Exploring Nudge Designs to Help Adolescent SNS Users Avoid Privacy and Safety Threats

Hiroaki Masaki[1], **Kengo Shibata**[1,2], **Shui Hoshino**[3], **Takahiro Ishihama**[3], **Nagayuki Saito**[4], **Koji Yatani**[1]
[1]University of Tokyo [2]University of Geneva [3]Nanameue Inc. [4]LINE Corporation
{masaki, kengo, koji}@iis-lab.org, {shui, ishihama}@nanamenue.jp, nagayuki.saito@linecorp.com

ABSTRACT
A nudge is a method to influence individual choices without taking away freedom of choice. We are interested in whether nudges can help adolescents avoid privacy and safety threats on social network services (SNS). We conducted online surveys to compare how 11 different nudge designs influence decisions in 9 scenarios featuring various privacy and safety threats. We collected 29,608 responses from adolescent SNS users (self-claimed high school and university students), and found that nudges can help to reduce potentially risky choices. Participants were more likely to avoid potentially risky choices when presented with negative frames (e.g., "90% of users would not share a photo without permission") than affirmative ones (e.g., "10% of users would"). Social nudges displaying statistics on how likely other people would make potentially risky decisions can have a negative effect in comparison to a nudge with only general privacy and safety suggestions. We conclude by providing design considerations for privacy/safety nudges targeting adolescent SNS users.

Author Keywords
Social nudges; adolescent SNS users; online privacy and safety; large-scale survey.

CCS Concepts
•**Security and privacy** → **Social network security and privacy;** *Social aspects of security and privacy;* •**Human-centered computing** → *Empirical studies in interaction design;*

INTRODUCTION
Social network services (SNS) provide a wide variety of opportunities, connecting a vast network of friends and strangers around the world. While SNS generally benefits users, it can lead to substantial negative effects. Online risks adolescents face range from privacy breaches and cyberbullying to sexual predation [31]. These risks may result in serious consequences, including emotional distress, damaged reputation, severe isolation, and suicide [18]. In 2017, over 1,800 adolescents in Japan became victims of crimes, such as indecent assaults committed by perpetrators they came to know through the use of SNS, and this number is still increasing [17]. It is thus important to investigate new mechanics to prevent adolescents from making potentially risky decisions.

Many studies have investigated ways to assist people in engaging in more privacy/safety-concerned behavior. A nudge refers to a method of predictably influencing individual choices towards more desirable options without taking away freedom of choice [24]. Researchers have applied nudges for online privacy and security in scenarios of software installation [4], mobile app installation [7, 11], password creation [26], personal information sharing [20], and post sharing on SNS [19, 27]. According to behavioral economics and human-computer interaction literature, nudges that inform people of public opinions, called "social nudges," may be more effective than other types of nudges to prevent potential risky behavior [8, 26].

We believe that nudges may be a possible approach to help adolescents avoid threats to privacy and safety on SNS, but little quantitative evidence has been reported. The objective of this research is to examine how nudges potentially affect adolescent SNS users' decisions toward scenarios related to privacy and safety through large-scale online surveys conducted on SNS for adolescent users. We investigated diverse nudge designs: a nudge that includes a general suggestion for privacy/safety-concerning actions, a social nudge using data extracted from actual surveys, and a social nudge using fictitious data. With 29,608 responses collected through a series of surveys, we contribute novel insights on how different designs of nudges may help adolescent SNS users avoid risky decisions related to privacy and safety. We also identify scenarios where particular designs of nudges are the most and least effective.

The contributions of this work are summarized as follows:

- Large-scale online surveys with adolescent SNS users examining benefits of nudges for privacy/safety-related scenarios,
- Statistical analysis to identify scenarios where particular designs of nudges are the most and least effective, and
- Design implications on nudges for protecting privacy and safety of adolescent SNS users.

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376666

RELATED WORK

Online Risks

As social media became ubiquitous, researchers began to identify potential privacy and safety risks [10]. Livingstone addressed the issue of child online safety by reviewing its complexity given new social technology and the intricate nature of harm in everyday lives [14]. Harm is the objectively measured outcome, but the risk lies in its probability. Common risks found online that can lead to harm include privacy breaches, cyberbullying, and sexual predation [31]. Additionally, Best et al. identified issues of psychological well being, depression, and social isolation in relation to social media and online communication [2].

Privacy is a major concern for SNS usage as a lack thereof can jeopardize user's online safety. Privacy has been operationalized in research as the individual ability to control information disclosure. On the contrary, online safety can be operationalized as a transactional process of risky behavior, risk event, risk response, and risk result [31].

To address risks faced online, researchers have highlighted policy amendments, well-crafted interfaces, and the involvement of social media providers to be important [15, 29]. We aim to explore a novel interface design to prevent adolescents from engaging in potentially risky actions with the help of a social media platform. This approach builds on the hypothesis that interface design is critical in influencing how people interact with others and disclose information on SNS.

Nudges for Privacy and Safety

Developing an intervention to help individuals engage in more responsible behavior is crucial to reduce the prevalence of online harm. However, restrictions can be inefficient and may curtail the benefits of social media [16]. Instead of enforcing restrictions through parental control, systems that can help self-monitoring, impulse control, and risk coping have been highlighted to promote responsible behavior. This type of "teen-centric" solution may be effective in promoting a sense of personal agency to manage online risk and build resilience in online contexts [30]. Technology that supports such goals are sought after.

Thaler et al. popularized the concept of "nudging," a method of predictably altering individual choices towards more desirable options without taking away the freedom of choice [24]. This soft paternalistic intervention can be applied in online contexts by altering the information displayed, providing additional information, or highlighting risks.

Studies have investigated the effect of nudges in privacy and safety related decisions like those involved in the installation of applications and disclosure of personal information online. Bravo-Lillo et al. tested modifications to user interfaces to draw attention to essential information while installing software [4]. Participants exposed to certain nudges were more likely to find clues hinting that the software was malicious. Harbach et al. investigated the effect of displaying examples of private data that may be at risk in the context of Android app installation, which helped users make more privacy-conscious decisions [11]. To discourage users from installing privacy-invasive apps, Choe et al. developed visual representations of the mobile app's privacy rating [7]. They compared two visuals framed in either a positive or negative way and found positive framing was more effective than negative framing to emphasize an app that may harm a user's privacy. In parallel, Samat et al. investigated the effects of different presentations of privacy notices [20]. They found that framing impacted on information disclosure decisions particularly when a potential risk is perceived as high.

SNS interface modifications which include privacy/safety nudges can help users make more privacy/safety-concerned decisions about the contents and audience of their posts. To avoid unintended disclosures on social media, Wang et al. developed a Chrome browser plug-in to integrate nudge interfaces into Facebook pages [27]. The interface showed users a list of people who could see their Facebook post before sharing. Additionally, they developed an interface that delayed the publication of their post to give users time to reconsider uploading. These nudges were powerful tools to reduce unintended disclosures on Facebook. A further study using Facebook investigated Privacy Wedges, an interface that visually categorizes friends by interpersonal distance to encourage privacy-respecting posting [19]. This custom audience setting allows restricted sharing of posts to a specific category of friends. Prabhu designed Rethink[1], which is a keyboard application downloaded over 10,000 times. When a user tries to post an offensive message, the app automatically detects the offensive word and gives an alert. This app, developed specifically for teenagers, successfully reduced the use of abusive words.

The related work above informs us that nudges are practical tools in online privacy and security contexts. Building on this prior research, we aim to explore the potential of nudges for adolescent SNS use.

Social Nudges

Studies in psychology and behavioral economics have shown that people's decisions change when they see other people's behavior. For example, people's judgment of the length of a line can vary depending on the responses of other people [1]. A radio fundraising campaign highlighted that people who received information about contributions made by other members of the community donated more money than people without this information [23]. This social nudge can also be considered as a form of anchoring bias, a well-studied behavioral bias in the decision-making process in psychology [25].

Some political decisions utilizing this aspect of conformity have successfully encouraged people to make certain desirable choices. For example, in California, a descriptive normative message to households detailing average neighborhood power usage led to power consumption decrease for people who had consumed more energy than average [21]. Similarly, field experiments in Southern California showed that the power consumption reduced among residents who received bimonthly notifications telling average usage of households

[1] http://www.rethinkwords.com/

in the recipient's neighborhood group and the average consumption of the efficient homes in the same group [3]. They also showed that the power consumption reduced among residents who received a call or email offering similar information right before and after peak load events. In 1995, Minnesota state surveyed to increase voluntary compliance with the individual income tax [8]. People who received information about the percentage of local taxpayers were more likely to pay the tax compared to people who simply received advice on the subject. People who received a rational argument explaining the importance of tax were as likely to pay tax as people who received simple advice.

Additionally, researchers have investigated the effect of nudge designs using social norms or real data in the field of human-computer interaction. An experiment with a fake photo-driven SNS showed that participants who have seen other users upload more revealing images would have biased personal views of appropriate information to share [6]. Ur et al. designed a novel interface for password creation that used a large set of data to provide numerical feedback of guessability and actionable recommendations [26]. They found that their interface led to more secure password creation than a normal interface showing a meter with only a bar as a strength indicator.

These studies indicate that nudges using social norms or real data are applicable tools and even more convincing than simple nudges, as shown in [8, 26]. However, nudges using real data have not yet been fully explored, and it is still unknown whether data-driven privacy/safety nudges for adolescent SNS users is effective. We aim to apply nudging techniques using real data for privacy/safety issues on SNS and provide insight into behavioral changes for adolescents.

METHOD

As we discussed in the introduction, how nudges can effectively promote risk-averse behavior in adolescents remains unanswered. To quantitatively examine which nudge designs are the most effective to encourage appropriate SNS use by adolescents, we conducted online surveys. Although integration of nudges in real SNS apps or systems would be ideal for investigating the effect of their designs, interface modification for research purposes is not practical. We delivered a 2-choice questionnaire concerning privacy/safety-related decisions on Himabu, a popular SNS platform for Japanese adolescents[2]. The content of the question about potentially risky actions is, for instance, "Would you upload photos of you and your friends without permission?" in a photo sharing scenario.

Upon access to the app, randomly selected Himabu users see a pop-up with an image of an interaction on a hypothetical SNS (Figure 1). We deliberately designed pop-ups in a way that they provide common interaction scenarios, but they would not be associated with particular existing SNS. The pop-ups are also accompanied by additional textual descriptions to clarify the scenarios. Below the image of an interaction at a

[2] https://himabu.com/
Note that the service was terminated at the end of December 2019.

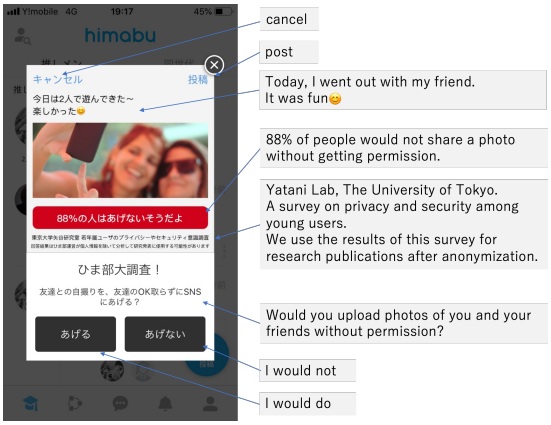

Figure 1: An example screenshot of a pop-up shown to our participants during our surveys.

hypothetical SNS, we declare our identity and an explanation about data use for informed consent. At the bottom of the pop-up, users can choose either of two possible actions: "I would do (the action suggested above)" or "I would not." The choice of "I would do" means that the user would engage in a potentially risky action. The other choice represents the opposite intention of actions. Users can also opt out from the survey by tapping the close button on the top right corner.

NUDGE DESIGN

We created multiple nudge designs informed by our literature review. Research on privacy/safety nudges suggests that giving privacy/safety-related information at appropriate timing can lead people to risk-averse behavior [4, 11]. We designed *NudgeGeneral* to investigate the effect of giving privacy/safety-related information at suitable timing. Psychology and behavioral economics have shown that people tend to follow other people's behavior [1, 23]. Existing literature also shows that data-driven nudges using the conformity of people can be more convincing than simple nudges [8, 26]. For instance, a message telling a user that most other users do not upload photos of her friends without her permission allows her to reconsider uploading. Therefore, we designed a data-driven nudge, *NudgeData*.

NudgeGeneral: Nudges Using General Guidelines

NudgeGeneral is a nudge design that displays a general suggestion to avoid privacy and safety threats (e.g., "Your friend may be uncomfortable with this." for a photo-sharing scenario). This can be considered a normal tutorial interface. We believe that it can potentially show benefits just by being displayed within the context of interactions (e.g., right before uploading a picture). According to the framework by Caraban et al. [5], *NudgeGeneral* is instantiation of a generic mechanism to remind consequences or confront.

NudgeData: Social Nudges

NudgeData is a nudge that contains survey data results taken from the same target user population. Suppose that 10% of adolescent users would share a picture on SNS without getting permission from their friends. In this case, a nudge for photo sharing would be "10% of users would (share a photo without getting permission)".

		No polarity	Polarity	
			Affirmative (Phrase how many would do)	Negative (Phrase how many would NOT do)
No intervention		*None*	—	—
Nudge with general suggestions only		*NudgeGeneral*	—	—
Nudge with real survey data		—	*NudgeData-Do*	*NudgeData-Don't*
Nudge with fictitious survey data	5%	—	*NudgeDummyData-5-Do*	*NudgeDummyData-5-Don't*
	10%	—	*NudgeDummyData-10-Do*	*NudgeDummyData-10-Don't*
	25%	—	*NudgeDummyData-25-Do*	*NudgeDummyData-25-Don't*
	40%	—	*NudgeDummyData-40-Do*	*NudgeDummyData-40-Don't*

Table 1: Eleven nudge design conditions studied in our work. *None* is the reference condition where no nudge was included.

(a) **(b)** **(c)**

Figure 2: Nudge presentations in our surveys. (a) a nudge with general privacy/safety suggestion (*NudgeGeneral*). (b) a nudge using actual survey data (*NudgeData*). (c) a nudge using fictitious survey data (in this case, the design assumes that 5% of people would choose a potentially risky action.) (*NudgeDummyData*). In (b) and (c), the nudges include negative presentations (i.e., "XX% of users would not.").

In addition, a nudge can present survey results affirmatively or negatively. The negative presentation of the previous example would be: "90% of users would not (share a photo without getting permission)". This is similar to positive and negative framing, and prior work has shown differences according to this polarity [9, 13, 28]. We thus decided to include both types of nudge descriptions as *NudgeData-Do* (affirmative presentations) and *NudgeData-Don't* (negative presentations).

NudgeDummyData: Social Nudges Using Fictitious Data

We further decided to explore the effect of presentations of survey results by introducing conditions using fictitious data (*NudgeDummyData*). There exist limitations of data-driven suggestions. First, data of specific actions or awareness is necessary. Second, real data may have an opposite effect, known as a boomerang effect [12]. For example, in California, a descriptive normative message to households detailing average neighborhood power usage led to undesirable power consumption increase for people who had consumed less energy than average [21].

Our motivation here is to examine how the effect of nudges would change depending on how far the fictitious data are from the real results. As we expected the statistics in NudgeData to vary depending on the scenarios, we also were concerned that such variance could influence the effect of such nudges. We set four different fictitious survey results: 5, 10, 25, and 40% instead of results taken from the same target user population we use in *NudgeData*. We decided to include 10, 25, and 40% to cover the spectrum of possible survey outcomes, and 5% to represent an extreme case where our participants may not believe our data. Same as *NudgeData*, we examined both affirmative and negative presentations of *NudgeDummyData*, resulting in eight conditions. In Caraban et al.'s framework [5], *NudgeData* and *NudgeDummyData* are considered as a mechanism to enable social comparisons or social influence.

We note that using fictitious data in nudges would not be possible in a real setting, and we are not advocating their use. Our intention was to investigate how different data could impact the nudging effect.

STUDY PROCEDURE

We conducted a between-subject study to quantitatively investigate the effects of different nudge designs. In total, we had 12 conditions summarized in Table 1. We included a condition where no nudge was provided (*None*) as the reference condition for comparison.

Scenarios

For studying the effect of different nudge designs, we developed common scenarios where adolescents on SNS can be at risk through a literature survey and interviews. We first conducted a literature review on existing computer literacy materials (e.g., textbooks and guidelines about Internet literacy issued in Japan). We then created a set of questions concerning potentially dangerous scenarios high school students had experienced or had heard of from their friends for our interviews with 15 high school students (12 women and 3 men). We extracted and grouped scenarios observed in our interviews, and selected nine scenarios based on two criteria for generalizability of the findings of this study: 1) scenarios would not require data or information from private communication chat channels; and 2) scenarios would not be specific to particular SNS. We also consulted with Himabu administrators to confirm that they reflected common problematic situations on their platform. Table 2 shows the nine privacy/safety-related scenarios with which we asked our participants what actions they would take.

ID	Question	Description used in *NudgeGeneral*
S1	Would you upload photos of you and your friends without permission?	Your friend may be uncomfortable with this.
S2	Would you meet a person online who you know for about a month in person ?	He/she may be a suspicious person.
S3	Would you upload a post that may show your address or the route to your school?	It may disclose your information.
S4	Would you upload a post that includes your face?	It may disclose your information.
S5	Would you pay money if you find a person who would sell a (concert) ticket you really want?	You may get ripped off.
S6	Would you communicate with a person you find on SNS if he/she has the same interests?	He/she may be a suspicious person.
S7	Would you accept a friend request from a stranger who seems to have the same interests as you?	He/she may be a suspicious person.
S8	If you have a boyfriend/girlfriend, would you upload photos with him/her?	Photos may be distributed beyond your control.
S9	Would you upload a post that shows your school uniform or school name?	It may disclose your information.

Table 2: Nine privacy/safety-related scenarios investigated in our work.

Participants

We distributed online surveys on Himabu. We consulted with the institutional review board at our institute, and confirmed that approvals from guardians are not necessary for people at the age of 15 and above. The institutional review board reviewed and approved the whole study reported in this paper. We randomly chose Himabu users who self-claimed as high school students or older as participants. They had the freedom not to participate in our study by simply clicking the close button shown at the top right corner of a pop-up. No compensation was offered in this study.

Survey Distribution

We distributed our surveys between June and August 2019. We ran surveys with each scenario for four weekdays in a week. We adopted a between-subject design because forcing our participants to respond to all of the surveys would greatly discourage participation. For the given week, the system distributed *None* and *NudgeGeneral* on Day 1. The data collected in the *None* condition were fed to *NudgeData* conditions. We then executed two *NudgeData* conditions (both *Do* and *Don't*) on Day 2. We tested *NudgeDummyData-5* and *NudgeDummyData-10* conditions on Day 3. We lastly employed *NudgeDummyData-25* and *NudgeDummyData-40* conditions on Day 4. We started our surveys at 7 pm, as many Himabu users are active during the night. We opened our surveys for 25–100 minutes depending on response rates. Each Himabu user who satisfied our participant selection criteria received one of the conditions randomly for the given day. We excluded participants who had already responded from participation in the other design conditions for that week. Thus, a participant could respond to our surveys once per scenario. In this manner, our examination guaranteed the between-subject design across the nudge design conditions.

Data Collection and Debriefing

As most of our participants were minors, we were extremely careful about data collection. We opted to collect the minimum information necessary for our investigation. Although age and geographical information would provide interesting analysis, we avoided collecting such information for privacy reasons. We did not have the exact number of unique users either because the company we collaborated with decided not to store such records for potential ethical reasons. The survey system instead internally ensured that a user was allowed to respond to a survey only once for each scenario.

As our investigation included the nature of a deception study, we conducted debriefing after our surveys were completed. The supplemental file presents our debriefing notice posted on our website[3]. The URL was shared through a post by the Himabu official Twitter as well as a pop-up notice on the Himabu app.

RESULTS

In total, we collected 38,444 answers to the online survey, along with Himabu user profile data. We checked all respondents' self-claimed educational stages with their profiles, and excluded those who we were unable to confirm that were at the age of 15 or above. This filtering led to 29,608 responses. 21,045 of them were from self-claimed high school students, and the rest were from those at universities and colleges. Table 3 shows the total number of responses per scenario and condition as well as the percentage of positive responses (i.e., "I would do").

In later analyses of our results, we use logistic regression. Logistic regression is a statistical method of modeling binary responses given a set of explanatory variables. Statistically-significant explanatory variables confirm above or below the odds ratio of 1 for the observed event of interest.

Frequency of Privacy-concerned Choices without Nudges

Table 3 shows the total number of responses and the percentage of the choice of "I would do" for each condition and scenario. We found that six scenarios (S1, S3, S4, S5, S8, and S9) received skewed responses toward privacy/safety-aware actions. On the other hand, three scenarios (S2, S6, and S7) resulted in rather polarized responses. These three scenarios were related to communication with people who participants would know only through online SNS. Our results confirm that participants were in general cautious about interactions on SNS, but approximately half of them were more open to connect with people who they do not know in person.

We further broke down the data by gender. Table 4 summarizes the results separated by gender[4]. In general, woman respondents tended to commit to privacy/safety-concerned actions. S2 and S9 were particular in gender differences, and all nudge conditions resulted in higher positive response rates in S9 than the reference condition for woman respondents.

Presence of Nudges

We first examined the contribution of the presence of nudges to committing to privacy/safety-aware actions. We created

[3] https://iis-lab.org/research/sns-nudge-debriefing/

[4] A reviewer suggested using "men/women" as opposite to "male/female". As we also found that gender study papers tend to use "men/women", we incorporated this advice in the camera-ready revisions.

Nudge conditions	S1		S2		S3		S4		S5		S6		S7		S8		S9	
	#	yes%	#	yes%	#	yes%	#	yes%	#	yes%	#	yes%	#	yes%	#	yes%	#	yes%
None	520	11.5%	691	44.7%	194	13.4%	376	30.6%	383	7.6%	419	51.6%	300	59.7%	248	29.0%	271	11.1%
NudgeGeneral	584	10.4%	740	36.5%	257	14.0%	404	25.2%	384	6.8%	391	48.8%	310	53.2%	252	23.8%	262	10.3%
NudgeData-Do	234	16.7%	576	46.4%	209	14.8%	234	32.1%	283	9.2%	364	54.9%	268	62.7%	248	26.2%	212	12.3%
NudgeData-Don't	276	13.0%	553	40.1%	173	9.2%	250	24.4%	325	9.2%	327	41.6%	282	50.7%	219	24.2%	220	10.9%
NudgeDummyData-5-Do	219	17.8%	470	41.3%	122	19.7%	171	28.1%	242	8.3%	289	53.6%	232	55.6%	148	29.7%	181	10.5%
NudgeDummyData-5-Don't	239	13.4%	512	36.7%	133	13.5%	194	19.6%	257	9.7%	304	40.1%	249	48.6%	182	26.4%	201	13.9%
NudgeDummyData-10-Do	222	16.2%	472	43.6%	119	17.6%	193	28.0%	262	5.3%	247	48.2%	250	53.6%	168	33.9%	204	18.1%
NudgeDummyData-10-Don't	250	9.6%	518	39.4%	139	12.9%	188	26.6%	238	10.5%	264	35.6%	267	44.2%	187	24.6%	175	11.4%
NudgeDummyData-25-Do	235	18.3%	474	47.0%	165	15.8%	170	26.5%	248	12.9%	271	46.5%	248	52.4%	183	27.9%	146	19.2%
NudgeDummyData-25-Don't	286	12.9%	508	36.2%	141	14.2%	174	30.5%	214	7.5%	265	41.1%	259	39.8%	184	26.1%	194	15.5%
NudgeDummyData-40-Do	217	18.4%	449	45.7%	126	23.8%	181	27.1%	251	8.8%	287	55.1%	231	60.2%	132	26.5%	164	14.0%
NudgeDummyData-40-Don't	241	12.4%	510	39.4%	126	14.3%	194	25.3%	283	10.2%	279	43.0%	267	48.3%	183	26.2%	175	10.9%

Table 3: The total numbers of responses and percentages of the "I would do" choices ("yes" responses) for each scenario and nudge condition.

five models to compare each of the five nudge design conditions of *NudgeGeneral*, *NudgeData-Do*, *NudgeData-Don't*, *NudgeDummyData-Do*, and *NudgeDummyData-Don't* against *None*, shown in Table 5. Participants were more likely to choose a privacy/safety-concerned choice in conditions of:

- *NudgeGeneral* in S2,
- *NudgeData-Don't* in S6 and S7, and
- *NudgeDummyData-Don't* in S2, S6, and S7.

Alternatively, participants were more likely to choose a risky choice in conditions of *NudgeDummyData-Do* in S1.

The results above suggest that the presence of nudges can influence on participants' choices toward privacy and safety in scenarios people do not exhibit clear consensus. They also imply that affirmative social nudges can negatively impact on privacy/safety-aware actions.

Differences by Nudge Designs

General Suggestion vs. Affirmative Social Nudges

To compare the contributions between nudges with general guidelines and affirmative social nudges, we compared the conditions of *NudgeData-Do* and *NudgeDummyData-Do* against *NudgeGeneral*. Table 6 shows logistic regression results where the default condition was set to *NudgeGeneral*. This analysis confirms the negative contributions of affirmative social nudges in three of the nine scenarios tested. We did not find any significant explanatory variable which positively contributed to privacy/safety-aware actions. The results above thus suggest that affirmative social nudges should not be used for encouraging privacy/safety-aware actions.

General Suggestion vs. Negative Social Nudges

We built a similar logistic regression model with data in the conditions of *NudgeData-Don't* and *NudgeDummyData-Don't* against *NudgeGeneral*. Table 7 presents our models for each scenario. Unlike the cases of affirmative social nudges, we found a few contributions toward privacy/safety-aware actions. Participants were more likely to choose a privacy/safety-concerned choice:

- *NudgeDummyData-5-Don't* in S6,
- *NudgeDummyData-10-Don't* in S6,
- *NudgeDummyData-10-Don't* in S7, and
- *NudgeDummyData-25-Don't* in S7.

S6 and S7 were scenarios where our participants were polarized. Thus, social nudges using actual data did not effectively persuade actions toward privacy and safety. But social nudges showing data in favor of privacy/safety-aware actions successfully led our participants to commit to privacy/safety-concerned choices.

Affirmative vs. Negative Social Nudges

We next compared contributions of the two different expressions of social nudges. We built a logistic regression model with five explanatory variables representing the condition of negative social nudges and the four conditions of *NudgeDummyData*. Table 8 shows our resulted models. We found statistically significant positive effects of negative social nudges for 5 of the 9 scenarios. In addition, participants were more likely to choose a privacy/safety-concerned choice in the conditions of:

- *NudgeDummyData-5* in S2,
- *NudgeDummyData-10* in S6,
- *NudgeDummyData-10* in S7, and
- *NudgeDummyData-25* in S7.

On the contrary, participants were more likely to choose a risky choice in *NudgeDummyData-40* in S3 and *NudgeDummyData-25* in S9.

The results above suggest that descriptions should be written in a negative presentation if future systems use social nudges. They also revealed that social nudges are not necessarily powerful when people already have one-sided opinions in privacy/safety-related scenarios.

Differences by Gender

Our interviews with high school students suggested that women users tended to be more cautious on SNS than men. We thus expected that they would be more conservative about risky actions on SNS than men. Breakdowns by gender led to similar results for all scenarios except S9 (see our supplemental document). Table 9 summarizes logistic regression results for S9 separated by gender. Man participants were more likely to choose a privacy/safety-concerned choice with social nudges in negative presentations whereas woman participants were swayed in the opposite direction.

DISCUSSION

Our study results reveal design implications for nudges for adolescent SNS users to avoid privacy/safety-risky actions.

Nudge conditions	S1		S2		S3		S4		S5		S6		S7		S8		S9	
	#	yes%	#	yes%	#	yes%	#	yes%	#	yes%	#	yes%	#	yes%	#	yes%	#	yes%
Men																		
None	245	12.7%	340	58.8%	94	16.0%	197	28.9%	184	10.9%	212	53.8%	155	62.6%	134	27.6%	127	19.7%
NudgeGeneral	299	10.0%	368	50.0%	119	14.3%	205	22.9%	201	9.0%	194	48.5%	164	55.5%	122	18.9%	130	11.5%
NudgeData-Do	131	17.6%	279	63.1%	101	16.8%	111	29.7%	150	12.0%	207	54.1%	137	62.0%	141	26.2%	90	17.8%
NudgeData-Don't	138	12.3%	280	55.0%	80	10.0%	117	19.7%	171	11.1%	170	41.8%	155	56.1%	104	24.0%	102	8.8%
NudgeDummyData-5-Do	113	19.5%	235	55.3%	46	26.1%	91	27.5%	124	12.1%	161	52.8%	130	58.5%	75	29.3%	98	12.2%
NudgeDummyData-5-Don't	130	18.5%	245	51.8%	67	20.9%	100	18.0%	129	12.4%	149	43.6%	119	56.3%	95	25.3%	99	15.2%
NudgeDummyData-10-Do	115	14.8%	252	57.9%	51	25.5%	93	26.9%	143	6.3%	115	53.0%	125	56.8%	80	41.3%	96	21.9%
NudgeDummyData-10-Don't	133	9.0%	249	53.0%	58	13.8%	96	18.8%	127	14.2%	147	38.1%	145	46.2%	90	21.1%	101	8.9%
NudgeDummyData-25-Do	122	18.9%	256	60.5%	62	19.4%	80	27.5%	129	17.8%	157	43.9%	118	49.2%	90	22.2%	74	20.3%
NudgeDummyData-25-Don't	153	15.0%	255	51.4%	67	23.9%	96	27.1%	126	6.3%	140	37.9%	140	39.3%	91	23.1%	95	14.7%
NudgeDummyData-40-Do	110	19.1%	229	57.6%	61	29.5%	104	26.9%	140	10.0%	162	50.6%	123	62.6%	56	30.4%	78	16.7%
NudgeDummyData-40-Don't	138	13.0%	268	50.4%	62	24.2%	101	24.8%	135	11.1%	148	43.2%	141	57.4%	96	21.9%	86	12.8%
Women																		
None	220	12.7%	300	29.3%	80	8.8%	155	32.3%	163	4.9%	182	50.0%	122	59.0%	95	28.4%	128	3.9%
NudgeGeneral	232	11.6%	316	22.2%	121	11.6%	177	27.1%	158	4.4%	163	47.2%	127	52.8%	104	27.9%	108	9.3%
NudgeData-Do	86	17.4%	247	28.7%	94	12.8%	99	35.4%	111	6.3%	125	56.8%	116	62.9%	88	27.3%	110	7.3%
NudgeData-Don't	110	14.5%	228	24.1%	76	3.9%	107	28.0%	125	8.0%	129	43.4%	106	45.3%	98	26.5%	97	11.3%
NudgeDummyData-5-Do	85	15.3%	208	25.0%	66	13.6%	61	31.1%	98	4.1%	107	57.9%	83	56.6%	55	30.9%	70	10.0%
NudgeDummyData-5-Don't	91	6.6%	240	20.8%	51	5.9%	75	21.3%	103	4.9%	136	37.5%	105	46.7%	75	29.3%	86	10.5%
NudgeDummyData-10-Do	87	17.2%	194	25.8%	54	9.3%	82	29.3%	96	4.2%	104	47.1%	105	53.3%	73	28.8%	90	16.7%
NudgeDummyData-10-Don't	104	7.7%	226	24.8%	66	10.6%	78	34.6%	97	5.2%	99	33.3%	105	44.8%	77	26.0%	62	9.7%
NudgeDummyData-25-Do	91	16.5%	184	30.4%	82	13.4%	77	24.7%	101	6.9%	103	50.5%	106	58.5%	78	35.9%	57	19.3%
NudgeDummyData-25-Don't	108	12.0%	224	21.4%	61	4.9%	63	34.9%	78	9.0%	105	43.8%	105	41.0%	73	26.0%	85	17.6%
NudgeDummyData-40-Do	94	19.1%	187	32.6%	55	18.2%	67	29.9%	98	8.2%	106	58.5%	86	60.5%	59	27.1%	74	12.2%
NudgeDummyData-40-Don't	81	12.3%	204	23.0%	54	3.7%	86	27.9%	122	10.7%	116	43.1%	112	37.5%	79	31.6%	78	10.3%
Gender Unspecified																		
None	55	1.8%	51	41.2%	20	20.0%	24	33.3%	36	2.8%	25	44.0%	23	43.5%	19	42.1%	16	0.0%
NudgeGeneral	53	7.5%	56	28.6%	17	29.4%	22	31.8%	25	4.0%	34	58.8%	19	36.8%	26	30.8%	24	8.3%
NudgeData-Do	17	5.9%	50	40.0%	14	14.3%	24	29.2%	22	4.5%	32	53.1%	15	66.7%	19	21.1%	12	16.7%
NudgeData-Don't	28	10.7%	45	28.9%	17	29.4%	26	30.8%	29	3.4%	28	32.1%	21	38.1%	17	11.8%	21	19.0%
NudgeDummyData-5-Do	21	19.0%	27	44.4%	10	30.0%	19	21.1%	20	5.0%	21	38.1%	19	31.6%	18	27.8%	13	0.0%
NudgeDummyData-5-Don't	18	11.1%	27	40.7%	15	6.7%	19	21.1%	25	16.0%	19	31.6%	25	20.0%	12	16.7%	16	25.0%
NudgeDummyData-10-Do	20	20.0%	26	38.5%	14	21.4%	18	27.8%	23	4.3%	28	32.1%	20	35.0%	15	20.0%	18	5.6%
NudgeDummyData-10-Don't	13	30.8%	43	37.2%	15	20.0%	14	35.7%	14	14.3%	18	27.8%	17	23.5%	20	35.0%	12	41.7%
NudgeDummyData-25-Do	22	22.7%	34	35.3%	21	14.3%	13	30.8%	18	11.1%	11	45.5%	24	41.7%	15	20.0%	15	13.3%
NudgeDummyData-25-Don't	25	4.0%	29	17.2%	13	7.7%	15	33.3%	10	10.0%	20	50.0%	14	35.7%	20	40.0%	14	7.1%
NudgeDummyData-40-Do	13	7.7%	33	36.4%	10	20.0%	10	10.0%	13	0.0%	19	73.7%	22	45.5%	17	11.8%	12	8.3%
NudgeDummyData-40-Don't	22	9.1%	38	50.0%	10	10.0%	7	0.0%	26	3.8%	15	40.0%	14	42.9%	8	25.0%	11	0.0%

Table 4: The total numbers of responses and percentages of the "I would do" choices ("yes" responses) for each scenario and nudge condition separately by gender. Note that responses from users who did not specify gender were excluded.

- Nudges with general suggestions or negative social nudges can be effective in scenarios to which adolescent SNS users exhibit polarized attitudes,
- Nudges would not be powerful in scenarios where a large majority of adolescent SNS users are already aware of privacy/safety-concerned choices, and
- Affirmative social nudges should be avoided for privacy and safety promotion purposes.

We confirmed that the presence of nudges can help the avoidance of potentially risky actions when people have polarized opinions (Table 5). In S6 and S7, social nudges using fictitious survey data in negative presentations may further reduce the likelihood to choose potentially risky actions when compared to nudges with general suggestions (Table 7). This result suggests that an understanding of how often people would make risky choices is important to identify scenarios where nudges can be useful.

Our analysis does not confirm the benefits of nudges in cases where a large majority of users already support privacy/safety-concerned choices. In these scenarios, participants were likely to be aware that they should commit to privacy/safety-concerned choices. As a result, nudges would not be effective compared to scenarios where responses without nudges were polarized. This suggests that SNS may refrain from using nudges for such scenarios to avoid overwhelming users unnecessarily.

Our results also reveal the negative effect of affirmative social nudges. In some cases where we found statistical significance, affirmative social nudges could greatly increase the likelihood of potentially risky choices (e.g., *NudgeDummyData-40-Do* in S1 and *NudgeDummyData-25-Do* in S9 in Table 6) in comparison to *NudgeGeneral*. These results clearly suggest a boomerang effect of social nudges [12]. We thus conclude that affirmative social nudges should be avoided for any case of privacy and safety promotion purposes.

The advantage of social nudges using fictitious data was clear only in S6 and S7 (i.e., *NudgeDummyData-10* in S6 and S7 and *NudgeDummyData-25* in S7 in Table 8). In these cases, fictitious data were more favorable to privacy/safety-concerned actions than actual survey results. However, using fictitious data can be misleading even if it is intended to be for good purposes. Furthermore, the advantages of fictitious data were unclear in the other scenarios. We thus suggest using actual survey data in negative presentations or general privacy/safety suggestions in practical settings.

Differences by gender observed in S9 (Table 9) show a large deviation from the other results. In this particular instance, women participants were more likely to choose

	S1	S2	S3	S4	S5	S6	S7	S8	S9
NudgeGeneral against *None*									
OR	0.89	**0.71**	1.05	0.77	0.89	0.90	0.77	0.76	0.92
	[0.61,1.30]	**[0.57,0.88]**	[0.61,1.81]	[0.56,1.05]	[0.51,1.54]	[0.68,1.18]	[0.56,1.06]	[0.51,1.14]	[0.53,1.60]
p	0.56	<**0.01**	0.85	0.10	0.67	0.44	0.11	0.19	0.78
NudgeData-Do against *None*									
OR	1.53	1.07	1.13	1.07	1.23	1.15	1.14	0.87	1.12
	[0.99,2.37]	[0.86,1.33]	[0.64,1.97]	[0.75,1.52]	[0.71,2.15]	[0.86,1.52]	[0.81,1.59]	[0.59,1.29]	[0.64,1.96]
p	0.05	0.56	0.68	0.70	0.45	0.34	0.46	0.48	0.68
NudgeData-Don't against *None*									
OR	1.15	0.83	0.66	0.73	1.24	**0.67**	**0.70**	0.78	0.98
	[0.74,1.79]	[0.66,1.04]	[0.34,1.27]	[0.51,1.05]	[0.73,2.12]	**[0.50,0.90]**	**[0.50,0.97]**	[0.52,1.18]	[0.56,1.74]
p	0.54	0.11	0.21	0.09	0.43	<**0.01**	<**0.05**	0.24	0.95
NudgeDummyData-Do against *None*									
OR	**1.65**	0.99	1.51	0.86	1.17	0.98	0.84	1.03	1.46
	[1.20,2.27]	[0.83,1.18]	[0.95,2.41]	[0.65,1.13]	[0.76,1.82]	[0.78,1.23]	[0.64,1.09]	[0.75,1.42]	[0.95,2.25]
p	<**0.01**	0.88	0.08	0.27	0.47	0.85	0.19	0.86	0.08
NudgeDummyData-Don't against *None*									
OR	1.06	**0.76**	1.03	0.77	1.29	**0.63**	**0.56**	0.85	1.20
	[0.76,1.47]	**[0.63,0.90]**	[0.64,1.66]	[0.59,1.01]	[0.84,1.99]	**[0.50,0.79]**	**[0.43,0.72]**	[0.62,1.17]	[0.78,1.86]
p	0.75	<**0.01**	0.91	0.06	0.25	<**0.001**	<**0.001**	0.32	0.41

Table 5: Odds ratios (OR) and their 95% confidence intervals observed in the logistic regression comparing each of the nudge conditions (*NudgeGeneral*, *NudgeData*, and *NudgeDummyData*) against *None*.

	S1	S2	S3	S4	S5	S6	S7	S8	S9
NudgeData-Do against *NudgeGeneral*									
OR	**1.71**	**1.50**	1.07	1.40	1.39	1.28	**1.48**	1.14	1.22
	[1.11,2.65]	**[1.20,1.88]**	[0.64,1.80]	[0.98,1.99]	[0.79,2.46]	[0.96,1.70]	**[1.06,2.06]**	[0.76,1.70]	[0.69,2.16]
p	<**0.05**	<**0.001**	0.80	0.06	0.25	0.09	<**0.05**	0.54	0.50
NudgeDummyData-5-Do against *NudgeGeneral*									
OR	**1.86**	1.22	1.5	1.16	1.24	1.21	1.10	1.35	1.02
	[1.20,2.87]	[0.97,1.55]	[0.85,2.65]	[0.77,1.73]	[0.68,2.28]	[0.89,1.64]	[0.78,1.55]	[0.86,2.14]	[0.55,1.90]
p	<**0.01**	0.10	0.16	0.48	0.49	0.22	0.58	0.19	0.95
NudgeDummyData-10-Do against *NudgeGeneral*									
OR	**1.66**	**1.35**	1.32	1.15	0.78	0.97	1.02	**1.64**	**1.93**
	[1.06,2.59]	**[1.07,1.71]**	[0.73,2.37]	[0.78,1.69]	[0.40,1.52]	[0.71,1.34]	[0.73,1.42]	**[1.07,2.53]**	**[1.13,3.29]**
p	<**0.05**	<**0.05**	0.36	0.48	0.46	0.87	0.93	<**0.05**	<**0.05**
NudgeDummyData-25-Do against *NudgeGeneral*									
OR	**1.92**	**1.55**	1.15	1.07	**2.04**	0.91	0.97	1.24	**2.07**
	[1.26,2.93]	**[1.22,1.96]**	[0.66,1.98]	[0.71,1.60]	**[1.18,3.52]**	[0.67,1.24]	[0.69,1.35]	[0.80,1.91]	**[1.16,3.66]**
p	<**0.01**	<**0.001**	0.62	0.76	<**0.05**	0.55	0.85	0.34	<**0.05**
NudgeDummyData-40-Do against *NudgeGeneral*									
OR	**1.94**	**1.46**	**1.92**	1.10	1.32	1.28	1.33	1.15	1.42
	[1.26,2.99]	**[1.15,1.86]**	**[1.12,3.29]**	[0.74,1.64]	[0.73,2.39]	[0.94,1.74]	[0.94,1.88]	[0.71,1.87]	[0.78,2.57]
p	<**0.01**	<**0.01**	<**0.05**	0.64	0.35	0.11	0.11	0.56	0.25

Table 6: Odds ratios (OR) and their 95% confidence intervals observed in the logistic regression comparing affirmative social nudges (*NudgeData-Do* and *NudgeDummyData-Do*) against *NudgeGeneral*.

	S1	S2	S3	S4	S5	S6	S7	S8	S9
NudgeData-Don't against *NudgeGeneral*									
OR	1.29	1.17	0.63	0.96	1.40	0.75	0.90	1.02	1.07
	[0.83,2.00]	[0.93,1.46]	[0.34,1.17]	[0.66,1.38]	[0.81,2.42]	[0.55,1.00]	[0.65,1.25]	[0.67,1.56]	[0.60,1.91]
p	0.26	0.18	0.14	0.81	0.23	0.05	0.54	0.92	0.83
NudgeDummyData-5-Don't against *NudgeGeneral*									
OR	1.33	1.01	0.96	0.72	1.48	**0.70**	0.83	1.15	1.41
	[0.84,2.09]	[0.80,1.28]	[0.52,1.77]	[0.47,1.10]	[0.84,2.63]	**[0.52,0.95]**	[0.59,1.16]	[0.74,1.78]	[0.80,2.48]
p	0.23	0.93	0.90	0.13	0.18	<**0.05**	0.28	0.54	0.23
NudgeDummyData-10-Don't against *NudgeGeneral*									
OR	0.91	1.13	0.91	1.07	1.62	**0.58**	**0.70**	1.04	1.12
	[0.55,1.50]	[0.90,1.43]	[0.50,1.68]	[0.72,1.59]	[0.91,2.87]	**[0.42,0.80]**	**[0.50,0.97]**	[0.67,1.62]	[0.61,2.07]
p	0.71	0.30	0.77	0.73	0.10	<**0.001**	<**0.05**	0.85	0.71
NudgeDummyData-25-Don't against *NudgeGeneral*									
OR	1.27	0.99	1.01	1.30	1.11	0.73	**0.58**	1.13	1.59
	[0.82,1.97]	[0.78,1.25]	[0.56,1.83]	[0.88,1.92]	[0.58,2.12]	[0.53,1.00]	**[0.42,0.81]**	[0.73,1.75]	[0.91,2.78]
p	0.28	0.92	0.96	0.20	0.75	0.05	<**0.01**	0.59	0.10
NudgeDummyData-40-Don't against *NudgeGeneral*									
OR	1.22	1.13	1.02	1.00	1.57	0.79	0.82	1.14	1.06
	[0.77,1.94]	[0.90,1.43]	[0.56,1.88]	[0.67,1.48]	[0.90,2.73]	[0.58,1.08]	[0.59,1.14]	[0.73,1.76]	[0.57,1.97]
p	0.40	0.29	0.94	1.00	0.11	0.14	0.24	0.56	0.85

Table 7: Odds ratios (OR) and their 95% confidence intervals observed in the logistic regression comparing negative social nudges (*NudgeData-Don't* and *NudgeDummyData-Don't*) against *NudgeGeneral*.

	S1	S2	S3	S4	S5	S6	S7	S8	S9
NudgeDummyData-5 against *NudgeData*									
OR	1.06	**0.84**	1.45	0.79	0.98	0.94	0.83	1.16	1.08
	[0.74,1.50]	**[0.71,1.00]**	[0.92,2.28]	[0.58,1.08]	[0.65,1.48]	[0.76,1.18]	[0.65,1.06]	[0.84,1.60]	[0.70,1.64]
p	0.76	<**0.05**	0.11	0.14	0.92	0.60	0.14	0.36	0.74
NudgeDummyData-10 against *NudgeData*									
OR	0.84	0.93	1.32	0.96	0.84	**0.77**	**0.73**	1.22	1.34
	[0.58,1.21]	[0.78,1.11]	[0.83,2.09]	[0.71,1.29]	[0.55,1.28]	**[0.61,0.97]**	**[0.57,0.93]**	[0.90,1.67]	[0.89,2.02]
p	0.35	0.43	0.24	0.77	0.42	<**0.05**	<**0.05**	0.21	0.16
NudgeDummyData-25 against *NudgeData*									
OR	1.06	0.93	1.27	1.02	1.15	0.83	**0.65**	1.10	**1.59**
	[0.75,1.49]	[0.78,1.11]	[0.82,1.96]	[0.75,1.38]	[0.77,1.73]	[0.66,1.04]	**[0.51,0.83]**	[0.80,1.50]	**[1.06,2.39]**
p	0.76	0.43	0.29	0.91	0.50	0.11	<**0.001**	0.55	<**0.05**
NudgeDummyData-40 against *NudgeData*									
OR	1.04	0.97	**1.72**	0.91	1.04	1.03	0.90	1.08	1.08
	[0.73,1.48]	[0.82,1.16]	**[1.10,2.66]**	[0.67,1.23]	[0.70,1.55]	[0.82,1.29]	[0.71,1.15]	[0.78,1.50]	[0.70,1.68]
p	0.83	0.75	<**0.05**	0.52	0.84	0.80	0.42	0.65	0.72
Using negative presentations									
OR	**0.66**	**0.77**	**0.66**	0.84	1.08	**0.63**	**0.65**	0.84	0.83
	[0.53,0.83]	**[0.69,0.86]**	**[0.49,0.88]**	[0.69,1.03]	[0.83,1.41]	**[0.54,0.73]**	**[0.56,0.76]**	[0.69,1.04]	[0.63,1.08]
p	<**0.001**	<**0.001**	<**0.01**	0.09	0.58	<**0.001**	<**0.001**	0.11	0.16

Table 8: Odds ratios (OR) and their 95% confidence intervals observed in the logistic regression comparing *NudgeDummyData* against *NudgeData*. We also included an explanatory variable representing whether a nudge was in a negative presentation.

S9 by gender	Men	Women
NudgeGeneral against *None*		
OR	0.53	2.51
	[0.27,1.06]	[0.83,7.59]
p	0.07	0.10
NudgeData-Do against *None*		
OR	0.88	1.93
	[0.44,1.77]	[0.61,6.08]
p	0.72	0.26
NudgeData-Don't against *None*		
OR	**0.39**	**3.15**
	[0.18,0.89]	**[1.06,9.38]**
p	<**0.05**	<**0.05**
NudgeDummyData-Do against *None*		
OR	0.87	**4.15**
	[0.52,1.47]	**[1.60,10.75]**
p	0.61	<**0.01**
NudgeDummyData-Don't against *None*		
OR	0.60	**3.42**
	[0.35,1.02]	**[1.32,8.91]**
p	0.06	<**0.05**

Table 9: Odds ratios (OR) and their 95% confidence intervals observed in the logistic regression comparing each of the nudge conditions against *None* separated by gender. Please refer to our supplemental document for the analysis results of the other scenarios. Note that the "yes" response rate by participants whose gender was unspecified in the *None* condition was 0%, and logistic regression analysis did not show meaning results.

potentially risky actions with three kinds of social nudges. The rate to choose potentially risky actions in the conditions without nudges was 3.9% whereas it was 11.1% for the entire participant group. One explanation of this deviation was that data shown in social nudges were higher than the actual statistics, resulting in persuasion to potentially risky actions.

LIMITATIONS

There are several limitations of this work to be discussed. Our study is not intended to demonstrate the actual effects of nudges on interaction and communication on SNS. Due to the hypothetical nature of the study, the survey responses represent intended actions by our participants on given scenarios. Instead of directly integrating nudges to SNS, we conducted online surveys, collecting large-scale data from adolescent SNS users, to investigate the potential effect of different nudge designs to promote privacy and safety. Meta analysis by Schwenk and Möser reports a correlation of 0.54 between intention and behavior in environmental studies [22]. We thus believe that our study still offers strong insights on nudge designs even if its design included a hypothetical nature and observed people's intention.

Our study was conducted on one particular SNS (Himabu), where most users are Japanese. Thus, this homogeneous cultural background may have influenced our results. Future work should examine how different cultural backgrounds could impact nudges on privacy/safety action choices. Pop-ups used in our surveys deliberately avoid screenshots of widgets or features or interactions which people would be able to immediately associate with particular existing SNS. People may exhibit different privacy/safety-concerned actions depending on SNS (e.g., not sharing photos on Instagram, but on Facebook). Future work should examine how different SNS could impact on the benefits of nudges.

Profile data used in our analysis were all self-claimed by each user, and we did not have any method to validate this data. There may be users who deliberately enter fake profiles, which can compound our results and analysis. The administration team of Himabu is proactively removing users with fake profiles. Our results can be re-validated in the future with more targeted qualitative examinations (e.g., focus groups with high school students).

CONCLUSION

Protecting adolescents from privacy and safety threats is critical as SNS is very popular. We conducted large-scale online surveys examining behavior choices of adolescent SNS users in potentially privacy and safety-risky scenarios. With 29,608 responses collected through online surveys, we found that nudges with general descriptions or negative social nudges can be effective in scenarios to which adolescent SNS users exhibit polarized attitudes. Our study also revealed that nudges would not be powerful in scenarios where a large majority of users already lean toward privacy/safety-concerned choices. Social nudges in affirmative presentations should be avoided for purposes of protecting privacy and safety.

ACKNOWLEDGEMENTS

We greatly appreciate Kiyotaka Eguchi, Hatsue Arima, and Fuyuko Kido at LINE Cooperation for their huge support on this research. We also thank Carla F. Griggio, Hidenori Matsui, Tatsuhiko Sakaguchi, Arissa J. Sato, Zefan Sramek, Asahi Takenouchi, and Zhongyi Zhou for their insightful feedback on this project and paper. This research was partly supported by a collaboration funding "An investigation on the effect of interface designs to encourage privacy/safety-aware actions on SNS" provided by LINE Cooperation.

REFERENCES

[1] Solomon E. Asch. 1951. Effects of group pressure upon the modification and distortion of judgments. In *Groups, leadership and men; research in human relations*. Carnegie Press, Oxford, England, 177–190.

[2] Paul Best, Roger Manktelow, and Brian Taylor. 2014. Online communication, social media and adolescent wellbeing: A systematic narrative review. *Children and Youth Services Review* 41 (2014), 27–36.

[3] Alec Brandon, John A. List, Robert D. Metcalfe, Michael K. Price, and Florian Rundhammer. 2019. Testing for crowd out in social nudges: Evidence from a natural field experiment in the market for electricity. *Proceedings of the National Academy of Sciences* 116, 12 (2019), 5293–5298. DOI: http://dx.doi.org/10.1073/pnas.1802874115

[4] Cristian Bravo-Lillo, Saranga Komanduri, Lorrie Faith Cranor, Robert W. Reeder, Manya Sleeper, Julie Downs, and Stuart Schechter. 2013. Your Attention Please: Designing Security-decision UIs to Make Genuine Risks Harder to Ignore. In *Proceedings of the Ninth Symposium on Usable Privacy and Security (SOUPS '13)*. ACM, New York, NY, USA, Article 6, 12 pages. DOI:http://dx.doi.org/10.1145/2501604.2501610

[5] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 2019. 23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction. 1–15. DOI: http://dx.doi.org/10.1145/3290605.3300733

[6] Daphne Chang, Erin L. Krupka, Eytan Adar, and Alessandro Acquisti. 2016. Engineering Information Disclosure: Norm Shaping Designs. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 587–597. DOI: http://dx.doi.org/10.1145/2858036.2858346

[7] Eun Kyoung Choe, Jaeyeon Jung, Bongshin Lee, and Kristie Fisher. 2013. Nudging people away from privacy-invasive mobile apps through visual framing. In *IFIP Conference on Human-Computer Interaction*. Springer, 74–91.

[8] Stephen Coleman. 1996. The Minnesota Income Tax Compliance Experiment. (05 1996).

[9] Mark A. Davis and Philip Bobko. 1986. Contextual effects on escalation processes in public sector decision making. *Organizational Behavior and Human Decision Processes* 37, 1 (1986), 121–138. DOI: http://dx.doi.org/https://doi.org/10.1016/0749-5978(86)90048-8

[10] Ralph Gross and Alessandro Acquisti. 2005. Information Revelation and Privacy in Online Social Networks. In *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society (WPES '05)*. ACM, New York, NY, USA, 71–80. DOI: http://dx.doi.org/10.1145/1102199.1102214

[11] Marian Harbach, Markus Hettig, Susanne Weber, and Matthew Smith. 2014. Using Personal Examples to Improve Risk Communication for Security & Privacy Decisions. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2647–2656. DOI: http://dx.doi.org/10.1145/2556288.2556978

[12] Carl I. Hovland, Irving L. Janis, and Harold H. Kelley. 1953. *Communication and persuasion.* Yale University Press, New Haven, CT, US.

[13] Irwin Levin and Gary Gaeth. 1988. How Consumers Are Affected by the Framing of Attribute Information Before and After Consuming the Product. *Journal of Consumer Research* 15 (02 1988), 374–378. DOI: http://dx.doi.org/10.1086/209174

[14] Sonia Livingstone. 2013. Online risk, harm and vulnerability: Reflections on the evidence base for child internet safety policy. 18 (11 2013), 13–28.

[15] Sonia Livingstone and David R Brake. 2010. On the rapid rise of social networking sites: New findings and policy implications. *Children & society* 24, 1 (2010), 75–83.

[16] Sonia Livingstone and Ellen Helsper. 2008. Parental Mediation of Children's Internet Use. *Journal of Broadcasting Electronic Media - J BROADCAST ELECTRON MEDIA* 52 (11 2008), 581–599. DOI: http://dx.doi.org/10.1080/08838150802437396

[17] National Police Agency. 2018. Report of incidents and countermeasures of underaged victims of malicious online interaction through SNS etc. in 2017. http://www.npa.go.jp/safetylife/syonen/H29_sns_shiryo.pdf. (2018). [Online; accessed 26-August-2019].

[18] Gwenn Schurgin O'Keeffe, Kathleen Clarke-Pearson, and Council on Communications and Media. 2011. The Impact of Social Media on Children, Adolescents, and Families. *Pediatrics* 127, 4 (04 2011), 800–804. DOI: http://dx.doi.org/10.1542/peds.2011-0054

[19] Frederic Raber, Alexander De Luca, and Moritz Graus. 2016. Privacy Wedges: Area-Based Audience Selection for Social Network Posts. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. USENIX Association, Denver, CO. https://www.usenix.org/conference/soups2016/workshop-program/wpi/presentation/raber

[20] Sonam Samat and Alessandro Acquisti. 2017. Format vs. Content: The Impact of Risk and Presentation on Disclosure Decisions. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*. USENIX Association, Santa Clara, CA, 377–384. https://www.usenix.org/conference/soups2017/technical-sessions/presentation/samat-disclosure

[21] P. Wesley Schultz, Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius. 2007. The Constructive, Destructive, and Reconstructive Power of Social Norms. *Psychological Science* 18, 5 (2007), 429–434. DOI: http://dx.doi.org/10.1111/j.1467-9280.2007.01917.x PMID: 17576283.

[22] Gero Schwenk and Guido Möser. 2009. Intention and behavior: a Bayesian meta-analysis with focus on the Ajzen–Fishbein Model in the field of environmental behavior. *Quality & Quantity* 43, 5 (09 2009), 743–755. DOI:http://dx.doi.org/10.1007/s11135-007-9162-7

[23] Jen Shang and Rachel Croson. 2009. A Field Experiment in Charitable Contribution: The Impact of Social Information on the Voluntary Provision of Public Goods.

[24] Richard Thaler and Cass Sunstein. 2009. *NUDGE: Improving Decisions About Health, Wealth, and Happiness*. Vol. 47.

[25] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* 185, 4157 (1974), 1124–1131. DOI: http://dx.doi.org/10.1126/science.185.4157.1124

[26] Blase Ur, Felicia Alfieri, Maung Aung, Lujo Bauer, Nicolas Christin, Jessica Colnago, Lorrie Faith Cranor, Henry Dixon, Pardis Emami Naeini, Hana Habib, Noah Johnson, and William Melicher. 2017. Design and Evaluation of a Data-Driven Password Meter. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3775–3786. DOI: http://dx.doi.org/10.1145/3025453.3026050

[27] Yang Wang, Pedro Giovanni Leon, Alessandro Acquisti, Lorrie Faith Cranor, Alain Forget, and Norman Sadeh. 2014. A Field Trial of Privacy Nudges for Facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2367–2376. DOI: http://dx.doi.org/10.1145/2556288.2557413

[28] Dawn K. Wilson, Robert M. Kaplan, and Lawrence J. Schneiderman. 1987. Framing of decisions and selections of alternatives in health care. *Social Behaviour* 2, 1 (1987), 51–59.

[29] Pamela Wisniewski, Heng Xu, Mary Beth Rosson, Daniel F. Perkins, and John M. Carroll. 2016. Dear Diary: Teens Reflect on Their Weekly Online Risk Experiences. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 3919–3930. DOI:http://dx.doi.org/10.1145/2858036.2858317

[30] Pamela J. Wisniewski. 2018. The Privacy Paradox of Adolescent Online Safety: A Matter of Risk Prevention or Risk Resilience? *IEEE Security Privacy* 16, 2 (March 2018), 86–90. DOI: http://dx.doi.org/10.1109/MSP.2018.1870874

[31] Pamela J. Wisniewski, Heng Xu, Jack M. Carroll, and Mary Beth Rosson. 2013. Grand Challenges of Researching Adolescent Online Safety: A Family Systems Approach. In *AMCIS*.