A Field Trial of Privacy Nudges for Facebook

Yang Wang,[‡] Pedro Giovanni Leon,[*] Alessandro Acquisti,[*]
Lorrie Faith Cranor,[*] Alain Forget,[*] and Norman Sadeh[*]

[‡]Syracuse University
ywang@syr.edu

[*]Carnegie Mellon University
{pedrogln, acquisti, lorrie, aforget, sadeh}@cmu.edu

ABSTRACT

Anecdotal evidence and scholarly research have shown that Internet users may regret some of their online disclosures. To help individuals avoid such regrets, we designed two modifications to the Facebook web interface that nudge users to consider the content and audience of their online disclosures more carefully. We implemented and evaluated these two nudges in a 6-week field trial with 28 Facebook users. We analyzed participants' interactions with the nudges, the content of their posts, and opinions collected through surveys. We found that reminders about the audience of posts can prevent unintended disclosures without major burden; however, introducing a time delay before publishing users' posts can be perceived as both beneficial and annoying. On balance, some participants found the nudges helpful while others found them unnecessary or overly intrusive. We discuss implications and challenges for designing and evaluating systems to assist users with online disclosures.

Author Keywords

Behavioral bias; Online disclosure; Social media; Facebook; Nudge; Privacy; Regret; Soft-paternalism

ACM Classification Keywords

H.5.m Information Interfaces and Presentation (e.g., HCI): Miscellaneous

INTRODUCTION

Online social networks such as Facebook are designed to encourage sharing, facilitating the seamless and immediate broadcasting of all kinds of information. While sharing information through social networks generally benefits users, seemingly innocuous disclosures can lead to substantial negative consequences. Lack of awareness of the potential audience, posting while in highly emotional states, and hasty disclosures have been shown to lead social media users to experience regret [28]. Research in the fields of psychology, behavioral economics, and behavioral decision making has uncovered cognitive and behavioral biases that affect decision making. These biases are systematic deviations from what traditional economists call rational decisions. Furthermore, when limited resources (e.g., time or information) are available to make a decision, human beings often rely on heuristics or shortcuts. These biases and heuristics have been shown to impact privacy decisions [1, 4, 5] and privacy blunders in social media are vivid examples of the hurdles users face.

Behavioral economists have proposed the use of soft paternalistic interventions to help people overcome behavioral biases that affect decision making. These interventions are designed to "nudge" (instead of force) people towards behaviors that have been shown to be publicly desired, but difficult to follow, without limiting people's autonomy [24]. Acquisti has proposed to use soft paternalistic interventions to improve security and privacy decisions [2]. We refer to soft-paternalistic mechanisms that nudge people towards more thoughtful and informed privacy-related decisions as *privacy nudges*.

Inspired by the literature on behavioral decision research and nudging, as well as by our prior work on regrettable Facebook behavior [28], we investigated the impact of Facebook privacy nudges. In this paper, we describe the design and evaluation of mechanisms that nudge Facebook users to consider more carefully the content and context of their online disclosures through visual cues and time delays. We developed a platform that enables us to deploy nudges and evaluate them with users in longitudinal field trials.

We conducted an exploratory 6-week field trial with 28 Facebook users. Our goal was to gain an understanding of how users perceive and interact with the features of our nudges. We analyzed both quantitative and qualitative data about participants' interactions with the nudges, the content of their posts, as well as their opinions collected through a final survey. We found that reminders about the audience of posts can prevent unintended disclosures without major burden; however, introducing a delay before publishing users' posts can be perceived as both beneficial and annoying. While many participants found the nudges helpful, others found them unnecessary or overly intrusive, suggesting that nudges may not be appropriate for everyone.

Our work makes two contributions. First, we developed an experimental platform that modifies Facebook's interface and collects users' behavioral data to operationalize and evaluate the concept of Facebook privacy nudges. Second, we identified key aspects worth considering when designing and evaluating a privacy nudging system.

CHI 2014, Apr 26 – May 01 2014, Toronto, ON, Canada
ACM 978-1-4503-2473-1/14/04.
http://dx.doi.org/10.1145/2556288.2557413

BACKGROUND AND RELATED WORK

Our work was inspired by scholarly research on problematic disclosures on social media, cognitive and behavioral biases, privacy decision making, and nudging.

Potential for Regret

Offline, people are naturally good at tailoring comments, gestures, and actions to specific audiences [14]. However, online (and in particular on social media such as Facebook), users tend to communicate with many groups (e.g., friends, co-workers) simultaneously, and as a result encounter difficulties in adapting messages for different audiences. Thus, shared content is often visible across groups, leading to a phenomenon called "context collapse" [21]. These issues are exacerbated by the fact that even experienced users have difficulties with Facebook privacy settings [3, 20]. Hence, a variety of dynamics lead to regrettable Facebook posts. Content is sometimes viewable by unintended audiences and users create posts "in the heat of the moment," which can lead to unintended disclosure and regret [28]. Unintended disclosures can lead to a range of consequences, including stalking, identity theft, blackmail [15], and reputation damage [9].

Biases, Heuristics, and Privacy Decision Making

Behavioral biases and heuristics can lead to systematic errors in decision making [26]. While, biases and heuristics have been studied in many contexts, here we focus on those that are closely related to our nudges. Bounded rationality forces individuals to rely on heuristics to simplify the choices available; however, sometimes choices with the best outcomes may be inadvertently discarded [22]. The economic effects of asymmetric information, which limits rational decision making, has been studied in the market of used cars [6]. Nearly half a century later, problems of bounded rationality and incomplete information are alive in social media.

Bounded rationality and asymmetric information prevent individuals from anticipating the audience for their posts. While it is easier for users to think in terms of broad audiences (e.g., friends, friends of friends, public), more granular groupings (e.g., parents, neighbors, church) can help mitigate unintended disclosures. Similarly, as Facebook comments inherit the audience of the original status update, it may be impossible for the person commenting to determine the audience of his or her comment—an example of asymmetric information. In fact, recent research by Bernstein et al. found that Facebook users "consistently underestimate their audience size for their posts, guessing that their audience is just 27% of its true size" [7]. As a result, Facebook users often post content that can be viewed by an unintended audience, which may lead to regret [28]. One of the nudges we present here attempts to mitigate problematic online disclosures associated with bounded rationality and asymmetric information.

Another relevant bias is known as hyperbolic time discounting, the fact that individuals use variable and inconsistent discount rates over time and often assign higher utility to present choices than to future ones [18]. For instance, people tend to procrastinate because they over-estimate the enjoyment of not doing work now and under-estimate the future consequences of delaying work [13]. In the privacy domain, Acquisti has shown that people often trade their personal information for immediate gratification [1]. The work on dual process theory is also relevant. For instance, Kahneman posits the existence of two processing systems in our brains: intuition (System I) and reasoning (System II). Intuition tends to be fast, automatic, and rely on heuristics, while reasoning is slower and involves more conscious judgement [17]. Our prior work on regrets found evidence of impulsive behavior, often driven by highly emotional states [28]. The second nudge we present was designed to mitigate problematic disclosures potentially due to hyperbolic time discounting and impulsive behavior.

Soft-paternalism and Privacy Nudges

Soft paternalistic interventions attempt to help individuals by mitigating behavioral biases (or, in some cases, exploiting them) to achieve the outcomes that better align with users' preferences. Thaler and Sunstein popularized the idea of nudging. They defined a nudge as "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives" [25]. Acquisti described the role of hyperbolic discounting and immediate gratification in the so-called privacy calculus [1]. He further proposed that nudges could be used to influence privacy decision-making in a manner that decreases users' regret [2].

Although not explicitly referred to as nudging, researchers in the fields of HCI and Persuasive Technology have explored mechanisms to assist users with privacy and security decision making. Forget et al. built a system to nudge users to create stronger passwords [12]. Ur et al. showed that certain password meter designs may encourage users to create stronger passwords [27]. Wilson et al. studied the effect of predetermined privacy profiles to assist users with location sharing disclosures [30]. Choe et al. investigated the impact of the "framing" heuristic (i.e., people would prefer alternatives that are framed as gains over those framed as loses, even when the two alternatives are equivalent) in the context of mobile apps selection; and found that the framing effect had minimal impact on participants' trust perceptions [10]. A recent longitudinal Facebook study highlighted how profoundly Facebook interface changes can impact users' information sharing — indicating a potential for Facebook privacy nudges [23].

Several mechanisms have also been proposed to improve privacy decision making for social media. For example, Fang et al. described a wizard that creates sharing categories automatically [11]. Lipford et al. examined interfaces for online social network privacy controls, comparing expandable grids to visual policies [19]. Besmer et al. built a tool that allows Facebook users to negotiate about photo tagging [8]. Unlike these mechanisms, our approach aims to proactively nudge users away from posting potentially regrettable content.

PRIVACY NUDGE DESIGN

Our study focused on two types of nudges: one that reminds users about the audience for their post, and one that encourages users to pause and think before posting. In line with

the concept of soft paternalism, neither of these nudges limits users' ability to disclose information, or affects the trade-offs associated with their disclosures. Instead, both nudges provide contextual clues intended to assist users in making better informed information-disclosure decisions. The nudge designs were improved based on two pilot studies conducted in April and July of 2012, respectively. We conducted a third pilot study in February 2013 to confirm that our experimental platform was stable enough to use in a field trial.

Audience Nudge

Inspired by the literature on bounded rationality and asymmetric information, the audience nudge aims to help Facebook users anticipate the audience for their posts.

This nudge initially included only the message: "these [number] people can see this post" based on the privacy setting of the post. However, after testing this feature in the first pilot study, we realized that participants did not find the intervention particularly useful, because the intervention message hardly varied over time — users often use the same privacy setting for their posts, and thus the message stayed the same.

In order to make the nudge more dynamic and salient, we considered using visual, rather than merely textual, cues. We designed a profile picture feature that displays five profile pictures, randomly selected from the pool of people who can view the post, based on that post's current privacy setting. This feature was inspired by Jenni and Loewenstein's work on the "identifiable victim effect," which finds individuals are willing to expend more resources on identifiable than unidentified victims [16]. In our context, the profile pictures provide some form of identifiability to both familiar contacts such as family members, close friends, and co-workers, as well as unfamiliar acquaintances or strangers, prompting Facebook users to think about who should see their posts.

We then combined both features (textual and visual information) into one "audience nudge," shown at the top of Figure 1. The nudge addresses two complementary aspects of audience perception: specific members of the audience, and the size of that audience. The pool of profile pictures and the audience size were determined by the privacy setting of the post. Our nudge implementation was able to detect and work with complex privacy settings of status updates such as "friends except certain people or groups" and "friends of friends." However, given the restrictions imposed by Facebook, the nudge cannot always precisely measure the audience size. In such cases, the message would provide qualitative rather than quantitative information, e.g., "These people, your friends, AND FRIENDS OF YOUR FRIENDS can see your post."

Timer Nudge

The timer nudge was inspired by literature on hyperbolic time discounting and dual process theory. It was designed to encourage users to pause and think (i.e., switch from System I to System II) before posting. It introduced a visual delay of 20 seconds after a user clicked the "post" button before publishing the submitted post. During the countdown, the user could cancel this post; otherwise the post was made automatically at the end of the countdown.

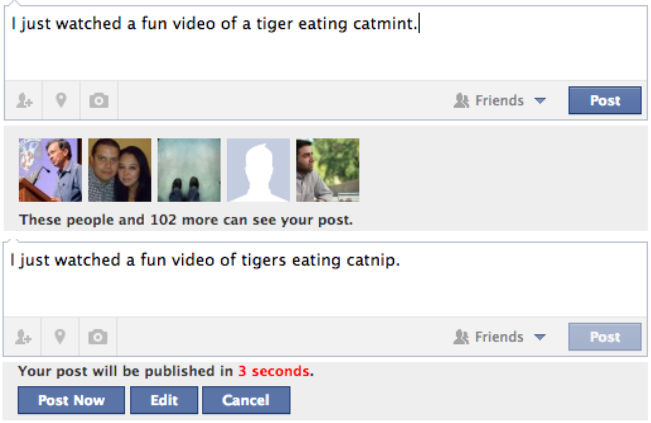

Figure 1. Audience+timer nudge. As the user types a post, five profile photos are displayed, selected randomly from people who will be able to see this post (top). After the user clicks "Post," a countdown timer appears and delays the post for ten seconds (bottom). During the count-down period users could edit or cancel the post, click a button to go ahead and post immediately, or do nothing and when the count-down expires the post will be made automatically.

In the first pilot, participants found this nudge interesting, but also suggested reducing the time delay as well as allowing users to bypass the delay or edit the post. Accordingly, we reduced the delay to 10 seconds and added three links: "post now," "edit," and "cancel."

"Audience+Timer" Nudge

In the second pilot study, we tested the audience and timer nudges. Participants of this pilot study found both the picture and timer nudges somewhat helpful. However, some participants did not realize that the *post now, edit* and *cancel* links could be clicked [29]. To address this issue, we changed the design of these links to buttons that mimicked the Facebook look-and-feel. Furthermore, to create an improved Facebook privacy nudge, we combined the best aspects of two nudges we previously tested during the pilot into a single, new "audience+timer" nudge (Figure 1).

We built an experimental platform to both implement the nudge on Facebook and monitor how users interact with it. The platform consisted of a Facebook application to access users' Facebook data and a Chrome browser plugin to insert the nudge interface seamlessly into the Facebook interface.

In the second pilot study, we encountered a few technical issues. First, our nudges were not always shown. Second, our system did not reliably log users' Facebook behavior and interactions with the nudges. We discovered that both issues were partly due to significant changes introduced by Facebook after deploying our system. The lack of reliable behavioral logs prevented us from doing any quantitative analysis of the nudges [29]. We fixed these issues and implemented reliable logging of system events (e.g., display the nudge UI) and user behavior (e.g., click cancel). We then tested the updated system in the third pilot study. We identified some minor technical issues (e.g., Facebook pages loaded more slowly than usual, and some participants were unable to comment on certain public pages), but otherwise the system appeared

stable and was able to log events reliably. We fixed these minor issues before the 6-week field trial. However, Facebook continued to make changes during our field trial and the comment problem reoccured.

STUDY METHODOLOGY

To evaluate the "audience+timer" nudge, we conducted a 6-week field trial with 28 Facebook users during April and May, 2013. We posted study ads on the Craigslist pages of the 12 most populated US cities as well as Syracuse and Pittsburgh. The ads directed prospective participants to a screening survey. The survey invited respondents to participate in the study if they met the following criteria: active adult US Facebook users who posted or commented at least once per day on average; native English speakers who posted in English and used Chrome, primarily, to access Facebook (because our platform was implemented for Chrome).

Study participants were required to install a Chrome plugin and an associated Facebook application. The 6-week field trial was divided into two phases. During the first three weeks (the *control period*) data collection took place without nudging interventions. At the end of the control period we asked participants to complete a mid-term survey to allow us to better understand participants' Facebook behavior during this period. We asked about unusual events, whether posts had caught the attention of unexpected audiences, and any regrets since the start of the study. During the remaining three weeks (the *treatment period*), in addition to data collection, all 28 participants were also shown our nudges. At the end of the treatment period, we asked participants to complete a final survey about their Facebook experiences during this period and their opinions of our nudges.

Participants were compensated with a $10 Amazon gift card for each week of participation and a $10 bonus for study completion. The study received IRB approvals at both Syracuse University and Carnegie Mellon University.

Since our nudges were highly dependent on the Facebook platform, we had to keep up with Facebook's frequent and unpredictable interface changes. To maintain our system during the 6-week study, we spent about an hour every day testing our system with all Facebook features it might interact with. We also had a programmer on standby to update the system if we found any issues. Despite these efforts, the system still encountered two technical problems during the six-week study. First, some participants were intermittently unable to post comments. This problem occurred most frequently for comments on public Facebook pages (e.g. for a company or celebrity). Second, participants experienced slow performance caused by our Chrome browser plugin.

RESULTS

In this section we present the study results. We first describe participants' demographics and their posting frequency during the study. We then report participants' interactions with our nudges and changes to privacy settings made while our nudges were active. Finally, we present a detailed participant-level analysis looking into whether and how our nudges impacted each of our study participants.

ID	Gender	Age	Days	Days	Status updates		Comments	
			Control	Treat.	Control	Treat.	Control	Treat.
P1	F	49	18	22	1	0	171	68
P2	F	31	18	22	2	9	14	17
P3	F	28	18	22	2	5	57	158
P4	M	27	20	20	35	16	69	16
P5	F	30	18	22	20	24	31	50
P6	M	31	18	22	223	244	396	176
P7	F	45	19	21	45	87	822	950
P8	F	44	19	21	17	17	45	55
P9	F	39	18	22	1	7	14	18
P10	F	20	18	22	32	47	363	386
P11	M	23	18	22	5	9	52	27
P12	M	36	18	22	3	0	235	53
P13	F	50	18	22	1	3	14	15
P14	F	20	18	22	22	41	60	21
P15	F	23	18	22	4	17	4	12
P16	F	32	18	22	7	16	65	102
P17	M	28	21	21	2	5	144	17
P18	F	45	21	21	13	6	22	2
P19	F	19	21	21	86	37	290	132
P20	M	51	21	21	3	6	45	34
P21	F	28	21	21	34	33	77	48
P22	F	23	21	21	1	2	14	4
P23	F	27	21	21	10	7	34	42
P24	F	26	19	22	10	6	26	8
P25	M	27	19	22	4	1	3	3
P26	M	21	19	22	4	10	7	31
P27	F	22	19	22	20	1	46	11
P28	M	49	19	22	30	1	126	1
	Min	19	18	20	1	0	3	1
	Max	51	21	22	223	244	822	950

Table 1. Summary of participants' demographics and number of status updates and comments made during the control and treatment periods.

Demographics and Facebook Posting Activities

Our participants included 19 females and 9 males between the ages of 19 and 51 (M=32, SD=10) from 16 U.S. states. All of them reported being active Facebook users, posting status updates or comments daily. Twelve (43%) self-reported having posted something on Facebook that they later regretted. Our participants came from a variety of occupations including medical staff, engineers, students, managers, teachers, homemakers, retired, and unemployed. Two had completed high school and the rest had at least some college education.

Table 1 summarizes participants' demographics, and the number of status updates and comments made during the control and treatment periods. We use a combination of one letter and one number to refer to each participant. As shown in Table 1, there is a large variability in the frequency of posting across participants. Overall, there is no obvious difference in posting frequency between the control and treatment periods. While our nudges might have impacted posting frequency, we cannot attribute those changes exclusively to the nudge. For example, participants explained that they had posted with unusual frequency, during both the control and treatment periods, due to factors such as vacations, illness, or new jobs. Thus, we do not use posting frequency to evaluate the impacts of our nudges.

Interactions with Nudges

There are many ways a participant could interact with our nudges. We focused on four types of interactions with our nudges: hovering over profile pictures displayed by the audience nudge, and clicking *post now*, *edit*, and *cancel* buttons displayed by the timer nudge.

Hovering Over Profile Pictures

When a user hovers over the five profile pictures displayed by the audience nudge, the corresponding Facebook user's profile name appears. Twenty-four out of 28 participants hovered over profile pictures at least 3 times, and half of them did that throughout the treatment period, suggesting that most participants saw the pictures and interacted with them. However, participants only rarely clicked *edit* or *cancel* or changed the privacy setting after hovering over a profile picture. This suggests that participants were interested in identifying the people shown in the profile pictures, but generally did not feel the need to exclude them from seeing their posts.

Clicking Post Now

In the timer nudge, users could either wait for the post to be submitted automatically after the 10-second delay, or click the *post now*, *edit*, or *cancel* buttons. Twenty-four and 26 participants clicked *post now* at least once for status updates and comments, respectively. Participants clicked *post now* more often for status updates (64%) than comments (25%).

Clicking Edit

Participants used the *edit* button less than *post now,* but when they clicked *edit*, they did it more for comments (4.7%) than for status updates (1.7%). Eighteen participants clicked *edit* for comments, while only five participants did that for status updates. Seven participants clicked *edit* only once, while another 11 participants clicked it at least three times.

We also found that participants used the *edit* button in different manners. Clicking *edit* did not necessarily result in a different post, since in many cases the final post was not modified, suggesting that participants were using the *edit* feature to stop the timer and review their posts. A handful of participants ended up canceling their posts after clicking *edit*. Some participants used the *edit* option to correct typos, slightly rephrase, or complement their posts with additional information, while others made major changes to their original posts.

Clicking Cancel

We logged only seven cancellations for status updates (1.0%) and 15 for comments (0.6%) from eight participants. In some cases, participants refrained from submitting their post altogether, while in other situations they started a new rephrased post. In a few cases, as detailed in the per-participant analysis that follows, participants seemed to cancel potentially sensitive posts.

Privacy Settings Changes

Inline settings allow Facebook users to specify the audience for their status updates. Facebook users can select from a set of predefined groups (e.g., only me, friends, friends of friends, and public), create groups (e.g., high school classmates, co-workers, neighbors), and customize the setting to include or exclude specific people or groups. The setting remains the default for future status updates until it is modified again. The privacy setting of a comment inherits the setting from the corresponding status update. We expected that our nudges would help participants to more carefully select privacy settings for their status updates.

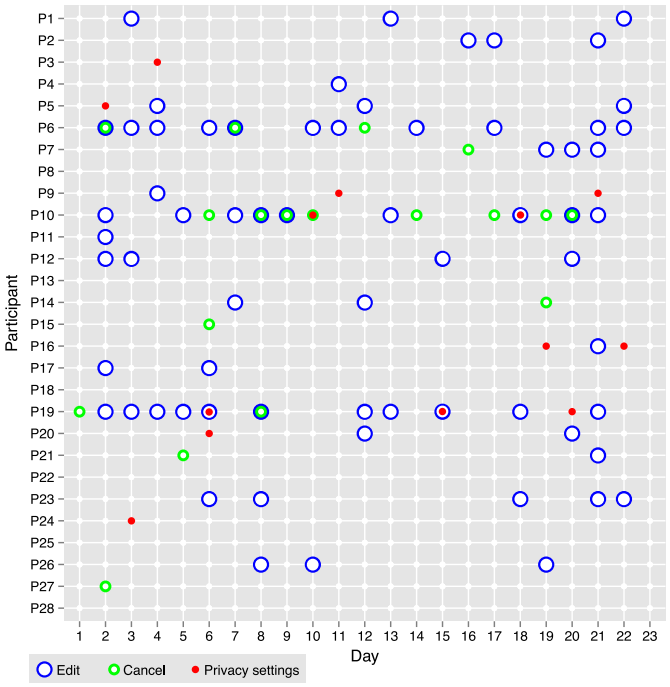

Figure 2. The days on which each participant clicked edit or cancel, or changed their privacy settings during the treatment period. The blue and green circles denote a participant who clicked edit or cancel at least once on that day, respectively. The red dot denotes a participant who has changed the inline privacy setting at least once on that day.

Six and eight participants modified their inline privacy settings during the control (number of changes per user: M=.6, SD=1.8) and treatment (M=.5, SD=.9) periods, respectively. Some of them changed the privacy settings both during the control and treatment periods, suggesting that our nudges were not necessarily associated with those changes. However, four participants changed their inline privacy settings only during the treatment period. In one of these cases a participant made his privacy settings more restrictive after hovering over a profile picture.

Interactions Over Time

To investigate novelty and habituation effects, we analyzed the temporal distribution of interactions with our nudges. Figure 2 shows the days on which participants clicked the edit and cancel buttons or made changes to their privacy settings during the treatment period. While there were six participants who did not exhibit any of these interactions and four who only interacted in the first few days, the majority of participants paid attention to and interacted with our nudges throughout the treatment period.

Participant-Level Analysis

To investigate the impact of our nudges on each participant, we analyzed each participant's interactions with our nudges as well as their survey responses. This allows us to understand why some participants liked the nudges and found them useful while others did not, and also helps us tease apart the impact of the audience and timer nudges. We categorized each participant into one of five descriptive groups defined by two dimensions: participants' attitudes and participants' level of interaction with our nudges.

Frequent Interactions and Positive Attitude

This group includes four participants (P4, P10, P20 and P23) who made extensive use of our nudges, and believed that at least one of the nudges could be helpful for themselves or others. These interactions include clicking the *cancel* or *edit* buttons and hovering over profile pictures.

P4 said, "I didn't post more or less, but I did post more cautiously. The constant reminder of who would be seeing my post was kind of an eye opener." He reported having used the time delay "to correct grammatical errors or statuses that looked 'off.'" We found that while he often clicked the *post now* button, for a few posts he waited several seconds before clicking it, yet for others did not click it. He clicked *edit* for several comments, for example, he changed "long out" for "sign out" when writing about logging off of Facebook. He also hovered over pictures for many posts. For instance, after looking at the pictures, he checked the privacy setting without any change before posting, "I just started selling gold in the RMAH...I got a few auctions sold beforehand... :(" He told us that the nudges made him "a bit more aware. Especially the first day. That was almost the 'Oh wow' moment when I realized that more people could see my posts than I thought about." He also suggested that the nudges "should be default for all users."

P10 found the time delay helpful because it helped her avoid "getting into fights on Facebook because you have to stop and think." Despite experiencing some technical issues, her overall opinion was positive. She summarized, "I generally made better decisions." She normally clicked *post now* within 3 seconds, but later on she waited longer for a few of her posts. Besides, she clicked *cancel* for a few posts and then posted edited versions. For example, she canceled the status update, "not excited about still being sick wtf" and then posted, "not excited about still being sick after spending all afternoon in bed not doing my paper or having fun."

P20 canceled and edited a few posts. For example, he clicked *edit* and then ended up not posting: "Traded your Z in for that? :P." He also changed the privacy setting from public to only me after hovering over the profiles pictures shown when posting, "I've been nice up to this point, but the guy has to go! Eating all the bird seed. Where's my bebe gun?" In the final survey he said, "I think I was careful of what I said."

P23 found both nudges helpful. She explained, "I did like knowing when posts were going to be made public (like if a friend's wall is not protected to only their friends, etc.)" and elaborated, "I was going to respond to something snarky... I cancelled it because the application informed me that the entire internet could see my post." She also found the timer nudge helpful when she posted on other people's walls because it prevented her from "entering a discussion ... on someone's wall who posts religious or other annoying stuff all the time." However, she said she "was annoyed when I was using it to post to my own wall or to my close friends wall - those are not when I need the reminder." This suggests she might have preferred a nudge she could customize to fit her needs. Despite some technical glitches that prevented her from posting some comments, P23 had a positive opinion of the nudges overall because "It made me think twice about what I posted and who might see it." Our log data also show that she often hovered over the audience profile pictures and used the *edit* button to make minor changes to her posts.

Limited Interactions and Positive Attitude

Ten participants who had few interactions with our nudges stated that they thought at least one of the nudges could be useful for them or others (P8, P9, P11, P12, P15, P18, P21, P24, P25, and P26). For instance, P8 did not consider our nudges helpful to her, but said they could benefit "young people who are more likely to fly off the handle." She explained, "I didn't benefit, but trouble makers and kids would since it's an extra step and not just post and go-it may make someone think twice before posting hurtful comments." This participant used the friends privacy setting for all her posts and often clicked *post now* within 2 seconds. While she hovered over some profile pictures, she did not perceive any benefit from them since, "just wanna hit post and be done, not mess around with the delay or figure out who may or may not see it since I have my privacy settings the way I want them." Nonetheless, she did edit some of her posts when she caught spelling or grammar errors.

While our logs show that P9 hovered over profile pictures, she reported not seeing any profile pictures. She based her opinion on the timer nudge and said she did not benefit much, but "it would be good for someone with a short temper." She said the nudge might be useful, "if I'd made a spelling error or tagged the wrong person." She clicked *post now* within 3 seconds on all her status updates and mostly used the friends privacy setting. In one case where she changed her privacy setting, she selected her Farmville friends group and explicitly excluded two friends before posting a Farmville-related request. She once clicked *edit* for a comment, but ended up reposting the same comment.

P11 reported that the profile pictures "helped me shape some of my posts." He further reported having canceled some posts because "I didn't want other people to think I'm stupid." However, he also expressed that, "the countdown timer annoys me a bit." He clicked *post now* within 2 seconds for all his status updates. Overall, he felt the nudge "made me think about what I was going to say."

P12 did not post any status updates during the treatment period and he used the public privacy setting for all his status updates during the control period. He clicked *post now* for most of his comments within 3 seconds and did not hover over pictures or click the *cancel* button. However, he clicked *edit*

five times for his comments, two of which he ended up not posting, and three of which were reposted without changes. He concluded, the nudges are "not really for me, but I can see how it could be useful for others."

P24 found the audience nudge useful: "I think seeing all of the profile pictures made me rethink what I was going to post if it was slightly offensive or using curse words." However, the only interactions she had with ours nudges were hovering over pictures a few times. In the treatment period, she did change her privacy setting to exclude three friends when she posted, "So I bought a Rick Pitino Makers bottle for $50 and turned around and sold it for $180, lol..." Overall, she thought the nudge "made very little impact."

Limited Interactions and Negative Attitude

Three participants (P7, P14, and P22) neither interacted with our nudges, nor liked them. P7 was an active user who made over a thousand posts during the treatment period. She thought clicking *post now* was necessary to send her posts, which frustrated her. She explained, "I found it to be a pain because some of my posts I just like to post and go." She often clicked *post now* within a few seconds. Furthermore, while she hovered over profile pictures for about one third of her status updates, it did not seem to have any effect on what she posted as most of her status updates were public posts related to products, coupons, and promotions.

P14 also disliked the nudges. She did not click any buttons or hover over profile pictures when posting status updates. She clicked *post now* for three comments, and all were within three seconds. She also canceled one comment because "I was impatient!!" She explained, "I could see how the timer and the profile pictures would be beneficial but I just thought it was annoying."

P22 also did not like the nudges or interact with them. She only made six posts during the treatment period. She remembered seeing our nudges, but said that she did not pay attention to them. Our logs confirmed that she had no interaction with our nudges. When asked if there was any situation where the nudge negatively affected her, she replied, "for me, I don't care, so every time." She also seemed to post on Facebook just to get paid for participating in our study, posting the comment: "i'm only commenting on stuff b/c im being paid to by some app to spy on me and if i dont do enough social stuff they'll stop letting me do it."

Frequent Interactions and Negative Attitude

Seven participants did not like our nudges, but had extensive interactions with them (P2, P3, P6, P16, P17, P19, and P27). P2 experienced technical problems with our nudges and strongly disliked the timer. She complained, "the delayed posting thing which I HATE... makes posting harder." She believed that the timer is "not needed" and she did not "need anyone editing or censoring me." Although she indicated that she preferred the profile pictures over timer, she based her negative opinions on the timer. She explained, "I try to not put anything too embarrassing or horrible. I didn't mind that you were watching or anyone. I say what I want or feel," adding that, "there is no way to protect people from posting embarrassing or life impacting information online while mad or upset or whatever. It's human nature to be stupid sometimes." She often clicked the *post now* button when commenting, sometimes waiting several seconds before doing it and sometimes clicking it right away. For example, she waited 3 seconds before clicking *post now* for the comment: "David's burrito defeated him. It was HUGE," but waited 8 seconds for: "Off subject but the worst my arm pits have felt was during laser hair removal...Most painful area so far." She also clicked *edit* a few times without changing any comments, suggesting that she was using the *edit* option to take a second look at her comments. She also clicked *edit* to change one of her comments from "Or its like saying it transcends life. Negative Nelly aka Autumz..." to "Or its like saying it transcends life. I guess we all know your glass is half empty."

P6 hovered over profile pictures and clicked the *cancel* and *edit* buttons a few times. In addition, he often clicked *post now*, sometimes waiting a few seconds and sometimes clicking it right away. An active user, he made more than 400 posts during the treatment period. He thought that the plugin "made posting slightly more frustrating, but did not affect output." He explained that he was annoyed by "the fact I had to keep re-confirming that I wanted to post something I was sure about posting." This was another participant who thought that clicking *post now* was necessary to send his posts and did not realize that after the 10 seconds delay, his posts would be automatically posted.

P19 was another participant who frequently interacted with the nudges but did not like them. She said, "I didn't care about this feature at all, it did not affect my facebook usage at all, I just ignored it, like most people would." She also complained about encountering technical issues with the timer nudge. Interestingly, she acknowledged that she often regrets her posts, but said she solves the problem by just deleting them: "I almost always post things that I wish I wouldn't have, then I just delete them and the problem is solved." But, in some cases damage may have already been done before deleting a post. P19 did not seem to be concerned about this, and thus did not find utility in the nudges. While she disliked the nudge idea, she hovered over the audience nudge profile pictures 10 times, clicked *cancel* for a status update and a comment, clicked *edit* for 21 comments, and made two privacy setting changes during the treatment period. Most of her edits were minor rewording of the posts. For one lengthy status update about love and betrayal she changed the privacy setting to exclude one particular friend before she started typing her post, suggesting that she had a clear idea of her intended audience before she posted. Overall, the nudges did not seem to help her avoid making potentially controversial posts, but did give her a chance to make minor edits of her posts.

Indifferent

Four participants expressed indifference about the nudges (P1, P5, P13, and P28). They either did not receive enough exposure to the nudges to form an informed opinion or simply expressed a neutral opinion even after having interacted with the nudges. Lack of exposure was due to participants posting less frequently or using different browsers or devices to post.

P1 reported having seen profile pictures and using the time delay to review her posts. She did not post any status update during the treatment period and often clicked *post now* within 3 seconds for her comments. She hovered over pictures before making several comments and clicked *edit* twice to correct typos. She was among those participants who had problems posting some of their comments which likely affected her overall opinion of the nudges. Although she used the nudges while posting and mentioned that the nudge could be useful "in case you are commenting on the wrong post," she was also not impressed and "could take it or leave it."

P5 neither expressed a positive nor a negative opinion and her behavioral data did not show any relevant interactions with the nudge as she only clicked *post now* right away. Similarly, P13 often clicked *post now* right away and did not have other interaction with the nudges. She explained, "In most cases I changed the time to posting to now so I didn't have to wait, I also didn't have to edit my posts because I wasn't saying anything I didn't want anyone to see." P28 hardly noticed the nudge as he only posted one status update and one comment using the plugin during the treatment period.

DISCUSSION

The goal of our nudging interventions was to help users be more thoughtful when posting on Facebook, in order to reduce the potential for posting status updates or comments they might later regret. Consistent with the tenets of soft paternalism, we designed our nudges to encourage users to think about the audience and content of their posts without limiting their ability to post on Facebook. Here, we discuss participants' perceptions of our nudges and the challenges of conducting longitudinal field trials. We also offer recommendations for designing and testing nudges.

Varied Perceptions

Participants varied in how they perceived the general nudging idea and the two specific nudge features (audience and timer). Some participants were positive about the nudging intervention, e.g., P26 said "i like the plugin and think it is a great idea, i would love using this as the final product." In contrast, others had negative opinions. For example, P2 disliked the timer nudge because she did not "need anyone editing or censoring me."

We found that how participants use Facebook often played a role in their perceptions of our nudges. Generally, we found that those who use Facebook to post personal thoughts perceived the nudges as more beneficial than those who use it to broadcast news articles and other public information. On the other hand, those who use it for commercial or money making purposes (e.g. to share information about products and coupons) had negative opinions. We also found that those participants who had prior experience adjusting privacy settings and seemed to be careful about what they posted recognized the benefits of the nudges, but believed they did not need them.

Nudging toward Audience Awareness

The profile picture feature of the nudge was designed to remind Facebook users of the prospective audience for their posts. We found that most participants paid attention to and interacted with the profile pictures and several valued this feature, stating that it made them think about whether there post might offend someone.

Profile pictures were accompanied by a an indication of the number of people who could potentially see a post. Some participants said they found this information helpful, especially when posting comments on friends' posts.

In this study, we bundled the profile pictures and audience size information together. Further work is needed to determine their effectiveness in isolation.

None of the participants complained about the profile pictures, as they were less intrusive than the countdown timer. Users can ignore them, as some of our participants did, and go about their posting as usual. We found that most participants hovered over the profile pictures, and anecdotes from the final survey suggest that some of the participants benefited from having seen profile pictures. For example, one participant reported having decided not to post something after seeing the profile pictures. We also observed a participant change to a more restrictive privacy setting after hovering over a profile picture. This suggests that profile pictures can assist users in making better privacy decisions, but sometimes their effect can be subtle or difficult to measure.

Nudging with A Countdown Timer

The countdown timer was designed to encourage participants to stop and reflect on the content of their posts in order to avoid regrettable, "heat of the moment" posts. Participants were quite divided in their views. Some participants found the countdown timer valuable for giving them a chance to review their posts "a little more carefully" and "catch misspelled words or grammar errors." On the other hand, some participants voiced frustration about what they perceived as a requirement to take an extra step or wait 10 seconds to complete their posts.

We observed that the nudge was successful in helping some participants reconsider their posts. It had an additional benefit of helping users catch typos and other minor errors in their posts. A number of participants rephrased or even canceled their posts during the delay. However, this benefit came at the cost of delaying every post. The timer countdown was both the most liked and disliked nudge feature we implemented.

For the most part, participants were not as concerned about their posts being delayed for 10 seconds as they were about having to wait 10 seconds before they could move on to something else. In reality there was no need for them to watch the countdown, but some participants seemed unwilling to trust that their post would get posted after 10 seconds and others thought they *had* to click *post now*. Our timer nudge seemed to leave some participants feeling uneasy and afraid to move on until their massage got posted. We could design an interface where the post would appear on the user's screen as if it had been posted (perhaps by posting it to "only me") while the timer is counting down. Even though the nudge would still have the same functionality, a change in the visual display might make it more acceptable.

The idea of delays may be applied in other scenarios where people may benefit from some extra time to think about their actions. However, since the time delay interrupts primary tasks (in this case Facebook posting), it should be used selectively and with caution. Future research should explore other ways to slow people down and encourage them to think about their actions, as well as ways to introduce time delays more selectively.

Challenges and Limitations

Conducting our investigation as a longitudinal field trial allowed us to investigate the impact of our nudges under real life conditions, but that made our study more challenging to run and resulted in a number of limitations.

One source of limitations stemmed from implementation challenges. Since Facebook changes its interface frequently and unpredictably, we had to constantly monitor, test, and adapt our code to keep up with those changes. Despite this concerted effort, our nudges malfunctioned a few times during our field trials. For instance, some participants experienced some of their comments not being posted. In addition, since our nudge was not an integral part of the Facebook platform, we had to work around the Facebook UI to embed the nudge features. We also had to add logging functionality to capture all possible user-driven events. However, this extra logging slowed down our nudges and indirectly affected some participants' Facebook experience. These technical challenges made it difficult to run our field trial for an extended period of time. In addition, some participants disliked our nudges primarily because of these issues.

As with any research involving observation of participant behavior, one methodological concern is the Hawthorn effect: participants may change their behavior simply because they are in a study. To mitigate this we minimized our interactions with participants once the study began. In the mid-term survey, we explicitly asked them whether their posting behavior had changed. Some participants noted posting with different frequencies due to various reasons, but un-related to our study. Several participants also reported in the final survey that they thought our nudges were introduced by Facebook rather than us.

External factors beyond our control and observation likely affected participants' posting behavior, making it difficult to determine causality. Similarly, measuring the effectiveness of our nudges in preventing regret is also challenging because generally only a small fraction of the posts made by users may lead to regret, and arguably even fewer lead to the short-term regret we may detect in this study. 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.

While our combination of different nudge features might increase the chance that we detect some effect of our nudges, it makes it difficult for us to isolate the effects of individual nudge features or account for interactions between features.

Implications for Designing and Testing Privacy Nudges

We identified a set of key aspects to consider when designing and evaluating privacy nudges. First, designers should consider the intrusiveness of the nudges. We found that our less-intrusive audience nudge was better received by users than our more-intrusive timer nudge. On the other hand, we observed more direct benefits from the timer nudge. It would be useful to investigate whether the timer nudge could be improved by making it less visually intrusive — for example, showing a user's post actually posted but visible to "only me" until the end of the countdown.

Second, designers should keep in mind that some users will dislike the sense of being watched. Designers should look for ways to nudge people without making them feel that a new "big brother" is watching. In our second pilot study users disliked being judged by a nudge that provided subjective feedback on the sentiment of the users' posts [29]. The nudges we tested in this study, on the other hand, were not perceived as judgmental by participants.

Third, designers should consider the extent to which they should allow users to control or customize a nudge. In our system, we did not give users any ability to control the nudges except for the *post now* button that allowed them to skip the time delay. Some users wanted to be able to turn off the nudges or personalize them according to their needs and preferences. Controls could allow users to configure nudges such that they are enabled only under certain circumstances, such as at specific times, when certain people can see their posts, or when they type certain sensitive words.

Fourth, it is critical that the nudges function properly and do not interfere with the usability and reliability of the system in which they are embedded. Our nudges suffered from technical glitches that decreased their perceived value for some participants. However, without help from Facebook, we found it difficult to improve the reliability of our system.

Lastly, but importantly, nudges are difficult to evaluate both quantitatively and precisely when they are designed to impact behaviors that may occur only occasionally, or that may be hard to observe. And yet, when it comes to privacy, it could be precisely occasional, rare behaviors that end up causing the most damage — for example, a spur-of-the-moment status update that leaves a long and painful trail of unintended consequences. Collecting enough measurable, quantitative data to compute aggregate results from a small sample of users is difficult under such circumstances. Unless it is feasible to study a large number of users, an evaluation strategy including qualitative participant-level analysis is likely to provide more informative results than a quantitative analysis.

CONCLUSIONS

While the field study we presented in this paper should be considered exploratory, our results suggest that privacy nudges have the potential to be a powerful mechanism to assist users in avoiding unintended disclosures. Although our findings come from a Facebook case study, the principles underlying the privacy nudges we tested may be extended to similar services such as Twitter or to other types of services

such as e-commerce, location sharing, and smart phone applications.

ACKNOWLEDGEMENTS

We thank Jeff Dyer, Colleen Eagan, Emily Forney, Eric Balebako, Yao Li, and members of the Privacy Nudge group at Carnegie Mellon for their help; CHI reviewers for their feedback; and our participants for their insights. This paper is based upon work supported by the IWT SBO Project on Security and Privacy for Online Social Networks (SPION), the NSF Grant CNS-1012763 (Nudging Users Towards Privacy), and a Google Focused Research Award on Privacy Nudges.

REFERENCES

1. A. Acquisti. Privacy in electronic commerce and the economics of immediate gratification. In *Electronic commerce*, pages 21–29. ACM, 2004.
2. A. Acquisti. Nudging privacy: The behavioral economics of personal information. *IEEE Security and Privacy*, 7(6):82–85, 2009.
3. A. Acquisti and R. Gross. Imagined communities: Awareness, information sharing, and privacy on the facebook. In *PETS*, pages 36–58. Springer, 2006.
4. A. Acquisti and J. Grossklags. Privacy and rationality in individual decision making. *Security & Privacy, IEEE*, 3(1):26–33, 2005.
5. I. Adjerid, A. Acquisti, L. Brandimarte, and G. Loewenstein. Sleights of privacy: Framing, disclosures, and the limits of transparency. In *SOUPS*, pages 9:1–9:11. ACM, 2013.
6. G. A. Akerlof. The market for "lemons": Quality uncertainty and the market mechanism. *The quarterly journal of economics*, pages 488–500, 1970.
7. M. S. Bernstein, E. Bakshy, M. Burke, and B. Karrer. Quantifying the invisible audience in social networks. In *SIGCHI*, pages 21–30. ACM, 2013.
8. A. Besmer and H. Lipford. Tagged photos: concerns, perceptions, and protections. In *Proc. CHI Ext. Abs.*, pages 4585–4590, 2009.
9. d. boyd and N. Ellison. Social network sites: Definition, history, and scholarship. *J. of Computer-Mediated Commun.*, 13(1), 2007.
10. E. Choe, J. Jung, B. Lee, and K. Fisher. Nudging people away from privacy-invasive mobile apps through visual framing. In *INTERACT*, pages 74–91. Springer Berlin Heidelberg, 2013.
11. L. Fang and K. LeFevre. Privacy wizards for social networking sites. In *WWW*, pages 351–360. ACM, 2010.
12. A. Forget, S. Chiasson, P. C. van Oorschot, and R. Biddle. Improving text passwords through persuasion. In *SOUPS*, page 112. ACM, 2008.
13. S. Frederick, G. Loewenstein, and O. T. Time discounting and time preference: A critical review. *J. of Econ. Lit.*, 40(2):351 – 401, 2002.
14. E. Goffman. The presentation of self in everyday life. 1959. *Garden City, NY*, 2002.
15. R. Gross and A. Acquisti. Information revelation and privacy in online social networks. In *WPES*, pages 71–80, 2005.
16. K. E. Jenni and G. Loewenstein. Explaining the identifiable victim effect. *Journal of Risk and Uncertainty*, 14(3):235–257, 1997.
17. D. Kahneman. A perspective on judgment and choice: Mapping bounded rationality. *American psychologist*, pages 697–720, 2003.
18. D. Laibson. Golden eggs and hyperbolic discounting. *The Quarterly J. of Econ.*, 112(2):443–478, May 1997.
19. H. R. Lipford, J. Watson, M. Whitney, N. Carolina, H. Lipford, K. Froiland, and R. W. Reeder. Visual vs . Compact : A Comparison of Privacy Policy Interfaces. *Interfaces*, pages 1111–1114, 2010.
20. M. Madejski, M. Johnson, and S. M. Bellovin. The failure of online social network privacy settings. Technical report, Columbia University, 2011.
21. A. Marwick and d. boyd. I tweet honestly, i tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1):114, 2011.
22. H. A. Simon. A behavioral model of rational choice. *The Quarterly J. of Econ.*, 69(1):99–118, Feb. 1955.
23. F. Stutzman, R. Gross, and A. Acquisti. Silent listeners: The evolution of privacy and disclosure on facebook. *J. of Privacy and Confidentiality*, 4(2), Mar. 2013.
24. R. H. Thaler and C. R. Sunstein. Libertarian paternalism. *Am. Econ. Rev.*, 93(2):175–179, 2003.
25. R. H. Thaler and C. R. Sunstein. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press, 1 edition, Apr. 2008.
26. A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131, 1974.
27. B. Ur, P. Kelley, S. Komanduri, J. Lee, M. Maass, M. Mazurek, T. Passaro, R. Shay, T. Vidas, L. Bauer, N. Christin, and L. Cranor. How does your password measure up? the effect of strength meters on password creation. In *Proc. USENIX Security*, 2012.
28. Y. Wang, S. Komanduri, P. G. Leon, G. Norcie, A. Acquisti, and L. F. Cranor. "I regretted the minute I pressed share": A qualitative study of regrets on facebook. In *SOUPS*, 2011.
29. Y. Wang, P. G. Leon, K. Scott, X. Chen, A. Acquisti, L. F. Cranor, and N. Sadeh. Privacy nudges for social media: An exploratory facebook study. *PSOSM*, 2013.
30. S. Wilson, J. Cranshaw, N. Sadeh, A. Acquisti, L. F. Cranor, J. Springfield, S. Y. Jeong, and A. Balasubramanian. Privacy manipulation and acclimation in a location sharing application. In *UbiComp*, pages 549–558. ACM, 2013.