily News. 2013. Florida teacher fired after she rented party penthouse for students that included alcohol, condoms. URL `http://www.nydailynews.com/news/national/fla-teacher-fired-allegedly-giving-students-alcohol-condoms-article-1.1488471`.

New York Times. 2003. Fame is no laughing matter for the 'star wars kid'. URL `http://www.nytimes.com/2003/05/19/business/compressed-data-fame-is-no-laughing-matter-for-the-star-wars-kid.html`.

New York Times. 2016. Don't post about me on social media, children say. URL `http://well.blogs.nytimes.com/2016/03/08/dont-post-about-me-on-social-media-children-say`.

The Independent. 2013. Facebook now charges you for messages sent to celebrities and people you aren't friends with. URL `http://www.independent.co.uk/news/uk/home-news/facebook-now-charges-you-for-messages-sent-to-celebrities-and-people-you-arent-friends-with-8563299.html`.

The Independent. 2015. Spotify reportedly under pressure from music labels to limit free streaming. URL `http://www.independent.co.uk/arts-entertainment/music/news/spotify-reportedly-under-pressure-from-music-labels-to-limit-free-streaming-10126106.html`.

The Telegraph. 2012. Drunk Twitter users unlikely to face criminal prosecution. URL `http://www.telegraph.co.uk/technology/twitter/9754007/Drunk-Twitter-users-unlikely-to-face-criminal-prosecution.html`.

Tufekci, Zeynep. 2008. Can you see me now? Audience and disclosure regulation in online social network sites. *Bulletin of Science Technology and Society* **28**(1) 20–36.

Van Zandt, Timothy. 2004. Information overload in a network of targeted communication. *The RAND Journal of Economics* **35**(3) 542–560.

Vasalou, Asimina, Alastair J. Gill, Fadhila Mazanderani, Chrysanthi Papoutsi, Adam Joinson. 2011. Privacy dictionary: A new resource for the automated content analysis of privacy. *Journal of the American Society for Information Science and Technology* **62**(11) 2095–2105.

Vickrey, William S. 1963. Pricing in urban and suburban transport. *The American Economic Review* **53**(2) 452–465.

Wang, Yang, Pedro Giovanni Leon, Kevin Scott, Xiaoxuan Chen, Alessandro Acquisti, Lorrie Faith Cranor. 2013. Privacy nudges for social media: An exploratory Facebook study. *Proceedings of the 22nd International World Wide Web Conference Committee (IW3C2)* 763–770.

Weitzman, Martin L. 1974. Prices vs. quantities. *The Review of Economic Studies* **41**(4) 477–491.

Weng, Jianshu, Ee-Peng Lim, Jing Jiang, Qi He. 2010. Twitterrank: Finding topic-sensitive influential twitterers. *Proceedings of the Third ACM International Conference on Web Search and Data (WSDM), ACM* 261–270.

Appendix. A

This appendix presents the optimal quota that maximizes the aggregate user surplus. We use the superscript $*$ to denote the associated outcome. ι is an infinitesimally small positive number.

(i) When $\epsilon \leq (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$:

(i.1) If $0 < v \leq (1-\epsilon)(1-\alpha)\lambda_L$, $\Lambda^* = \iota$, leading to $s_L^{q*} = 0$, $s_H^{q*} = 0$, and $s^{q*} = 1-\alpha$.

(i.2) If $(1-\epsilon)(1-\alpha)\lambda_L < v \leq (1-\epsilon)(1-\alpha\beta)\lambda_L$, $\Lambda^* = \iota$, leading to $s_L^{q*} = \frac{(v-\iota)/((1-\epsilon)\lambda_L)-(1-\alpha)}{\alpha(1-\beta)}$, $s_H^{q*} = 0$, and $s^{q*} = \frac{v-\iota}{(1-\epsilon)\lambda_L}$.

(i.3) If $(1-\epsilon)(1-\alpha\beta)\lambda_L < v \leq \bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H$, $\Lambda^* = \iota$, leading to $s_L^{q*} = 1$, $s_H^{q*} = 0$, and $s^{q*} = 1-\alpha\beta$.

(i.4) If $\bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H < v \leq (1-\epsilon)(2-\alpha\beta)\lambda_H - \bar{\lambda}$, $\Lambda^* = \frac{\gamma(1+\psi)[v-\bar{\lambda}+(1-\epsilon)\alpha\beta\lambda_H]}{c}$, leading to $s_L^{q*} = 1$, $s_H^{q*} = 0$, and $s^{q*} = 1-\alpha\beta$.

(i.5) If $(1-\epsilon)(2-\alpha\beta)\lambda_H - \bar{\lambda} < v \leq 2(1-\epsilon)(1-\alpha\beta)\lambda_H$, $\Lambda^* = \frac{2\gamma(1+\psi)[v-(1-\epsilon)(1-\alpha\beta)\lambda_H]}{c}$, leading to $s_L^{q*} = 1$, $s_H^{q*} = 0$, and $s^{q*} = 1-\alpha\beta$.

(i.6) If $2(1-\epsilon)(1-\alpha\beta)\lambda_H < v \leq 2(1-\epsilon)\lambda_H - \bar{\lambda}$, $\Lambda^* = \frac{\gamma(1+\psi)v}{c}$, leading to $s_L^{q*} = 1$, $s_H^{q*} = \frac{v/(2(1-\epsilon)\lambda_H)-(1-\alpha\beta)}{\alpha\beta}$, and $s^{q*} = \frac{v}{2(1-\epsilon)\lambda_H}$.

(i.7) If $v > \max\{2(1-\epsilon)(1-\alpha\beta)\lambda_H, 2(1-\epsilon)\lambda_H - \bar{\lambda}\}$, $\Lambda^* = \frac{\gamma(1+\psi)(v-\bar{\lambda})}{c}$, leading to $s_L^{q*} = 1$, $s_H^{q*} = 1$, and $s^{q*} = 1$.

(ii) When $\epsilon > (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$:

(ii.1) If $0 < v \leq (1-\epsilon)(1-\alpha)\lambda_L$, $\Lambda^* = \iota$, leading to $s_L^{q*} = 0$, $s_H^{q*} = 0$, and $s^{q*} = 1-\alpha$.

(ii.2) If $(1-\epsilon)(1-\alpha)\lambda_L < v \leq (1-\epsilon)(1-\alpha\beta)\lambda_L$, $\Lambda^* = \iota$, leading to $s_L^{q*} = \frac{(v-\iota)/((1-\epsilon)\lambda_L)-(1-\alpha)}{\alpha(1-\beta)}$, $s_H^{q*} = 0$, and $s^{q*} = \frac{v-\iota}{(1-\epsilon)\lambda_L}$.

(ii.3) If $(1-\epsilon)(1-\alpha\beta)\lambda_L < v \leq (1-\epsilon)(1-\alpha\beta)\lambda_H$, $\Lambda^* = \iota$, leading to $s_L^{q*} = 1$, $s_H^{q*} = 0$, and $s^{q*} = 1-\alpha\beta$.

(ii.4) If $(1-\epsilon)(1-\alpha\beta)\lambda_H < v \leq (1-\epsilon)\lambda_H$, $\Lambda^* = \iota$, leading to $s_L^{q*} = 1$, $s_H^{q*} = \frac{(v-\iota)/((1-\epsilon)\lambda_H)-(1-\alpha\beta)}{\alpha\beta}$, and $s^{q*} = \frac{v-\iota}{(1-\epsilon)\lambda_H}$.

(ii.5) If $(1-\epsilon)\lambda_H < v \leq \bar{\lambda}$, $\Lambda^* = \iota$, leading to $s_L^{q*} = 1$, $s_H^{q*} = 1$, and $s^{q*} = 1$.

(ii.6) If $v > \bar{\lambda}$, $\Lambda^* = \frac{\gamma(1+\psi)(v-\bar{\lambda})}{c}$, leading to $s_L^{q*} = 1$, $s_H^{q*} = 1$, and $s^{q*} = 1$.