Enhancing the privacy risk awareness of teenagers in online social networks through soft-paternalism mechanisms

J. Alemany[a], E. del Val[a,b], J. Alberola[a,c], A. García-Fornes[a]

[a]{jalemany1,edelval,jalberola,agarcia}@dsic.upv.es
Universitat Politècnica de València,
Camino de Vera s/n, Valencia (Spain)
[b]*Universidad de Zaragoza, Teruel (Spain)*
[c]*Florida Universitària,*
Rei en Jaume I, 2 Catarroja. Valencia (Spain)

Abstract

Privacy Risk in Online Social Networks (OSNs) is one of the main concerns that has increased in the last few years. Even though social network applications provide mechanisms to control risk, teenagers are not often aware of the privacy risks of disclosing information in online social networks. The privacy decision-making process is complex and users often do not have full knowledge and enough time to evaluate all potential scenarios. They do not consider the audience that will have access to disclosed information or the risk if the information continues to spread and reaches an unexpected audience. To deal with these issues, we propose two soft-paternalism mechanisms that provide information to the user about the privacy risk of publishing information on a social network. That privacy risk is based on a complex privacy metric. To evaluate the mechanisms, we performed an experiment with 42 teenagers. The proposed mechanisms were included in a social network called Pesedia. The results show that there are significant differences in teenagers' behaviors towards better privacy practices when the mechanisms are included in the network.

Key words: Behavioral bias, Online disclosure, Privacy, Soft-paternalism

1. Introduction

Teenagers constitute one of the main user groups of Online Social Networks [1]. The use of social networks is part of children's daily living routine. According to Livingstone et al. [2], 93% of 9-16-year-old users go online at least weekly (60% go online every day or almost every day). Although teenagers obtain a benefit from sharing and consuming information on OSN (i.e., instant messaging, watching videos, or playing games), they are also exposed to privacy risks (i.e., cyberbullying or experiences that make them feel uncomfortable) [3, 4]. Recent surveys have shown that users' privacy concerns regarding social networks have increased in the last few years [5, 6].

There are clear differences in behavior between teenagers and adults on social networks. The comparison carried out by Christofides et. al [7] reveals that teenagers spend significantly longer on SNS per day, and they have more contact with strangers [8] (17 percent of teens have become "friends" with people who they have never personally met, and 43 percent of teens have been contacted online by strangers). They can be easily convinced to share their personal information with the promise of a small prize or gift. Since children and teenagers tend to be trusting, naïve, curious, adventuresome, and eager for attention and affection, potential offenders and strangers have found that children and teenagers are perfect targets for criminal acts in cyberspace [9]. The combination of both factors (i.e., the number of friends and their vulnerability), makes the risk (i.e., the probability of reaching a broader audience) of a teenager's publication higher than an adult's publication. Therefore, the privacy risk of teenagers actions increases. The need of mechanisms oriented to increase privacy awareness when teenagers share information in social networks or applications becomes more relevant. Despite the importance and vulnerability of this demographic group, this subset of the community has hardly been researched in the context of privacy in social networks.

In this research, we focus on teenage users' behavior regarding online privacy. In this context, three processes are considered to be important [10]: risk

assessment (i.e., calculating risk probability and magnitude); risk evaluation (i.e., determining the acceptability of a given risk); and risk management (i.e., the process of reducing risk to an acceptable level). When users are going to publish a message on an OSN, they should evaluate the benefits and risks of performing that action. The privacy decision-making process is complex and users often do not have full knowledge of the audience that will see the publication or how other users are going to use the disclosed information. In addition, the evaluation of all the possible scenarios of a disclosure could be overwhelming for a user, especially for teenagers [11].

Several approaches have been proposed to facilitate the decision-making process of users in OSN that may affect their privacy. For instance, some social network applications offer privacy-settings controls. However, in some cases, these controls are complex for non-expert users that are unable to fully understand the implications of their own settings. In other cases, the configuration of privacy settings is considered by users to be a tedious task, so they prefer to maintain the default settings [12]. In addition, privacy controls in OSN are more focused on protecting the information related to the user profile than on protecting the privacy of the user's publications [12, 13, 14]. There are other approaches that address the problem of users' privacy with the automation of privacy settings configuration [15, 16, 17, 18]. However, these proposals usually require an initial user intervention. Other approaches try to improve user awareness about the misalignment of users' expected audience with the actual audience to reduce the negative effects of performing an action in an OSN [19, 20, 21]. Several works also propose privacy risk metrics to asses users in the management of their privacy just before performing a sharing action [22]. However, to facilitate the decision-making process of users, it is not only important to measure the privacy risk (i.e., risk assessment), but also the way the metric will be shown to users. The way the information is shown can influence the users' decision-making process (i.e., risk evaluation and management).

According to Staksrud and Livingstone [10], it is relevant to assist teenagers to cope with risk without restricting their freedom of online exploration that

society promotes for children in other contexts. In recent years, there has been growing interest in the use of mechanisms from behavioral economics to improve decision-making processes where lack of information or cognitive overload may unfavorably affect user privacy [23]. These mechanisms are known as soft paternalistic interventions (i.e., nudges). They attempt to influence decision making to improve individual well-being, without actually limiting users' ability to choose freely, thus, preserving freedom of choice [24].

In this paper, we present two soft-paternalism mechanisms to assist users (especially teenagers) to make better decisions about actions in social networks that may increase their privacy risks. The aim is to increase their privacy awareness. In this paper, privacy awareness refers to the users' knowledge about the potential audience that might see a user's publication disclosure. The proposed mechanisms "nudge" users to reconsider the disclosure actions before performing them. The proposed mechanisms use information from a Privacy Risk Score (PRS) metric that considers different levels of friendship and the potential audience that may have access to the disclosed message [22]. The first mechanism shows the profile images of users that are part of the potential audience that may have access to the message and a risk-level alert. The second mechanism shows the number of users that are part of the audience that may have access to the message and a risk-level alert. We tested the mechanisms in a four-week experiment with 42 teenagers in an online social network called Pesedia. The results obtained through the analysis of the social network logs suggest that the use of soft-paternalism mechanisms could be a suitable option to assess in the decision-making process and prevent teenagers from privacy risk publications that could have negative consequences.

The rest of the paper is structured as follows. Section 2 presents previous works that are related to privacy protection and awareness. Section 3 describes in detail the nudging mechanisms proposed. Section 4 describes the methodology followed for the experiment (i.e., their study subjects, protocols, and types of evaluations). Section 5 presents the evaluations and results derived from the teenagers' activities and interactions during the study. Section 6 presents our

discussions about how the results obtained should be interpreted and what was learned from the research. Finally, Section 7 presents conclusions and future work.

2. Related work

As the number of activities in online social networks increases, teenagers have to deal with an increasing number of privacy decisions. These decisions are made with incomplete and asymmetric information (i.e., limited knowledge about the reachability of a publication) and with bounded rationality (i.e., limited resources to evaluate all possible options and their consequences). Previous studies [25] state that the limited attentional capability of humans results in their bounded capacity to be rational.

Several educational strategies have been carried out by education centers and public administrations to leverage teenage users' awareness of privacy risks and to reduce their exposure to associated negative experiences [1, 26, 27]. There are also some studies that evaluate the impact of educational initiatives which suggest that they are successful in increasing awareness about online risks [28, 29]. However, the research community considers that awareness and confidence do not necessarily promote less risky behavior among young people [2]. This result is in the line of the number of young people that report negative online experiences despite the initiatives carried out by education institutions [2].

As an alternative to educational materials, mechanisms from the field of behavioral research have been considered to be appropriate for designing systems that nudge users towards better decisions concerning privacy [24]. Specifically, soft-paternalism interventions have been considered as a suitable method to influence teenagers' privacy behaviors without losing freedom of choice or liberty.

In the context of privacy in mobile applications, Almuhimedi et al. [30] propose the creation of an application that includes soft-paternalism mechanisms with the aim of raising the awareness of data collected by other applications.

The authors carried out an 8-day experiment where the participants installed the proposed application. The application alerts consist of messages describing the number of apps accessing one information type and the total number of accesses in a given period. The alerts triggered changes of 58% in the data access permissions of other applications. The results suggest the positive effect of the soft-paternalism on the awareness of users' data that is being used by third-parties. This work monitorizes and informs once a third-party application has already accessed to user's data. However, our proposal is oriented to the creation of preventive action messages that would avoid future regrets about the sharing action.

Other works have used soft-paternalism mechanisms to deal with privacy in instant messaging applications. Patil et al. [31] carried out an experiment with 50 participants to evaluate whether privacy preferences of the social circles influence privacy setting configuration. When the participants were configuring their preferences for six privacy-relevant settings, they also had information about the privacy choices made by the majority of their contacts. The results of the experiment show that the primary driver in establishing a certain setting value is the privacy aspect. The privacy choice of user's social circle is a secondary source of guidance to establish privacy settings. The results also show that one's personal perception of privacy is an influential characteristic. Therefore, it could be considered appropriate provide information about the user's situation regarding privacy when he is going to perform an action in order to influence in his behaviour.

Soft-paternalism mechanisms have also been applied to online social networks. Konings et al. [32] present an approach that controls the access to information published on social networks and for how long it would be available. This proposal combines a policy-based cryptographic enforcement system with social signaling. Social signaling is used to label sensitive information. The authors propose a set of privacy icons to label the information shared on a social network. When users publish a message, they can select the users that will have access to the message, how long they will have access to it, and the social

icons that recommend how the message should be treated (i.e., private, keep information internal, not print, etc.). However, users do not have information about the potential audience that might see the message. Users only have the option to express their personal preferences about the audience.

Wang et al. [33] present the results of a 6-week experiment with 28 Facebook users. In the experiment, the authors introduced three types of nudges: audience nudge (contains textual and visual information of the audience), timer nudge (introduces a visual delay of 20 seconds after a user clicked the "post" button before publishing the submitted post), and the combination of the two. The results conclude that participants that use Facebook to post personal opinions perceive the nudges as being more beneficial than those who use it to broadcast news articles or for commercial purposes. Moreover, the users that have experience in the configuration of privacy settings considered that the nudges could be more useful for people without experience in social networks. However, in the case of the audience mechanism the privacy risk that a user could have if the expected audience re-share the user's publication is not considered. This information could provide him a broader view of the potential reachability of his publication. The results of the experiment suggest that these mechanisms can be useful for people who are starting to use social networks (e.g. children and adolescents).

A similar 12-day experiment with 21 participants was carried out in [34]. The authors propose different nudging mechanisms to be integrated into Facebook. The first mechanism "audience nudge" provides images of the audience that could see the post. Similarly to the audience mechanism proposed in [33], this mechanism also does not take into account the potential audience in the case of user with permissions re-shares the publication. The second mechanism "timer nudge" includes a time delay before a user posts a message on the social network. The third mechanism "sentiment nudge" consists of an estimation of the sentiment associated to the post that the user is going to publish. The authors analyzed the data collected from the experiment (i.e., number of changes in online privacy settings, number of canceled or edited posts, post frequency,

and topic sensitivity) and the data of a questionnaire after the experiment. They found clear evidence of changes in posting behavior for some of the participants. The participants mentioned that the "audience nudge" was useful for thinking about customized groups. For the "timer nudge", the users mentioned that the mechanism provided them the opportunity to stop and think about the publication. In general, the "sentiment nudge" was perceived as being a less useful nudge than the others. The authors mention that the reasons could be associated with the sentiment algorithm that was used.

It is important to provide mechanisms that facilitate the increase of privacy awareness. The concept of privacy awareness varies depending on the research work. Some authors consider privacy awareness to be the knowledge of privacy notices and understanding of privacy controls and settings [35]. Others define privacy awareness as the perception of the elements in an environment, threats, and implications from personal information disclosure [36]. In this work, we consider a more specific concept of privacy awareness to be the knowledge of the users about the potential audience that might see a user's publication disclosure. Specifically, we propose a nudge approach similar to the one proposed by Wang et al. [34, 33] to increase users' privacy awareness. However, the work presented here differs from the previous ones in several ways (see Table 1). First, we integrate a privacy risk metric in the nudges, which considers the potential audience of a publication (i.e., if a user of the intended audience re-shares the information). The current approaches consider the audience based only on the privacy policy defined by the user, without considering the potential re-sharing actions. Second, we introduce a new quantitative nudge that shows the number of potential users that may see the publication instead of showing the users' profile images. We also evaluate whether there are differences in the influence on users' behaviors between the visual nudge or the numeric nudge. Third, we evaluate the nudges in a population of teenagers between 12 and 14 years old.

	Nudges	Nudges for privacy	Media
Almuhimedi et al. [30]	• application message alerts (number of apps accessing one information type and the total number of accesses in a given period)	✓ Considers AppOps logs shown for each app-permission	Mobile app
Patil et al. [31]	• user's social circle	✓ Considers the actions performed by the user's social circle	Instant messaging app
Konings et al. [32]	• privacy icons for social signaling	✓ Considers the user's preferences	
Wang et al. [34] (2013)	• audience (profile images of the publication audience) • time (visual delay) • audience + time	✓ Considers the privacy policy of the publication ✓ Considers the privacy policy of the publication	Social network
Wang et al. [33] (2014)	• audience (profile images of the publication audience) • time (visual delay) • sentiment	✓ Considers the privacy policy of the publication	Social network
Our work	• visual audience + text message with a degree of privacy risk • numerical audience + text message with a degree of privacy risk	✓ Considers a privacy risk metric that estimates the potential audience of a publication ✓ Considers a privacy risk metric that estimates the potential audience of a publication	Social network

Table 1: Overview of approaches related to soft-paternalism mechanisms. We considered three main features: (i) the type of nudges used; (ii) if the nudges are applied to prevent privacy risk scenarios and what information was considered to establish the privacy risk; and (iii) the environment where the nudges are applied.

3. Nudging Mechanisms

Recent research works state that social media users underestimate their audience size, guessing that their audience is just 27% of its true size [37, 38]. Users usually do not remember which users are part of their direct audience in social networks, and, therefore, it is highly complicated to determine those users

that can be reached from their direct audience. Therefore, users do not usually apply privacy tools (e.g., audience selectors or access lists) to define who has access to their information. As a consequence, users post information that may reach undesired audiences without being conscious of it. This information can even reach other communities that were not in their intended audience.

The idea of nudging was popularized by Thaler and Sunstein [39] as a form of soft-paternalism to guide individuals toward certain behaviors. A nudge can be viewed as an intervention that can modify people's behavior without forcing them. Hansen [40] stated that users are not usually aware of biases that may result in choices that have potentially adverse outcomes. Therefore, nudges can be viewed as mechanisms oriented to mitigate human biases to provide more beneficial outcomes for users. In decision-making scenarios that are involved in social networks, nudge mechanisms can be focused to provide support for users to enhance their privacy and security.

According to this, we propose informing the users about the potential audience of their publications using soft-paternalism mechanisms based on a privacy risk metric. The metric used in this work to support the nudges is the Privacy Risk Score [22].

3.1. Privacy Risk Score (PRS)

We assume that there is a social network $\mathcal{G}$ that consists of N nodes, where every node $a_i \in \{a_1, ..., a_n\}$ represents a user of the social network. Users are connected through bidirectional links that represent friendship relationships and correspond to the edges $E \subseteq N \times N$ of $\mathcal{G}$. We define the Privacy Risk Score (PRS) [22] for a user a_i that publishes a message as an indicator of the potential risk of this message to be diffused over the social network (i.e., potential visibility). The higher the PRS value, the higher the threat to user a_i's privacy.

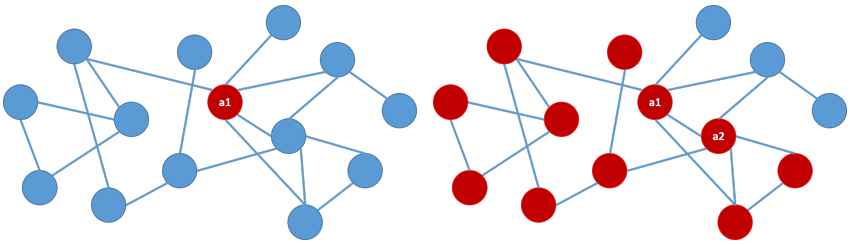
To estimate the PRS, two important factors are considered: (i) the user's position in the network. Those users located in paths where messages follow frequently, have a higher privacy risk than others; and (ii) the newness of a message. As stated in [41], the diffusion of a message in a social network is

dependent on the lifetime since this message was created. In our case, the message diffusion process of a message m is based on other models [42, 43], in which users are initially represented as deactivated nodes, since they did not received the message. Users become activated as they receive this message and the diffusion process finishes when no activations occur from time step s to $s+1$. The estimation of the PRS is described in more depth in [22].

According to this process, the privacy risk of a user by performing a message's diffusion is related to the amount of users that this user can activate. Figure 1a shows a social network in which user a_1 is publishing a new message. Blue nodes represent users that have not seen this message and can potentially see it (are deactivated nodes), while red nodes represent users that have already seen the message (are activated). The privacy risk associated with user a_1 for the diffusion of this message is high, since the probability to reach deactivated nodes (i.e., the rest of the users apart from a_1) is high too. Figure 1b shows a social network in which user a_1 is publishing a message that was forwarded by user a_2. However, this message has already been seen by a large number of users of the social network. In this case, the privacy risk associated with user a_1 for the diffusion of this message is low, since there are only 3 remaining deactivated users that can potentially be activated. Therefore, we say that the privacy risk associated with a user for a message diffusion process is high when a user publishes a new message since no other users have viewed it yet (i.e., they are deactivated). In contrast, the privacy risk is low when a user publishes a message that has already been viewed by others (i.e., they have become activated).

To represent this, we define $S = \{1, 2, \ldots, n\}$ to indicate the number of steps that a message has taken from its creation. Considering these two factors, we define a $S \times N$ reachability matrix $\gamma_{\mathbf{i}}$ associated to each user a_i to represent the number of messages that a_i has published at a certain step s and have been seen by other users. As an example, the value $\gamma_{i_{s,a_j}}$ represents the messages published by a_i in step s that were seen by a_j.

In a general view, the PRS value for a user a_i can be calculated as the per-



(a) High privacy risk of user a_1 for sharing a message non-seen yet.

(b) Low privacy risk of user a_1 for sharing a message seen by the majority.

Figure 1: Representation of user's privacy risk for different diffusion times of a message.

centage of agents of the social network that potentially see a message published by a_i at any stage (Equation 1).

$$PRS(a_i) = \frac{1}{S} \sum_{s=1}^{S} \left(\frac{\sum_{a_j \in N} \gamma_{i_{s,a_j}}}{\gamma_{i_{s,a_i}} \cdot |N|} \right) \tag{1}$$

The PRS takes a value in the interval [0..1]. If this value is close to 0 when a user is about to publish a message, it indicates that this message is expected to be seen by a small number of users. In contrast, if this value is close to 1, it indicates that the potential audience of the message is the majority of the social network. It is possible to define PRS value intervals. Each interval is associated with informative labels (i.e., none, low, medium, high) that will appear in the nudging mechanisms. The definition of the intervals depends on the domain.

This metric provides an estimation of the potential audience that could have access to a publication in terms of the users of the social network that potentially see a message published by another user. The goal of the PRS is oriented to helping users to manage their sensitive and non-sensitive information, thereby improving their experiences in the social network.

Figure 2 shows a scenario where the privacy risk score is calculated for users a_1 and a_2 in a social network. We assume for simplicity that all of the users in $\mathcal{G}$ have the privacy policy that only their direct friends can see their walls. The maximum value for parameter S cannot exceed the network diameter (i.e.,

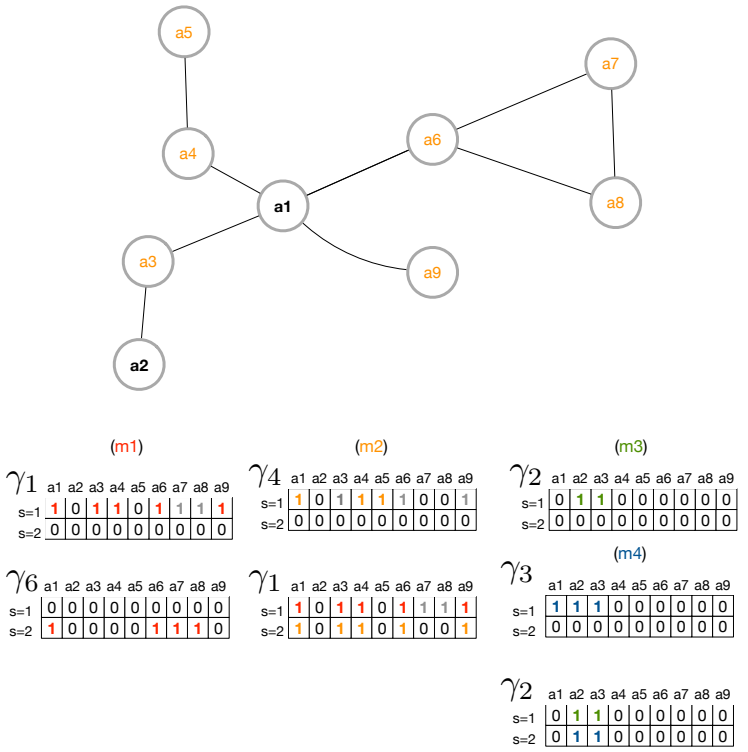


Figure 2: Example of social network activity and the PRS calculation process.

the longest of all of the shortest paths between two nodes). Therefore, for this example of PRS calculation, we use the value 2 for parameter S.

Following, we define a diffusion process of four messages (m1, m2, m3, m4):

m1) User a_1 publishes a message m_1 on its wall. Since a_1 sends this message at $s = 1$, γ_1 is updated at $s = 1$, adding a value of 1 to a_3, a_4, a_6, and a_9, which are the agents that can see the message. Then, a_6 decides to share m_1 on its wall. Users a_1, a_7, and a_8 can see m_1. The information about the users that can see m_1 is updated in γ_6. The interaction of user a_6 with m_1 occurs after user a_1 shares it (i.e., the interaction is produced in the step $s = 2$). Note that the values of γ_1 are updated at $s = 1$ because γ_1 measures the reachability of the messages when user a_1 has interacted with them. Therefore, in row $s = 1$, columns a_7 and a_8 have a grey 1.

m2) Then, user a_4 publishes m_2. This message is seen by a_1 and a_5. This information is updated in γ_4 at $s = 1$. After that, user a_1 decides to share m_2, and agents a_3, a_4, a_6, and a_9 can see m_2. Therefore, γ_1 is updated with this new information. However, in this case, the row $s = 2$ is updated since the sharing action of a_1 implies the second step of m_2.

m3) User a_2 decides to publish m_3 and only user a_3 can see it. The γ_2 matrix is updated accordingly.

m4) The message m_4 is generated by user a_3. This message is viewed by its direct neighbors a_2 and a_1, and the γ_3 matrix is updated with this information. Finally, user a_2 decides to share m_4, and only a_3 can see it. Its γ_2 its updated at $s = 2$ with this new information.

Considering these four messages, the PRS for a_1 and a_2 can be calculated as:

$$PRS(a_1) = \frac{1}{2}\left(\frac{6}{9} + \frac{4}{9}\right) = 0.6$$

$$PRS(a_2) = \frac{1}{2}\left(\frac{2}{9} + \frac{2}{9}\right) = 0.2$$

As can be observed, the PRS for a_1 is 0.6, indicating that messages published by a_1 are expected to be seen by a high number of users. In contrast, the PRS for a_2 is 0.2, indicating that messages published by this user are not expected to reach a lot of users. If we consider intervals for PRS values of size 0.25 (i.e., None [0, 0.25]; Low [0.25, 0.5]; Medium [0.5,0.75]; High [0.75,1]), the PRS(a_1) indicates that the risk is medium and the PRS(a_2) indicates that the risk is none.

3.2. Nudges

Considering the PRS, nudges are shown to users by means of two soft-paternalism mechanisms in order to propose more beneficial choices regarding the privacy of this publication. These mechanisms are *Picture Nudge*, which is based on profile images of the potential audience, and *Number Nudge*, which provides numerical information about the potential audience of a publication.

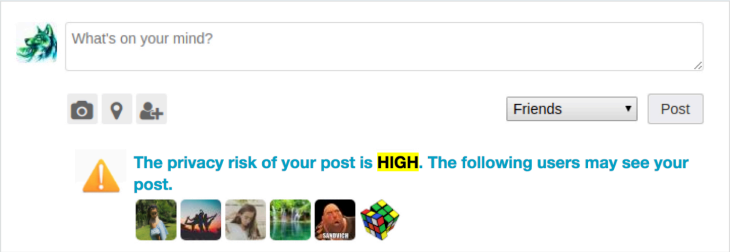


Figure 3: Picture Nudge. A notice indicates the privacy risk estimation associated with the action that the user is going to perform. The risk is categorized as high. The nudge shows the profile pictures of part of the audience that potentially could see the publication.

These nudge mechanisms try to increase the users' awareness about the reachability of their publications. Then, users can reconsider the privacy policy of a publication more carefully or can even decide not to publish that information.

Picture Nudge. The Picture Nudge is a mechanism that is triggered when a user is about to submit a publication (Figure 3). This mechanism consists of showing profile images of some users that are part of the audience that will have access to this publication. Users to be displayed are selected based on the PRS values of the post's audience and the probability of reaching new users. The probability increases if a user can be reached in more than one way. The selection of profile images to be displayed in the nudge prioritizes users outside of the intended audience. Although only six users are explicitly shown, the size of the audience can be very large. In addition, a warning is also shown according to the privacy risk estimation of this publication (high, medium, low, or none).

Unlike other proposals that provide mechanisms to detect and remove risky friends [44], the aim of the picture nudge proposed is to increase awareness about the potential audience that might see a user's publication. This does not imply that the users that appear in the images provided by the nudge are "risky" users. These users are part of the potential audience that may see the publication.

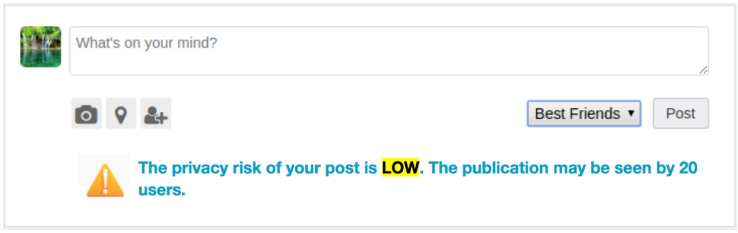


Figure 4: Number Nudge. A notice indicates the privacy risk estimation associated with the action that the user is going to perform. The risk is categorized as low. The nudge shows the number of users that eventually could see the publication.

Number Nudge. The Number Nudge is also triggered when a user is about to submit a publication (Figure 4). This mechanism consists of displaying the number of users that may have access to this publication. Similarly to the previous nudge, a warning related to the privacy risk estimation of this publication is also shown.

4. Experiment

We propose two research questions and two hypotheses to test the effects of the proposed nudging mechanisms in users' behaviors regarding privacy. We focus on the privacy aspect related to the content publishing, specifically, the selected audiences. First, we should consider that a new social network (or app) has a "learning curve" (i.e., a period of learning and discovery, until users start to use it regularly). This may influence the participants' behavior regarding privacy during the experiment. Therefore, we investigate the following research question:

RQ1 *How does the private privacy policy rate differ between the learning/discovery period and later when users publish content regularly?*

In other words, the private privacy policy for published content includes all of the private audiences ("only me", collections[1], and "friends").

[1]Collections are subsets of "friends" that are specialized and customized by users (e.g.,

Second, regarding the designed nudge mechanisms, we want to know if the nudge before publishing content and the information provided in it (about the potential audience) produce an effect towards better privacy practices. Therefore, two hypotheses are proposed:

H1 *The private privacy policy rate changes when teenage users publish content using the Picture Nudge mechanism.*

H2 *The private privacy policy rate changes when teenage users publish content using the Number Nudge mechanism.*

Finally, we investigate the differences between the effects of the designed nudge mechanisms in order to analyze which mechanism has a more powerful effect on users' behavior. Therefore, we investigate the following research question:

RQ2 *How does the private privacy policy rate differ between the Picture Nudge and the Number Nudge when teenage users publish content?*

To evaluate these effects, we performed an experiment in the context of the 2017 Summer School organized by the Universitat Politècnica de València. We focused the experiment on teenagers aged between 12 and 14 years old because they are starting with the use of social networking sites, and, at the same time, they are among the heaviest users of social networking [45]. Moreover, this particular group is developmentally vulnerable to privacy risks such as depression, sexting, and cyberbullying [46, 47, 48, 49]. Therefore, the effect of nudge mechanisms can be highly beneficial to them since these users may still not be aware of all of the consequences of their actions in social applications regarding their privacy. In the following sections, we describe the social network platform Pesedia where the experiment was performed and the methodology used for measuring the effect of the proposed nudges on real users.

best friends, family, acquaintances, etc.)

4.1. Platform

PESEDIA is an online social network for educational and research purposes that includes: (i) the design and development of new metrics to analyze and quantify privacy risks [22]; (ii) the application of methods to change users' behavior regarding their privacy concerns; (iii) the implementation of new features to improve the management of users' content; (iv) and the evaluation and testing of new proposals with real users.

The underlying implementation of PESEDIA uses Elgg [50], which is an open source engine that is used to build social environments. The environment provided by this engine is similar to other social networks (e.g. Facebook). Figure 5 shows the architecture of PESEDIA. The PESEDIA architecture has two main components: the *Platform Layer* and the *User Layer*. The *Platform Layer* is the core of the architecture. This layer contains the Social Network Services, which provides the main functionality of the social network, and the Storage System, which provides persistent storage of all of the information generated in the social network. Among other modules, the Social Network Services include the Privacy Risk Module, which is responsible for estimating the risk a user has when performing an action in the social network, and the Nudging Mechanism Module, which is responsible for providing a suitable visualization of the privacy risk associated to a user's action in order to influence his/her behavior. The *User Layer* is in charge of managing information associated to each user. This information is divided in three categories: contacts (grouped or non-grouped); information (e.g., profile items, publications, etc.); and settings, which are mainly focused on privacy settings, such as privacy policies and privacy thresholds.

4.2. Setup

The experiment was carried out on the PESEDIA social network. Nudging Mechanisms and Privacy Risk Module plugins were included in PESEDIA. We activated a log system to record all of the users' actions in order to analyze them after the experiment. Moreover, we also included a registry controller (by

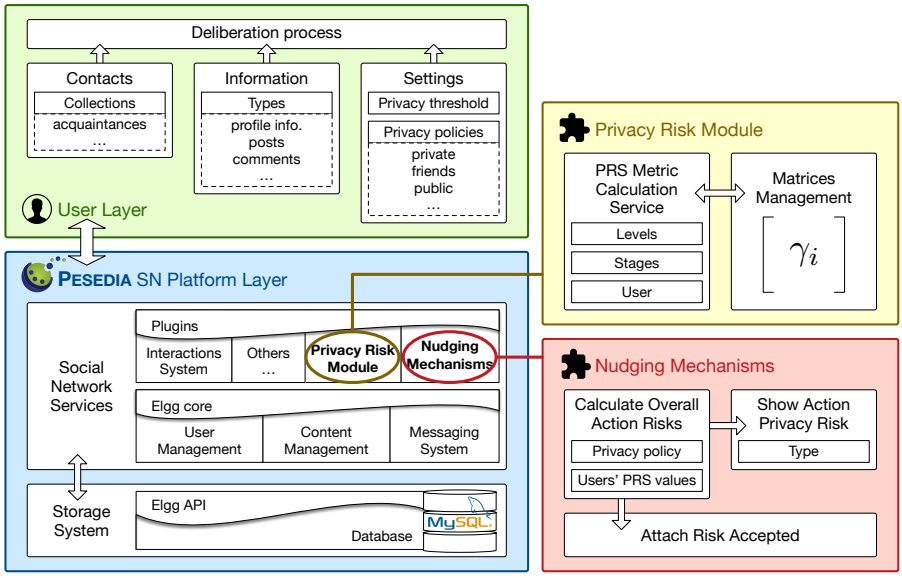


Figure 5: Block diagram that represents the architecture of PESEDIA SN. Also represented are the relevant plugins for this work: the Privacy Risk Module, and the Nudging Mechanisms.

a secret token) to avoid undesired registrations that could affect the security of the participants and the experiment.

The experiment period was 21 days. A total of 84 teenagers participated in it. During the period of the experiment, the participants had access to the PESEDIA social network to share their experiences and feelings about the Summer School. We organized three on-site sessions of 90 minutes in equipped labs at the university to use as control points of the experiment. These three on-site sessions were distributed at three points in time: session 1, at the beginning of the 21-day period; session 2, in the middle, and session 3, at the end. The aim of these sessions was to clarify any doubts that might arise among the participants about the social network functionality and new features introduced. In the first session, we introduced PESEDIA to the participants and they signed up on the social network. In the second session, the nudges were activated and introduced to the participants. During this session, we described how the Picture Nudge and Number Nudge mechanisms worked to all the participants.

We provided details about the information that each nudge provided and how they worked. We also explained both nudges through a set of examples to clarify any doubt about their performance. In the case of the Picture Nudge, we clarified that the users that appeared in the images provided are not "risky" users, they are part of the potential audience that may see the publication. The participants should evaluate, based on the potential audience shown by the nudges, whether their publication may reach more users than the initial expected audience. In the third (and last) session, the participants answered the questionnaire about the experience.

In order to test the research questions and hypotheses proposed in this work, we split the participants into three groups and considered two stages in the experiment (see Figure 6). The splitting of all the participants into the three groups was done before the second session (i.e., after completing stage 1), and based on the private privacy policy rate of users' posted content on Pesedia to have the groups balanced. The groups are explained below:

- Group G_1 did not have any nudges activated during the entire experiment. This group was created to evaluate whether the "learning curve" influences the users' privacy behaviors (RQ1).
- Group G_2 did not have any nudges activated during stage 1, but the Picture Nudge mechanism was activated during stage 2. Group $G2$ was created to evaluate whether the Picture Nudge influences users' privacy behavior (H1).
- Group G_3 did not have any nudges activated during stage 1, but the Number Nudge mechanism was activated during stage 2. Group $G3$ was created to evaluate whether the Number Nudge influences users' privacy behavior (H2).

Moreover, in order to reinforce the data obtained from social network activity, the teenagers completed a survey questionnaire about the experiment. This questionnaire was finally completed by 31 participants.

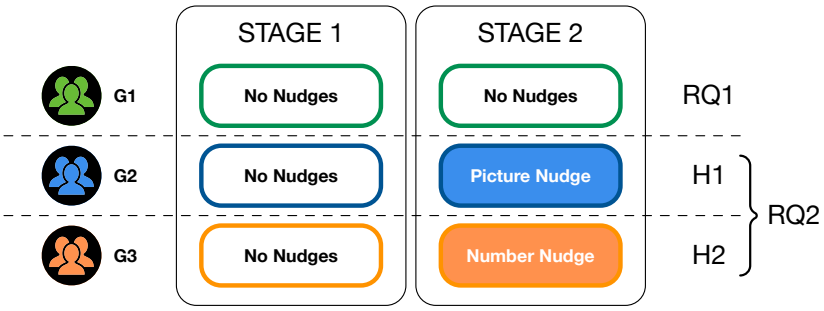


Figure 6: Structure of the experiment. Two stages and three groups of participants ($G1$, $G2$, and $G3$) were considered. In $G1$, the participants did not have any nudges activated during any stage. In $G2$, the participants did not have any nudges activated during stage 1, but the Picture Nudge was activated during stage 2. In $G3$, the participants did not have any nudges activated during stage 1, but the Number Nudge was activated during stage 2.

5. Results

In this section, we show the results obtained from the experiment. First, we introduce the participants' demographics and their initial attitude toward privacy as well as data related to posting behaviors. All of the information about participants was collected from the Pesedia platform through their profiles, activity, and settings. Second, we analyze the participants' activity during stage 1 (where none of the groups had the nudging mechanisms activated) and during stage 2 (where G_1 and G_2 had the nudging mechanisms activated) in order to quantify the impact of the nudges on the participants. We applied statistical significance tests to answer the research questions and to validate the hypotheses about the nudge effects on participants' behaviors. Finally, we present the participants' perception of the benefits and drawbacks of the nudges based on the survey results.

5.1. Demographics and activity

In this subsection, we provide an accurate description of the participants of the experiment. We show the participants' descriptive data and their performance in Pesedia. In addition, we focus on the privacy decisions made by the users during the experiment.

From the initial 84 participants that attended the experiment, we removed the participants who did not participate in both stages as well as participants who did not publish anything since either they did not attend or did not log into Pesedia (11 participants were removed). Also, there were participants that assisted to both sessions but they did not publish in both sessions any content with its corresponding privacy policy in the social network (e.g., they only performed "like" actions or comments). These users were also excluded from the experiment (23 participants were removed). In addition, we also performed a cleaning process of the data. This process consisted on removing those users who were extreme outliers from the data (8 participants were removed). "Outliers" are those points which stand out for not following a pattern which is generally visible in the data. To detect data outliers, we plotted the data points about users' activity. Figure 7 shows a boxplot representation of the distributions of the number of publications of participants in each stage. The activity of those users who lay far outside the general distribution (i.e., participants who fall more than 1.5 times the interquartile range above the third quartile) were analyzed in detail to detect if there was an anomalous behavior. In the context of the experiment, we considered users with an anomalous behavior those users whose activity was to repeat/create a message with nonsense content (i.e., random sequence of characters or empty messages) a disproportionate number of times. Once we had cleaned the data, the total number of participants included in the analysis was 42.

The following analysis is based on the behavior of the 42 participants. Table 2 includes information about the participants' age and gender and their behavior (previous to the experiment) on social networks, which is centered on the nature of the relationships and how active the participants are. That information was collected with an initial questionnaire during the first session to check the specific characteristics of teenagers that differentiates them from adults. The average age of the participants was 13.35 years old and the gender was balanced. The majority of participants had used the social network sites, and the proportion of unknown friends was high (see the acceptance threshold

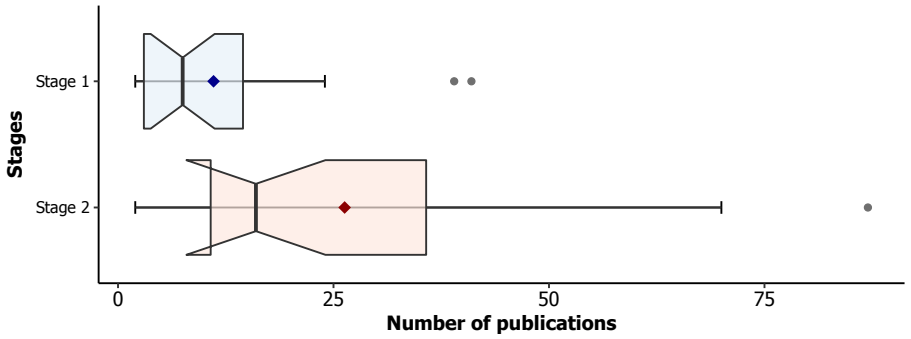


Figure 7: Distribution of the number of publications by the participants in PESEDIA during the experiment by stages.

of friendship requests and the values of real friends).

Figure 8 shows quantitative information related to their activity by gender, including log-in actions, friendship relations, and interactions considering different types (posts, likes, comments, shares, and private messages). These data are the result of the 21 days of the experiment. During that period, the participants did 317 log-in actions, established 220 relationships, and created 1976 pieces of content. In general, the most frequent activities were posts (641), likes (630), and direct messages (313). Taking into account the experiment duration, they carried out an average of 15 log-in actions per day, and 2.25 interactions per day and participant. Moreover, they performed a mean of 10.49 friendship relations. With regard to gender differences in activity, we highlight that the female participants were slightly more active creating content, especially with textual posts, share actions, and direct messages. In contrast, the male participants were more passive and performed more log-in actions.

With regard to the participants' attitudes towards privacy, we analyzed: (i) the participants' privacy policies assigned to social network dimensions such as profile, settings, and posts; and (ii) the participants' privacy concern through privacy setting changes, post updates, and collection creations. Collections are customized lists made by users (e.g., best friends, family, etc.). Figure 9 displays the distribution of the participants' privacy policy decisions grouped by dimen-

Demographic info.					Friendship info.				
Variables				Number (%)	Variables				Number (%)
Age	G_1	G_2	G_3	Total	Friends	G_1	G_2	G_3	Total
12	2	2	1	5 (11.90%)	$0-20$	2	3	2	7 (16.67%)
13	6	6	5	17 (40.48%)	$20-80$	4	3	3	10 (23.81%)
14	6	8	6	22 (47.62%)	$80-150$	3	4	1	8 (19.05%)
					>150	5	6	6	17 (40.47%)
Gender	G_1	G_2	G_3	Total	Real Friends	G_1	G_2	G_3	Total
Male	6	8	5	19 (47.62%)	$90-100\%$	3	5	1	9 (21.43%)
Female	7	8	7	22 (52.38%)	$60-70\%$	9	7	7	23 (54.76%)
					$30-40\%$	2	3	3	8 (19.05%)
					$10-20\%$	0	1	1	2 (4.76%)
Users of SNS	G_1	G_2	G_3	Total	Acceptance threshold	G_1	G_2	G_3	Total
Yes	13	15	11	39 (92.86%)	All	1	3	0	4 (9.52%)
No	1	1	1	3 (7.14%)	Some unknown people	4	4	4	12 (28.58%)
					Friends & Acquaintances	8	6	7	21 (50.00%)
					Close Friends	1	3	1	5 (11.90%)

Activity info.

Variable	4-point likert scale* - Number (%)															
	4				3				2				1			
Activity rate	G_1	G_2	G_3	Total	G_1	G_2	G_3	Total	G_1	G_2	G_3	Total	G_1	G_2	G_3	Total
Using SNS	4	5	5	14 (33.33%)	6	5	3	14 (33.33%)	3	5	3	11 (26.19%)	1	1	1	3 (7.15%)
Text posting	1	1	1	3 (7.15%)	0	2	1	3 (7.15%)	11	10	8	29 (69.05%)	2	3	2	7 (16.67%)
Photo posting	0	1	0	1 (2.38%)	0	1	0	1 (2.38%)	12	11	9	32 (76.19%)	2	3	3	8 (19.05%)
Video posting	0	1	0	1 (2.38%)	0	0	0	0 (0.00%)	3	2	6	11 (26.19%)	11	13	6	30 (71.43%)
Share	1	1	1	3 (7.15%)	3	3	2	8 (19.05%)	7	8	6	22 (52.38%)	3	3	3	9 (21.43%)
Comment	2	3	2	7 (16.67%)	4	1	5	10 (23.81%)	4	8	3	15 (35.71%)	4	4	2	10 (23.81%)
Like	7	5	7	21 (50.00%)	4	7	3	12 (28.57%)	2	3	1	6 (14.29%)	1	1	1	3 (7.15%)

Table 2: Participants' information organized by demographic, friendship, and activity categories. **Likert scale: 4 = extremely frequent, 1 = not frequent at all*

sions. The dimensions considered are: Profile, Settings, and Activity. In Profile, there are seven elements that contain the participants' profile information: age, gender, description, phone, location, school, and interests. In Settings, there are five general privacy setting options: default privacy option, tag visibility, friend list visibility, who can post on your wall, and the privacy policy for posts written on your wall. In Activity, we collected the privacy policy of all the posts

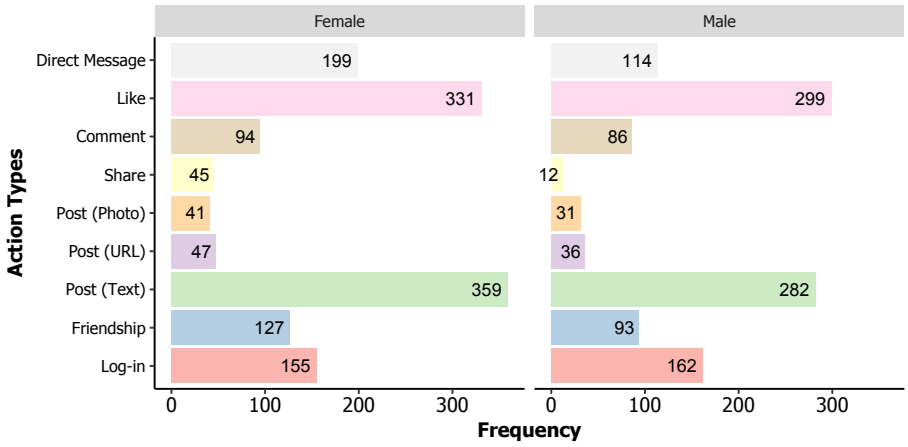


Figure 8: Information of participants' activity by gender in Pesedia.

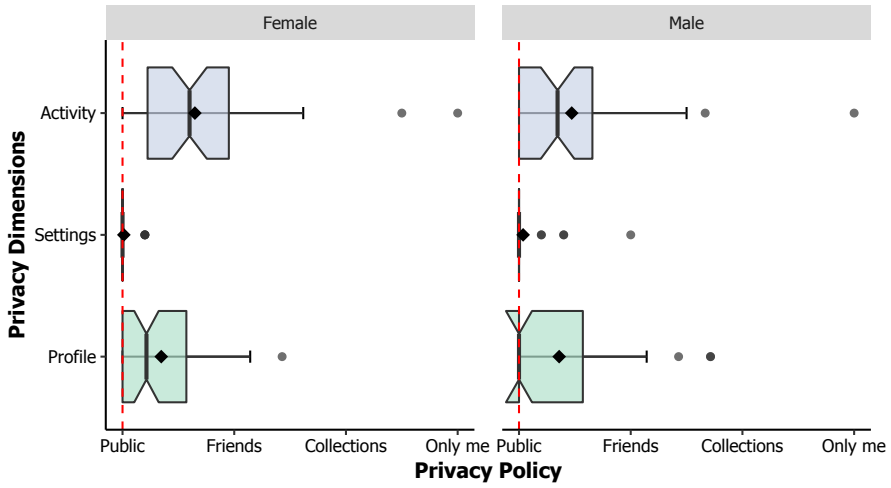


Figure 9: Distribution of privacy policies (represented as numbers: *Only me*, 0; Collections, 1; *Friends*, 2; and Public, 3) that were used by the participants in the different dimensions and disaggregated by gender.

published. The different privacy policies were: *Only me*, Collections, *Friends*, and Public; scored from 0 to 3, respectively. In Figure 9, the top red dashed line represents the mean privacy policy set as default in Pesedia.

The privacy policy defined by default in Pesedia for all of the participants

was public to be completely permissive. In the case of the Profile dimension, the expected behavior was that the participants would limit their privacy policies, and this occurred. However, as shown in Figure 9, this behavior was weaker than the expected, especially for profile items that contain sensitive information. In the case of the Settings dimension, the participants showed very little concern about it. Only 11.72% of the participants changed their privacy settings. However, in the case of the Activity dimension, the privacy policies that participants used for their posts were more restrictive than in the other dimensions. Since we are considering all of the posts published during the study, these results may be a consequence of the nudging mechanisms. Another point to highlight is the differences between gender participants; the female participants, on average, chose more restrictive policies than the male participants (for Activity and Profile dimensions). In general, we have observed that, although the participants modified their privacy options, they maintained permissive policies except for postings. In the following sections, we analyze these behaviors in more detail and how the nudging mechanisms influenced them.

5.2. Participants' posting behavior

In this section, we analyze the behaviors of the participants when they publish content in Pesedia by stages and groups. We also assess the accepted risk in privacy (i.e., the levels of privacy risk accepted for content published) by the participants who had the nudges enabled.

To understand the privacy behavior of the participants during posting activities, we extracted the participants' privacy policy for each post from the social network Pesedia. In Table 3, we show the privacy policies used for all of the participants. The participants are split into groups and stages to be able to compare behaviors with and without nudge mechanisms. We analyzed the total number of posts published and which percentage of these follow a specific privacy policy (i.e., *Only me*, Collections, *Friends*, or Public). Thus, we were be able to detect behavioral privacy changes between stage 1 and 2 and measure the effect of nudges.

Privacy Policies	Stage	G_1 Control Group	G_2 Picture Nudge	G_3 Number Nudge	All
Only me	S1	3.17% (4)	8.43% (7)	6.58% (5)	5.61% (16)
	S2	9.09% (14)	11.87% (33)	9.92% (13)	10.66% (60)
Collections	S1	0.00% (0)	0.00% (0)	0.00% (0)	0.00% (0)
	S2	0.00% (0)	3.24% (9)	4.58% (6)	2.66% (15)
Friends	S1	46.03% (58)	27.71% (23)	30.26% (23)	36.49% (104)
	S2	17.53% (27)	35.61% (99)	44.28% (58)	32.68% (184)
Public	S1	50.80% (64)	63.86% (53)	63.16% (48)	57.90% (165)
	S2	73.38% (113)	49.28% (137)	41.22% (54)	54.00% (304)
Total posts	S1	126	83	76	285
	S2	154	278	131	563

Table 3: Participants' posting behavior for the privacy aspect split into groups and stages. S1 and S2 denote stage 1 and stage 2, respectively.

In stage 1, the participants published and shared a total number of 285 posts on the Pesedia social network. Considering all the participants (42) and the duration of this stage (10 days), one out of every two participants published or shared a post per day. This participation was low, but it can be considered normal because the participants were new to Pesedia and they had to explore all of the functionality and services that our social network offers. From the privacy point of view, in general, the participants were not concerned about the privacy of postings. The majority of messages where published with public privacy policy (57.90%), followed by *friends* privacy policy (36.49%), then *only me* privacy policy (5.61%), and finally, no usage of collections by participants was done (0.0%). It is important to remember that the default privacy policy was public in the social network. Analyzing this information by groups, it can be observed that posts done by participants within G_2 and G_3 were less restrictive about privacy than other groups. For G_1, they used the friend policy more times than the other groups, with usage being close to that of public policy. In conclusion, during this stage of the study and with public policy as default, the

participants were able to change the privacy of their posts to adapt it to their needs. However, the majority of the participants maintained the public policy.

In stage 2, the participants published and shared a total number of 563 posts. The participants' activity increased considerably as a consequence of popularity, with a daily activity of about 51.18 posts per day (563 posts divided by 11 days), and more than 2 posts per user and day (51.18 posts divided by 42 participants). From the privacy point of view, if we analyze the average behavior of all of the participants and we compare it with the behavior in stage 1, it can be observed that privacy behaviors change. The use of more restrictive privacy policies such as *Only me* and Collections increased, while *Friends* and Public policy usage was reduced. When focusing on the behavior of each group, we found important differences between participants with and without nudge mechanisms. The participants in the G_2 and G_3 groups evolved their behaviors into more protective privacy policies, while G_1 did the opposite and evolved to more relaxed privacy policies. G_2 had a conservative privacy behavior since they shared half of the posts as public and the other half with private circles such as *Friends*, Collections, or *Only me*. When observing their previous activity in stage 1, there is a progression towards more secure privacy habits for social network activity. G_3 had the most conservative privacy behavior with values close to 60% of posts shared with private circles, while the rest of posts were shared as public. For both nudges, the use of the Collections policy for sharing posts increased, but it was still too low. The privacy behaviors of G_1 were less restrictive since they shared the majority (73.38%) of their posts as public. When considering the participants' behavior in stages 1 and 2 as the reference behavior, the nudging mechanisms seem to have a positive effect on the participants' privacy.

With regard to the posts published by nudged participants, Table 4 shows the proportion and quantity of posts labeled with different risk levels (calculated with the PRS metric and showed by the nudge mechanisms) that were accepted by participants when publishing posts on Pesedia. Quantities were shown as complementary information to the participants' acceptance of risk

Risk level	G_2 Picture Nudge	G_3 Number Nudge	All
NONE	1.90% (2)	0.00% (0)	2
LOW	27.62% (29)	42.59% (23)	59
MEDIUM	5.71% (6)	5.56% (3)	9
HIGH	64.76% (68)	51.85% (28)	96
Total	105	54	159

Table 4: The risk level of the posting action that participants took when nudges were activated.

since, as we mentioned in previous sections, the nudges were shown to them with a probability, thus avoiding upsetting the participants. The privacy risk accepted by participants was slightly higher in G_2 than G_3. That information is coherent with the data shown previously (Table 3), where the participants chose more protective privacy policies. Furthermore, the HIGH values of privacy risk (64.76% for Picture Nudge; 51.85% for Number Nudge) are greater than the public policies chosen. This reflects the risks of using friends policies for some users since these are not still enough to protect sensitive information.

5.3. Research questions and hypothesis testing

In this section, we test the research questions and the hypotheses proposed in this work in relation to the effects of nudges on users' privacy behavior. We use data collected from participants' posting activity to test whether there is a significant difference between the privacy behavior of participants between stages for the different conditions (G_1, G_2, and G_3 groups). In this way, we are able to measure the effect of nudges on participants' behavior.

Given the filtering of participants done by the conditions required to test the research questions and hypotheses (see subsection 5.1), we ran a samples equivalence test over the private privacy policy rate of users of the different groups at stage 1 to ensure that there were no existing differences between the samples. Kruskal-Wallis test of statistical significance to compare the mean of

the three groups (over the 42 participants) revealed no significant differences were founded in the private privacy policies rate between groups in stage 1 (p-value $> .05$). The samples equivalence test provides more confidence that these samples were equivalent on related the participants' privacy behavior in stage 1.

In order to ensure whether or not there is a significant difference in privacy behaviors between groups during stages, some research questions and hypotheses for testing. We collected data from the privacy policies of the participants' publications during stage 1 and 2 (see Table 3), and we normalized this data by the number of publications for each participant. Due to the continuous nature of the variable and the number of samples (less than 30 per group), we used the paired-sample t-test ($\alpha = .05$). For this test, we calculated the mid p-value since its Type I error rate is closer to the nominal level. In statistical hypothesis testing, a Type I error is the rejection of a true null hypothesis. Thus, we are able to reject the null hypothesis (H_0) to accept the alternative (H_1). We also measured the power test, which indicates the probability that the test correctly rejects the null hypothesis. And, thus, we obtain the Type II error, also referred to as the false negative rate (β) since the power is equal to $1 - \beta$. Therefore, we are able to accept the null hypothesis (H_0). Moreover, we measure the effect size to determine the magnitude of the phenomenon. To do this, we carried out a one-way MANOVA test. The sizes of effect can be classified as falling between small ($> .01$), medium ($> .06$) and large ($> .14$) [51]. Table 5 contains the results of the hypothesis testing methods carried out.

To answer the research question RQ1 about how the private privacy policy rate differs between the learning/discovery period and later when users publish content regularly (in the G_1 group), we tested mean differences between the two samples. In particular, we ran a paired-sample t-test ($\alpha = .05$) and the results ($t = -.348$, p-value$= .734$, partial $\eta^2 = .004$) revealed no significant differences between the samples. We also measured the power ($1 - \beta = .062$) and the effect size ($> .01$) using the one-way MANOVA test, which results also suggest no significant differences. Therefore, the results revealed that no significant

		t-Test					ANOVA			
		Mean	Std. D	t	df	p	F	p	partial η^2	$1-\beta$
RQ1	S1	.409	.395	$-.348$	13	.734	.111	.742	.004	.062
	S2	.457	.366							
H1	S1	.249	.350	-3.813	15	.002*	6.260	.002	.173**	.678
	S2	.557	.344							
H2	S1	.295	.385	-2.412	11	.035*	4.301	.044	.164**	.509
	S2	.613	.367							
RQ2	G_2	.557	.344	$-.416$	24	.681	.175	.679	.007	.069
	G_3	.613	.367							

Table 5: Tests for the differences in privacy behavior between nudged and non-nudged participants. *$p < .05$ **partial $\eta^2 > .01$ = small, $> .06$ = medium, $> .14$ = large effect [51]

differences were found in the privacy policies used by the participants in the G_1 group during stage 1 and 2.

H1 predicted that the Picture Nudge mechanism produces an effect on the participants' privacy behaviors (of the G_2 group), specifically in the private privacy policy rate of posting action. To address H1, we ran a paired-sample t-test ($\alpha = .05$) and the results ($t = -3.813$, p-value = .002, partial $\eta^2 = .173$) rejected the null hypothesis. Therefore, significant differences were found in the privacy policies used by participants in the G_2 group during stage 1 and 2, and also the effect size was large ($> .16$). Thus, H1 was supported.

H2 predicted that the Number Nudge mechanism produces an effect on the participants' privacy behaviors (of the G_3 group), specifically in the private privacy policy rate of posting action. To address H2, we ran a paired-sample t-test ($\alpha = .05$) and the results of the test ($t = -2.412$, p-value = .035, partial $\eta^2 = .167$) rejected the null hypothesis. Therefore, significant differences were found in the privacy policies used by participants in the G_3 group during stage 1 and 2, and also the effect size was large ($> .16$). Thus, H2 was supported.

To answer the research question RQ2 about how the private privacy policy

rate differs between the Picture Nudge (G_2) and the Number Nudge (G_3) when teenage users publish content (in stage 2), we tested mean differences between the two samples. In particular, we ran an independent-sample t-test ($\alpha = .05$) and the results ($t = -.416$, p-value = .681, partial $\eta^2 = .007$) revealed no significant differences between the samples. We also measured the power ($1-\beta = .069$) and the effect size ($> .01$) using the one-way MANOVA test, which results also suggest no significant differences. Therefore, the results revealed that no significant differences were found in the privacy policies used by the participants with the Picture Nudge mechanism enabled (G_2) and the Number Nudge mechanism enabled (G_3) in stage 2.

5.4. Participants' perception about nudges

We asked the participants directly about the privacy nudges using a survey embedded in PESEDIA. The results extracted from the survey represent the perceptions of the 31 participants who finally completed the survey. Of these participants, 11 participants were nudged with the Picture Nudge mechanism; 9 participants were nudged with the Number Nudge mechanism; and 11 participants were not nudged. The nudged participants were asked about the perceived benefits and drawbacks of the privacy nudges that they experienced. The non-nudged participants were asked about their desire to have tools (ours or similar ones) in social networks to inform them about privacy risks in order to improve their privacy awareness. Specifically, the following five questions were asked:

- Q1: Did you consider the nudges useful for preserving your privacy on the posting action?
- Q2: Did you consider the nudges irritating?
- Q3: Did you use the nudges for setting/fitting the audiences?
- Q4: Would you have liked to have a tool that informs you about privacy risks in order to improve your privacy (e.g., **showing the picture** of potential users that will have access to your publication)?

		Picture Nudge	Number Nudge	Non-Nudge	Total
# participants		11	9	11	31
Q1: Did you consider the nudges useful for preserving your privacy on the posting action?	Y	8	7		15
	N	3	2		5
Q2: Did you consider the nudges irritating?	Y	4	3		7
	N	7	6		13
Q3: Did you use the nudges for setting/fitting the audiences?	Y	8	6		14
	N	3	3		6
Q4: Would you have liked to have a tool that informs you about privacy risks in order to improve your privacy (e.g., **showing the picture** of potential users that will have access to your publication)?	Y		5	8	13
	N		4	3	7
Q5: Would you have liked to have a tool that informs you about privacy risks in order to improve your privacy (e.g., **showing the number** of potential users that will have access to your publication)?	Y	6		7	13
	N	5		4	9

Table 6: Opinion from a subset of participants about the privacy nudges.

- Q5: Would you have liked to have a tool that informs you about privacy risks in order to improve your privacy (e.g., **showing the number** of potential users that will have access to your publication)?

Table 6 shows the results of the participants' opinions about privacy nudges. The results are organized by the nudging mechanisms that the participants had during the study. The rows in the table represent the number of participants that responded (Yes or No) to a specific question. The empty values of the table are due to the fact that those participants were not asked the question (i.e., it made no sense to ask non-nudged participants about the inconveniences of the nudge). Questions Q1, Q2, and Q3, which targeted the nudged participants, evaluate whether the participants consider the nudges to be useful. Whereas questions Q4 and Q5, which targeted the non-nudged participants, evaluate whether the participants would like to have tools inform them about privacy risks in order to improve their privacy.

According to the participants' responses, both nudges had a good level of acceptance. For question Q1, three out of four participants considered the nudges to be useful for preserving privacy. For question Q2, about 65% of the participants did not consider the nudges to be irritating. These responses make sense if we consider question Q2 as being the opposite of question Q1. Nevertheless, the percentage is slightly lower than for question Q1, this may be because some users considered the nudge, though useful, should have been more appealing or less intrusive. For question Q3, almost three out of four participants considered the nudges to be helpful for setting the audiences. For the remaining questions (Q4 and Q5), we observed that non-nudged participants positively accepted the need fot tools to improve their privacy on social networks. Overall, the participants were satisfied with the nudging mechanisms that contain the PRS metric to improve their privacy awareness on social networks.

6. Discussion

This paper reports the results of a 21-day field experiment about the use of two types of nudging mechanisms to influence teenagers' posting privacy behavior in the social network platform Pesedia. Nudge mechanisms proposed in this paper did not limit participants' ability to share information in the social network. Instead, they encouraged the participants to reflect on their potential audience that may have access to the information. In general, previous soft-paternalism approaches not only in the context of social networks state that these mechanisms make users reflect and become more aware of their decisions, avoiding risky behaviors [30, 31, 33].

Initially, we thought that the "learning curve" of a new social network platform such as Pesedia would influence the users' privacy behaviors. However, after the analysis of the behavior of users without mechanisms during the period of the experiment, we found that there was not a significant difference in their posting privacy behavior between the initial days of the experiment and the last days.

There is significant evidence that users' privacy behavior for posting actions changed when the nudging mechanisms were activated. Independently of the mechanism used (i.e., picture or number nudge), when the nudging mechanisms were activated, the number of messages published with a private policy (i.e., *only me*, collections, or *friends*) was higher than the number of messages with a public policy. Therefore, this change could be driven by the nudges. Although users seem to publish with a more restrictive privacy policy, we noticed that most of them used friends or private policies without considering collections (i.e., a personalized subset of friends). This could be because the use of this policy in PESEDIA requires the manual creation of the collection or because it is a concept that is not present in the social network platforms that they are used to, and, therefore, they do not initially consider it as a possible option. Previous studies already showed the importance of nudges for increasing users' awareness about privacy and, thus, modify their behaviors. In this paper, we focused our experiments on teenagers, who are usually less concerned about privacy risk [46]. Although the effect of nudging mechanisms was appreciated, it is expected that more visible behavioral changes can appear if the experiment was extended in time [1].

Previous works that proposed the use of different types of nudge mechanisms do not pay attention to the differences between them on users behavior [34]. In this experimental study, we analyzed whether there is a significant difference between the effects on the privacy posting behavior of teenagers that had the Picture Nudge or the Number Nudge activated. The results revealed that there are no significant differences between mechanisms. This could be because the teenagers were focused mainly on the highlighted text about the risk level than on other details such as the profile pictures of users that may see the publication or the number of users that may see the publication. In the literature, we cannot find studies that sharply measure the effect of some type of nudge to be more beneficial in terms of changing the posting behavior. However, some authors such as [52] and [53] state that the design of nudges that are more tailored to users would cause these nudges to be more effective. This would require

aspects such as not receiving alerts about information that is already known or designing personalized nudges according to what is more effective for each specific user. This can be viewed as a limitation of our proposal that can be explored in future works.

With regard to the perception of users about the nudges, the majority of teenagers considered nudges to be useful mechanisms to preserve their privacy in posting. This follows the results obtained by Wang et. al. [33] where the users that were involved in a similar experiment with nudges in social networks mentioned that nudges could be more useful for people without experience in social networks (i.e., teenagers). Although the majority of the participants perceived nudges as beneficial, some of them considered them as irritating, and this is considered as a disadvantage towards the effective implementation of privacy nudges [54]. Wang et. al. [33] suggested that this behaviour can be associated to the profile of the publications (personal or not), but there is not any clear study that demonstrate this fact. In line with what is stated above, future research line should consider the design of more personalized nudges that really show information that is really valued by the specific user.

Regarding the ethical concerns of the mechanisms proposed, we would like to mention that the nudge mechanisms were designed to remind users of the potential audience that might see their publications. Previous research works detected that users often forget who are their friends in a social network or overwhelming the evaluation of all the possible scenarios when they share a message in the network [37]. The intended audience might be different to the final potential audience that could have access to the publication. The Picture nudge mechanism uses a list of public profile pictures of the users that may have access to a shared publication. The aim of this list is not to labeled or presented the members of the list as "risky" users. The final goal is to encourage users to be more aware of and more cautious about the privacy policy that they use when sharing information. Moreover, based on the conclusions provided by the research question RQ2, if the Profile Nudge mechanism were to be integrated into a social network platform where there was some concern about using a

list of profile images of users, Numeric Nudge mechanism could be used, as the results suggest that there are no significant differences in the effects they produce on user behavior.

The main reason for eliminating those users considered outliers within the experiment was to keep the population of users who attended the different sessions of the experiments proposed and followed the guidelines in each session. This caused the analysis of the effect of privacy nudges during posting actions on users is limited to users with a behavior within the average population. Previous research works as [55, 56, 53] highlight different kind of users taking into account their posting behavior in social networks (e.g., influencers). It would be interesting to apply different privacy nudges on different kind of users for comparing the changes in their behaviors (this, of course, for large enough population). Thus, identifying which factors and nudges improve the effect of privacy nudges for each kind of users, we would be able to maximize the effect produced on them.

The results of the experiments suggest that the use of nudge mechanisms seemed promising for assisting users in social networks activity. We encourage the inclusion of this type of mechanisms to commercial social network platforms as part of their functionality. Nudge mechanisms might be included as an optional functionality that can be activated by the users. These mechanisms could help their users to avoid any regrettable experiences disclosing information. We consider that nudges could especially help to those teenagers that start using these social platforms.

Despite the valuable conclusions extracted, the study carried out has several limitations. First, the current research was conducted for 21 consecutive days, and the nudging mechanisms were enabled only in the last 11 days. That is why only a short-term impact on users' privacy behaviors could be measured. As we stated above, we do not know the consequences of long-term usage of nudging mechanisms and their impact on behaviors. It could happen that after a certain period of time some users ignore or deactivate the nudge mechanisms. While the observed immediate effect of nudges was desirable, future research

extending the period of usage could be interesting to analyze if the effect of the nudges is stronger or if it is mitigated, and in that case, think of new nudge alternatives to maintain the effects. Second, the modeling of the experiment to test our hypotheses and the different mechanisms designed forced us to split the participants into groups. The limited number of participants in the experiments has consequences for the interpretation of the results since these cannot be generalized for the entire population of teenagers. Third, it is possible that other approaches of nudging mechanisms that are focused on the sensitivity of the post could produce more effective changes in behaviors regarding privacy. However, according to the research work described in [34], providing sentiment information about the message that is going to be published was not perceived as useful. In addition, it is often difficult to measure the effect of a nudge; users may not react to them in a noticeable way or the reaction might be gradual. Finally, the participants considered for the experiments have a certain age distribution (approx. 12-14). Therefore, these results cannot be extrapolated to users that are in other age range.

7. Conclusions

Teenagers are considered to be one of the vulnerable groups to suffer privacy risks because of their limited capacity for self-regulation and susceptibility to peer pressure. Most privacy approaches proposed in the literature try to deal with privacy in social networks to facilitate the configuration of privacy. However, there is still an open problem of making teenagers aware of the extent of disclosing information on social networks, even if users have defined a specific audience. In this paper, we focus on providing soft-paternalism mechanisms that integrate a privacy risk estimation (PRS) of the action that users are going to perform. The proposed mechanisms (nudges) attempt to influence users' decision making to improve their privacy, without actually limiting users' ability to choose freely. One of the mechanisms consists of displaying profile images of those users that might have access to the user's publication. The other mech-

anism consists of displaying the number of users that might have access to the user's publication. The proposed mechanisms are displayed when the user starts to write a message to disclose.

To evaluate the effect of mechanisms in a real context, we did a 21-day experiment with 42 teenagers ranging in age between 12 and 14 years. We included the proposed nudge mechanisms in the social network Pesedia. The experiment was divided into two stages. During stage 1, the nudges were not activated. During stage 2 the nudges were activated. We collected data about the teenagers' activity during the experiment and analyzed the privacy policy assigned to the publications. The results of the analysis show that there is a significant difference in teenagers' privacy behavior during stage 1 and stage 2. Therefore, the results suggest that the proposed nudges can be considered a useful tool for enhancing privacy awareness in social networks. The results of the analysis also show that there are no significant differences between the two nudges proposed. Finally, we analyzed the level of acceptance of the proposed nudges using a questionnaire. According to the participants' responses, nudges were not seen as irritating. The participants considered the proposed nudges to be useful for preserving their privacy.

As future work, we plan to propose new nudge mechanisms to increase privacy awareness. One of the extensions is the inclusion of an evaluation of the content of the message that users are going to disclose in order to provide a more accurate informative message about the privacy risk. Currently, we provide information about the reachability of the audience without considering the content. Another extension would be to design personalized nudges depending on what is more effective for each user. In addition, we expect to analyze the use of nudge mechanisms in two additional situations: (i) to assist users in the definition of the privacy policies associated to their profile items (i.e., profile photo, age, gender, city, etc.) and (ii) when users receive a friendship request. We also plan to do more experiments with a larger and more heterogeneous population to evaluate whether the mechanisms are appropriate for different user profiles. Moreover, we plan to introduce Pesedia in the educational context for

its continued use, so that, we can analyze the nudging effect on users' behavior regarding privacy in the long-term.

8. Acknowledgements

This work is partially supported by the Spanish Government projects and TIN2017-89156-R, the FPI grant BES-2015-074498, and the Post-Doc grant with the Ref. SP20170057.

References

[1] E. Vanderhoven, Educating teens about the risks on social network sites. an intervention study in secondary education/enseñar a los adolescentes los riesgos de las redes sociales: Una propuesta de intervención en secundaria, Comunicar 22 (43) (2014) 123.

[2] S. Livingstone, Taking risky opportunities in youthful content creation: teenagers' use of social networking sites for intimacy, privacy and self-expression, New Media & Society 10 (3) (2008) 393–411. doi:10.1177/1461444808089415.

[3] A. Lenhart, Teens, Online Stranger Contact & Cyberbullying: What the Research is Telling Us–, Pew Internet & American Life Project, 2008.

[4] A. Sengupta, A. Chaudhuri, Are social networking sites a source of online harassment for teens? evidence from survey data, Children and Youth Services Review 33 (2) (2011) 284–290.

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

[6] Y. Jeong, Y. Kim, Privacy concerns on social networking sites: Interplay among posting types, content, and audiences, Computers in Human Behavior 69 (2017) 302–310.

[7] E. Christofides, A. Muise, S. Desmarais, Hey mom, what's on your facebook? comparing facebook disclosure and privacy in adolescents and adults, Social Psychological and Personality Science 3 (1) (2012) 48–54.

[8] Y. Feng, W. Xie, Teens' concern for privacy when using social networking sites: An analysis of socialization agents and relationships with privacy-protecting behaviors, Computers in Human Behavior 33 (2014) 153–162.

[9] S. Chai, S. Bagchi-Sen, C. Morrell, H. R. Rao, S. J. Upadhyaya, Internet and online information privacy: An exploratory study of preteens and early teens, IEEE Transactions on Professional Communication 52 (2) (2009) 167–182.

[10] E. Staksrud, S. Livingstone, Children and online risk: Powerless victims or resourceful participants?, Information, Communication & Society 12 (3) (2009) 364–387.

[11] D. Albert, L. Steinberg, Judgment and decision making in adolescence, Journal of Research on Adolescence 21 (1) (2011) 211–224.

[12] K. Liu, E. Terzi, A framework for computing the privacy scores of users in online social networks, ACM Transactions on Knowledge Discovery from Data (TKDD) 5 (1) (2010) 1–6.

[13] R. K. Nepali, Y. Wang, Sonet: A social network model for privacy monitoring and ranking, in: Proc. of 33rd International Conference on Distributed Computing Systems Workshops (ICDCSW), 2013, pp. 162–166.

[14] M. Shehab, H. Touati, Semi-supervised policy recommendation for online social networks, in: Proc. of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2012, pp. 360–367.

[15] L. Fang, K. LeFevre, Privacy wizards for social networking sites, in: Proc. of the WWW, ACM, 2010, pp. 351–360.

[16] B. Vidyalakshmi, R. K. Wong, C.-H. Chi, Privacy scoring of social network users as a service, in: SCC, IEEE, 2015, pp. 218–225.

[17] I. Bilogrevic, K. Huguenin, B. Agir, M. Jadliwala, J.-P. Hubaux, Adaptive information-sharing for privacy-aware mobile social networks, in: Proc. of the Ubicomp, 2013, pp. 657–666.

[18] Z. Sun, L. Han, W. Huang, X. Wang, X. Zeng, M. Wang, H. Yan, Recommender systems based on social networks, Journal of Systems and Software 99 (2015) 109 – 119.

[19] G. Calikli, M. Law, A. K. Bandara, A. Russo, L. Dickens, B. A. Price, A. Stuart, M. Levine, B. Nuseibeh, Privacy dynamics: Learning privacy norms for social software, in: Proc. of the 11th SEAMS, ACM, 2016, pp. 47–56.

[20] Ö. Kafali, A. Günay, P. Yolum, Protoss: A run time tool for detecting privacy violations in online social networks, in: Proc. of ASONAM, 2012, pp. 429–433.

[21] Y. Mester, N. Kökciyan, P. Yolum, Negotiating privacy constraints in online social networks, in: Proc. of CARE, 2015, pp. 112–129.

[22] J. Alemany, E. del Val, J. Alberola, A. García-Fornes, Estimation of privacy risk through centrality metrics, Future Generation Computer Systems 82 (2018) 63–76.

[23] R. Balebako, P. G. Leon, H. Almuhimedi, P. G. Kelley, J. Mugan, A. Acquisti, L. F. Cranor, N. Sadeh, Nudging users towards privacy on mobile devices, in: Proc. CHI 2011 Workshop on Persuasion, Nudge, Influence and Coercion, 2011.

[24] A. Acquisti, I. Adjerid, R. Balebako, L. Brandimarte, L. F. Cranor, S. Komanduri, P. G. Leon, N. Sadeh, F. Schaub, M. Sleeper, et al., Nudges for privacy and security: Understanding and assisting users' choices online, ACM Computing Surveys (CSUR) 50 (3) (2017) 44.

[25] H. Simon, J. MARCH, ADMINISTRATIVE BEHAVIOR AND ORGANIZATIONS, New York: Free Press, 1976.

[26] Children and parents: Media use and attitudes report (2017).

[27] E. Vanderhoven, T. Schellens, R. Vanderlinde, M. Valcke, Developing educational materials about risks on social network sites: a design based research approach, Educational technology research and development 64 (3) (2016) 459–480.

[28] J. Davidson, M. Lorenz, J. Grove-Hills, E. Martellozo, Evaluation of ceop thinkuknow internet safety programme and exploration of young people's internet safety knowledge.

[29] T. Spielhofer, Children's Online Risks and Safety: A Review of the Available Evidence, National Foundation for Educational Research, 2010.

[30] H. Almuhimedi, F. Schaub, N. Sadeh, I. Adjerid, A. Acquisti, J. Gluck, L. F. Cranor, Y. Agarwal, Your location has been shared 5,398 times!: A field study on mobile app privacy nudging, in: Proceedings of the 33rd annual ACM conference on human factors in computing systems, ACM, 2015, pp. 787–796.

[31] S. Patil, X. Page, A. Kobsa, With a little help from my friends: can social navigation inform interpersonal privacy preferences?, in: Proceedings of the ACM 2011 conference on Computer supported cooperative work, ACM, 2011, pp. 391–394.

[32] B. Konings, D. Piendl, F. Schaub, M. Weber, Privacyjudge: Effective privacy controls for online published information, in: Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on, IEEE, 2011, pp. 935–941.

[33] Y. Wang, P. G. Leon, A. Acquisti, L. F. Cranor, A. Forget, N. Sadeh, A field trial of privacy nudges for facebook, in: Proceedings of the SIGCHI

conference on human factors in computing systems, ACM, 2014, pp. 2367–2376.

[34] Y. Wang, P. G. Leon, X. Chen, S. Komanduri, From facebook regrets to facebook privacy nudges, Ohio St. LJ 74 (2013) 1307.

[35] M. Bergmann, Testing privacy awareness, in: IFIP Summer School on the Future of Identity in the Information Society, Springer, 2008, pp. 237–253.

[36] I. Sim, D. Liginlal, L. Khansa, Information privacy situation awareness: construct and validation, Journal of Computer Information Systems 53 (1) (2012) 57–64.

[37] M. S. Bernstein, E. Bakshy, M. Burke, B. Karrer, Quantifying the invisible audience in social networks, in: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, 2013, pp. 21–30.

[38] P. B. Brandtzæg, M. Lüders, J. H. Skjetne, Too many facebook "friends"? content sharing and sociability versus the need for privacy in social network sites, Intl. Journal of Human–Computer Interaction 26 (11-12) (2010) 1006–1030.

[39] R. H. Thaler, C. R. Sunstein, Libertarian paternalism, The American Economic Review 93 (2) (2003) 175–179.

[40] P. G. Hansen, The definition of nudge and libertarian paternalism: Does the hand fit the glove?, European Journal of Risk Regulation 7 (1) (2016) 155–174.

[41] A. Davoudi, M. Chatterjee, Prediction of information diffusion in social networks using dynamic carrying capacity, in: Big Data (Big Data), 2016 IEEE International Conference on, IEEE, 2016, pp. 2466–2469.

[42] M. Kimura, K. Saito, Tractable models for information diffusion in social networks, in: European Conference on Principles of Data Mining and Knowledge Discovery, Springer, 2006, pp. 259–271.

[43] A. Guille, H. Hacid, C. Favre, D. A. Zighed, Information diffusion in online social networks: A survey, ACM Sigmod Record 42 (2) (2013) 17–28.

[44] J. L. Becker, H. Chen, Measuring privacy risk in online social networks.

[45] V. J. Rideout, Social media, social life: How teens view their digital lives, Common Sense Media, 2012.

[46] M. Madden, A. Lenhart, S. Cortesi, U. Gasser, M. Duggan, A. Smith, M. Beaton, Teens, social media, and privacy, Pew Research Center 21 (2013) 2–86.

[47] G. S. O'Keeffe, K. Clarke-Pearson, et al., The impact of social media on children, adolescents, and families, Pediatrics 127 (4) (2011) 800–804.

[48] S. Livingstone, Taking risky opportunities in youthful content creation: teenagers' use of social networking sites for intimacy, privacy and self-expression, New media & society 10 (3) (2008) 393–411.

[49] P. Byron, K. Albury, C. Evers, "it would be weird to have that on facebook": young people's use of social media and the risk of sharing sexual health information, Reproductive health matters 21 (41) (2013) 35–44.

[50] C. Costello, Elgg 1.8 social networking, Packt Publishing Ltd, 2012.

[51] J. Cohen, Statistical power analysis for the behavioral sciences. 2nd (1988).

[52] B. P. Knijnenburg, Simplifying privacy decisions: Towards interactive and adaptive solutions., in: Decisions@ RecSys, 2013, pp. 40–41.

[53] P. J. Wisniewski, B. P. Knijnenburg, H. R. Lipford, Making privacy personal: Profiling social network users to inform privacy education and nudging, International Journal of Human-Computer Studies 98 (2017) 95–108.

[54] L. Jedrzejczyk, B. A. Price, A. K. Bandara, B. Nuseibeh, On the impact of real-time feedback on users' behaviour in mobile location-sharing applications, in: Proceedings of the Sixth Symposium on Usable Privacy and Security, ACM, 2010, p. 14.

[55] D. Rozen, M. Askalani, T. Senn, Staring at the sun: identifying, understanding and influencing social media users, Research brief. Montreal, Canada: Aimia Inc.

[56] P. B. Brandtzaeg, J. Heim, A typology of social networking sites users, International Journal of Web Based Communities 7 (1) (2011) 28–51.