Decaying Photos for Enhanced Privacy: User Perceptions Towards Temporal Redactions and 'Trusted' Platforms

SABID BIN HABIB PIAS, Department of Computer Science, Indiana University, USA
IMTIAZ AHMAD, Department of Computer Science, Indiana University, USA
TASLIMA AKTER, Department of Computer Science, Indiana University, USA
APU KAPADIA, Department of Computer Science, Indiana University, USA
ADAM J. LEE, Department of Computer Science, University of Pittsburgh, USA

With the rising popularity of photo sharing in online social media, interpersonal privacy violations, where one person violates the privacy of another, have become an increasing concern. Although applying image obfuscations can be a useful tool for improving privacy when sharing photos, prior studies have found these obfuscation techniques adversely affect viewers' satisfaction. On the other hand, ephemeral photos, popularized by apps such as Snapchat, allow viewers to see the entire photo, which then disappears shortly thereafter to protect privacy. However, people often use workarounds to save these photos before deletion. In this work, we study people's sharing preferences with two proposed 'temporal redactions', which combines ephemerality with redactions to allow viewers to see the entire image, yet make these images safe for longer storage through a gradual or delayed application of redaction on the sensitive portions of the photo. We conducted an online experiment (N=385) to study people's sharing behaviors in different contexts and under different levels of assurance provided by the viewer's platform (e.g., guaranteeing temporal redactions are applied through the use of 'trusted hardware'). Our findings suggest that the proposed temporal redaction mechanisms are often preferred over existing methods. On the other hand, more efforts are needed to convey the benefits of trusted hardware to users, as no significant differences were observed in attitudes towards 'trusted hardware' on viewers' devices.

CCS Concepts: • **Security and privacy → Human and societal aspects of security and privacy**.

Additional Key Words and Phrases: Privacy, Image Obfuscation, Photo Sharing, Trusted Hardware

ACM Reference Format:
Sabid Bin Habib Pias, Imtiaz Ahmad, Taslima Akter, Apu Kapadia, and Adam J. Lee. 2022. Decaying Photos for Enhanced Privacy: User Perceptions Towards Temporal Redactions and 'Trusted' Platforms. *PACM on Human-Computer Interaction* 6, CSCW2, Article 437 (November 2022), 30 pages. https://doi.org/10.1145/3555538

1 INTRODUCTION

Photos have become one of the most dominant media on online social platforms [1, 81]. Posting personal photos is a way of sharing 'small moments' with close social contacts [13], engaging in impression management [82], or increasing social interaction [63, 71]. However, online photo sharing also raises several privacy concerns. In the context of interpersonal privacy concerns,

Authors' addresses: Sabid Bin Habib Pias, Department of Computer Science, Indiana University, USA; Imtiaz Ahmad, Department of Computer Science, Indiana University, USA; Taslima Akter, Department of Computer Science, Indiana University, USA; Apu Kapadia, Department of Computer Science, Indiana University, USA; Adam J. Lee, Department of Computer Science, University of Pittsburgh, USA.

2573-0142/2022/11-ART437 $15.00
https://doi.org/10.1145/3555538

where people violate the privacy of others [90], embarrassing moments can go viral online through rapid re-sharing of photos in social networks [37], thereby demeaning people's social impressions through rumors and gossip [27]. Such publicly shared photos can unintentionally include personally identifiable data [62] leading to security and privacy violations. Sometimes, the photos shared online can also raise privacy concerns for bystanders (those people who are incidentally captured in a photo) [68, 76].

Privacy measures can be taken in the form of obscuring a part of the photo before sharing the photo [36, 38]. In addition to strengthening a sense of social connection, image sharing on social media is often used to maintain or improve one's social standing [25, 65, 71]. Hence, it is important to preserve the satisfaction and experience of the viewers while obscuring a region in the photo [37]. Prior studies have found that strong obfuscation techniques such as 'masking' can lower viewers' satisfaction [37] whereas some 'beautification' on the sensitive region may preserve some of the viewers' satisfaction. On the other hand, applications such as Snapchat[1] provide a novel layer of protection by allowing people to share 'ephemeral' photos, which disappear soon after the viewer has opened the photo. Ephemeral photos let users share 'personal' moments, free of redactions, with their peers and convey 'self-expression' even more than the existing persistent versions of photo sharing [5, 21]. However, viewers can save these photos through workarounds [51], thereby putting the sender's privacy at increased risk. In this work, we consider a combination of the two approaches: 'temporal redactions', where redactions gradually appear, seek to allow viewers to see the entire photo, but then redact sensitive parts of the image for safer longer term storage. Although photo sharing in the context of redactions has been explored recently to address privacy threats, more work is needed to understand people's photo sharing experiences and interpersonal privacy concerns in the context of such 'temporal' redactions. In particular, we explore two types of temporal redactions, delayed and gradual applications of redactions, with respect to people's interpersonal privacy concerns, and whether such redactions are preferred over current approaches. Prior studies have shown that applying a redaction on a photo may decrease the aesthetic appeal of the photo to viewers. We explore if temporal redactions are preferred by users as they may help users preserve the aesthetic of the photo because redactions are not applied initially, and in addition, protect users' privacy by applying automated redactions to sensitive regions after, say, 24 hours of posting the photo.

Regardless of the privacy-preserving techniques used to obscure the sensitive regions of the photos, an interpersonal privacy threat arises from the lack of control over the viewer's actions and their control over the viewing platform. For example, sharing a selfie with a friend can turn into an interpersonal privacy violation if the friend reshares this photo without permission. Prior studies have shown that the viewer can save the media file when it's visible on the application [6, 19, 51]. One suggested approach is to enforce restrictions through the use of trusted hardware [29, 73] and Trusted Execution Environments (TEE) [77]. The TEE is a secured enclave that is isolated from the device's main operating system (OS) and protects its contents from software attacks [74]. When a viewer's device is equipped with a TEE, the sender of the photo can rely on the TEE on the viewer's device to prevent certain actions, e.g., preventing the viewer from taking a screenshot. With a TEE, the photo can be delivered securely to the trusted TEE on the viewer's device using encryption [24], and the sender has the assurance that the OS cannot be used to save or take a screenshot of the photo. Although TEEs are potentially a powerful, technical tool for privacy-enhanced photo sharing, it is not clear whether people will understand the additional protection they provide and trust viewers' devices more; even with a TEE, people may not change their photo sharing behaviors.

In this work we explore the following research questions in this study:

[1]https://www.snapchat.com/

RQ1: How do various temporal redaction techniques affect people's sharing likelihood for different types of sensitive content? How do these preferences change in the context of the target audience?

RQ2: Does the trustworthiness of the viewing platform affect people's photo sharing likelihood? Is there any effect of the trustworthiness of the viewing platform on the choice of temporal redaction or the content type in the photo?

We conducted an online, survey-based user experiment (N=385) to explore these research questions. In the survey, we studied people's likelihood of sharing photos that contain sensitive contents with several temporal redaction techniques to obfuscate those particular contents of the corresponding photos. We conducted a between-subjects survey with random assignment based on whether the device of the audience of these photos includes a Trusted Execution Environment (TEE) or not. Each of these two surveys had three within-subject temporal redactions (instant, delayed, and gradual) applied on the same photo along with two base conditions (sharing the original photo and sharing with ephemerality as in Snapchat, i.e., the photo will be deleted after some time). We asked the participants about their likelihood of sharing photos with these redaction techniques with four different target audiences – family, friends, colleagues, and other followers. Each photo contained one type of sensitive content from the following: personally identifiable data, a facial expression with an entertaining filter, and a bystander. All the photos were selected and refined based on feedback from pilot studies.

Our findings indicate that although participants continue to be comfortable with existing approaches, several participants (17–22% depending on the context) showed a preference for the delayed and gradual temporal application of redactions in different circumstances. Future studies should evaluate the sharing behavior of participants when they are provided with more control in the temporal application of redactions. We did not find a significant difference in participant behavior in terms of sharing photos on trusted and non-trusted viewing platforms. Although Trusted Execution Environments (TEE) can be a powerful tool to provide security and privacy assurances, future studies should focus on conveying the utility of TEEs to non-technical users.

2 RELATED WORK

Various approaches have been explored to find an optimal mechanism to protect privacy while sharing photos on social media. Researchers have looked into several privacy-preserving techniques while sharing information such as controlling the content or the recipient [20]. We will discuss privacy-preserving mechanisms in terms of content and recipient control.

2.1 Content control

Studies have shown that users' privacy perception in photo sharing behavior is affected by whether the photo contains location information, private information, or impression violating content [43, 44, 59]. One way to protect these contents from being revealed is to obscure sensitive regions of a photo [36, 60]. Prior studies have found that differing customized filters for obscuring the sensitive region based on the content type can be useful in terms of privacy protection and photo utility preservation [36]. Using cartoon stickers to hide unintended faces may also be another medium of obfuscation [58]. Yuan et al. built an iOS application ProShare which applied JPEG scramble on some parts of the photo in a client device [95].

Various approaches have been developed to address the privacy of the people depicted in the photo. Facial recognition can be used to detect people in the photo and ask for their permission before posting the photo online [91]. Some applications such as Face-Off [2] pixelate faces in an

[2]https://apps.apple.com/us/app/face-off/id1317161001

automated fashion when the viewer of the photo does not have permission to see those faces [45]. A privacy protection mechanism named "Cardea" can detect users' privacy preferences based on location, scene, other's presence, and hand gestures, and eventually blurs out parts of the photo based on user preferences [79]. The popular online platform Youtube also protects publicly identifiable information by blurring out faces and license plates [22]. Some researchers have explored replacing the full body of a bystander in a photo with neighboring views or the body of another person [32, 70]. Using another person's face has severe ethical implications that required to be reassessed. Dimiccoli et al. showed that bystander's privacy concerns can be mitigated through image degradation [28]. Realistically swapping faces of a person in a photo has been an area of interest for researchers [18, 53].

But these obfuscation techniques have some limitations in varying circumstances. Prior work has shown that while blurring and pixelating a region in the photo does not hamper photo quality significantly, they are vulnerable to the identification of the hidden face or element [60]. Moreover, masking with a colored rectangle has been effective against recognition by humans but is found to be less satisfactory to the viewers [36]. In addition to exploring different combinations of pixelating and masking to preserve the photo quality while protecting private information [37], prior studies have also shown that the latest image recognition algorithms can retrieve information from blurred and pixelated parts of the images [66]. Compared to blurring or pixelating, masking has been proven effective against image recognition techniques to identify faces and body parts in different scenarios in addition to human recognition [36, 60, 72]. Sun et al. [83] proposed a two-stage Head Inpainting mechanism to reduce the risk of identity detection by threat models. We have chosen blurring as an intermediate redaction technique keeping viewer satisfaction in consideration and gray masking as a permanent redaction technique to reduce the risk of content retrieval.

2.2 Temporal Content Control

For some time, content sharing on social media has been influenced by the idea of getting rid of the content after a short period. Bernstein et al. studied user behavior on the anonymous site 4chan [14] and found that users spend a very short time on each content posted by other users. Leavitt et al. showed that Reddit throwaway accounts are getting popular among women because of the concept of temporary identity [57]. Ayalon et al. showed in a study that Facebook [3] users lose interest in posted content over time [9] and suggested an expiry and extensive archival feature for Facebook posts by users. But Bauer et al. [12] found a gap between users' privacy preference prediction for their future self and suggested that extensive archival features would not be suitable for the users. The idea of "Temporary Social Media" [49] has been widely adopted by several SNS (Social Networking Service) platforms such as Snapchat and Instagram, where the photo is automatically removed after some time of showing it to the viewer. The ephemeral story-sharing feature has become widely popular and influential after Facebook integrated the feature into their SNS platform [94]. Younger SNS users are taking advantage of the ephemeral nature of sharing content as they find it easier to control the content flow [2]. Studies have shown that users prefer ephemeral social networking sites more for lightweight entertainment [50, 86] or as a medium of self-expression [87] than for privacy preservation.

Barua et al. [11] discussed elaborate forgetting mechanisms and interfaces where a decay mechanism of gradually removing obsolete contents would simulate the decay process in human memory. Gulotta et al. designed BitLogic and DataFade where users can upload Photos and those photos will be decayed over time [34]. They found that some users are protective of their self-representation and intrigued by the decaying nature of the system. Mohamed et al. studied user behavior on a

[3]https://www.facebook.com

fictitious system where social media posts were being decayed incrementally each month [67]. In the system, they faded, pixelated, or blurred the whole post including the image and the caption. They found that users are interested in the decaying digital artifact given some control over the decay feature.

In this study, we considered the degradation of a particular region of the photo that contains sensitive content. Two kinds of photo decaying techniques have been introduced – gradual redaction where the content of the photo will be blurred within 12 hours of sending the photo to the viewer and masked with a gray rectangle after 24 hours and delayed redaction where the content will be masked with gray box permanently after 24 hours of sending the photo to the viewers.

2.3 Access control

A common approach to control the access of photos is to allow a specific subset of users to view the photos. However, due to the complex architecture of the privacy settings, users hardly manage the privacy settings frequently and tend to make unintentional mistakes of sharing photos with the wrong audiences [61, 64]. Cook et al. categorized different audience groups in online sharing platforms such as Family and Friends, Professional Peers, and General Audience in the internet [23]. Knijnenburg et al. [52] explored different categorizations for granularity and found that categories with normal granularities (Family, Friends, Classmates, Colleagues, and Acquaintances) are more appropriate for social media users in terms of risk of oversharing and managing privacy settings. Furthermore, a combination of recipient control and content control has been adopted as a measure of protecting photo privacy in several studies. For example, PUPPIES [39] is a mechanism where the users can obscure a region of the photo and specify the privacy policy for specific audiences at the same time.

While social media applications adopt various security mechanisms to protect the user's data, it's possible for the viewers to save media files in their device [51]. Third-party apps such as SnapKeep were rumored to be used to save snaps without notifying the sender [6, 19]. In addition, there have been incidents of leaked photos from renowned social media and photo-sharing sites [33]. P3, a photo encoding algorithm has been proposed where the sensitive region of the photo can be encrypted and extracted in a secure part of a device while preserving the quality of the public part of the photo [75] to ensure that the viewer cannot save the photo in their device. To address unwanted media retrieval, Trusted Execution Environments (TEE) have been adopted among hardware manufacturers. TEE is a mechanism to make sure that the media file cannot be retrieved by the device owner if the device is equipped with a TEE enclave and the media file is encrypted in the enclave [77].

In this study, we considered family members, friends, colleagues and classmates, and other followers as target audiences with whom the participants may share photos. We conducted a between-subject study between two groups of participants where one group was briefed that their photos can only be viewed by those audiences whose devices are equipped with a security chip. Another group was briefed that their photos can be viewed on any device.

3 METHOD

We now describe our survey and data analysis procedures.

3.1 Survey study

3.1.1 Selection of Images. Each survey presented the participants with four types of photos, where each photo contained one of three different content types: information content, facial expression, and bystanders. For information content, we have considered common personally identifiable information (PII) such as a vehicle with the license plate number visible or photos of the front

of a house, where the house number was visible [80]. Prior studies suggested that these PIIs are considered sensitive content by users when they share photos online [59]. In some photos with vehicles, a person was posing with the vehicle. For facial expressions, we showed a forward-facing photo of a person where they have applied an entertaining filter (distortion or color) on their faces; facial expressions are also considered sensitive content for online photo sharing [59]. The bystander content type contained a bystander (a secondary person who is not the subject of the photo) in the photo where the main focus was on another person. All these scenarios were described to the participants with text below the photo. Although the participants did not own or capture the photos, they were instructed to consider the photos as if they owned those photos. The photo descriptions contained the following text.

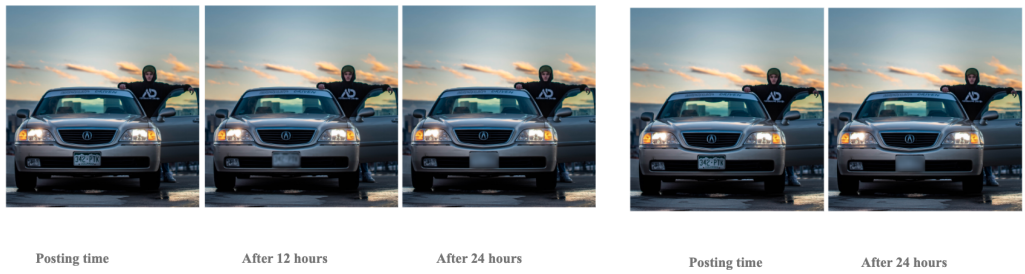
- **Information**: Assume this is a photo of your car (the front of your house), and you would like to share it
- **Facial Expression**: Assume this is a photo of your face.
- **Bystander**: Assume this is a photo of you with another person behind you.

We selected five images each for facial expression and bystander content types and ten images for information content including five images containing vehicles and five images containing the house number. As personally identifiable objects (PII) can be used to trace a person's identity in online networks [54], we wanted to pick situations that most participants might relate to in their personal lives. We collected the images by conducting searches for copyright-free images on Google Images [4] and Unsplash.[5] Each of the images contains exactly one of the three content types. The images were selected to ensure they were respectful of race, age group, and gender. We excluded images with children. In addition, we tried to keep the quality, illumination, and image dimensions consistent across all images.

3.1.2 Selection of redaction techniques: For each participant, four photos were randomly selected, among which two photos contained information content (one with a car, one with a house), one photo containing a facial expression of a person, and one photo had a bystander in the background. Each participant was shown each photo five times (for the following five redaction scenarios).

- **Instant Application of Redaction:** The content of the photo will be masked with a gray rectangle before being sent to the audience.
- **Delayed Application of Redaction:** The original photo will be sent without any obfuscation being applied initially. The particular content of the corresponding photo will be masked out automatically on the recipient's device 24 hours after sending the photo (Figure 1b).
- **Gradual Application of Redaction:** The original photo will be sent without any redaction at first. After 12 hours, a particular content of the photo will be blurred and some aspects of that content will be vaguely recognizable. After 24 hours, that content will be masked out (Figure 1a).
- **Original photo:** The original photo will be sent without any redaction applied. There would be no further automated action by the app.
- **Delayed Deletion:** This condition is similar to that of Snapchat stories,[6] in which no redaction technique is applied, but the entire photo is deleted automatically after 24 hours of posting the photo.

[4]http://images.google.com
[5]https://unsplash.com/license
[6]https://www.snapchat.com/



(a) Gradual redaction (b) Delayed redaction

Fig. 1. A timeline-based visual depiction of gradual and delayed obfuscation techniques (Original Photo by Nima Sarram on Unsplash)

Prior studies have compared different forms of obfuscation techniques such as masking, pixelation, blurring, and silhouette, and have suggested that masking performs better when the objective is to conceal content in the photo. In contrast, blurring is better when the viewer's satisfaction is considered in addition to hiding content in the photo [36, 60]. Hasan et al. found that 'cartooning', or 'coloring' the photo in addition to masking sensitive content does not show any significant improvements in privacy protection [37]. Therefore, we have selected only masking as the permanent form of obfuscation so that the participants' privacy is protected the most in the instant application of redactions and at the final stage of gradual and delayed application of redactions. Also, we have selected blurring as the intermediate obfuscation step as prior studies have found blurring as visually pleasing and somewhat privacy-preserving [36, 60].

To make the participants understand the obfuscation techniques better, we included both visual and text explanations of how the obfuscations would work. For the visual explanation, we showed a timeline (Fig 1) of the image conditions, which showed the transformed image at different times for each corresponding obfuscation technique. For the text explanation, the image condition was comprehensively described for each condition as follows:

"Assume this is a photo of you and your car, and you would like to share it with the license plate of the car being obscured by the app gradually over 24 hours in the recipient's device."

The explanation text varied according to the photo content and the application of redaction of the photo. The visual and text explanations were briefed to the participants both before starting to respond to any image, and while they responded to the images. Overall, each participant was asked questions about a total of twenty scenarios in random order (Appendix B.4.1).

3.1.3 Type of target audiences: For each scenario in the survey, we asked the participants about their likelihood of sharing the photo with four kinds of audiences (family, friends, colleagues, and other followers). These four recipient categories were chosen based on prior literature [59]. In the initial pilot studies, the participants expressed difficulty in understanding the term 'acquaintances' with respect to social media scenarios. In the subsequent pilot studies, the participants indicated that they understood the term 'other followers' better in the context of photo sharing on social media. Therefore, the fourth category of 'acquaintances' was rephrased as 'other followers' in the main study.

3.1.4 Device type of the image viewers: The participant pool was divided into two groups randomly to conduct a between-subject study based on the device condition of the image viewer. Participants were briefed that their images contain sensitive content. To convey that the original images with

sensitive content cannot be easily saved or captured as a screenshot on the viewers' devices, we described two different approaches to these two groups. The instructions contained the following text for the two conditions:

- **Trusted device**: "Assume the recipient's phone is equipped with a special security chip that will make it impossible for the viewer to save or take a screenshot of the photo."
- **Non-trusted device**: "Assume the recipient's social media app does not allow the viewer to save the photo and (like Snapchat) notifies the sender of the photo if the viewer takes a screenshot."

To arrive at this particular operationalization of 'TEEs', the research group went through multiple iterations to design these instructions. This text was further refined through our pilot studies. The objective of the instructions was to convey the implications of software and hardware security in the context of photo sharing to non-technical participants. At the same time, we deliberately avoided directly briefing the participants about the possible risks involving their photos being saved by third-party applications in the viewers' devices so as to not bias their responses towards more privacy.

The full instruction is portrayed in Appendix B.3.1.

3.1.5 Measuring sharing likelihood. We asked the following question (paraphrased) for each scenario (see Appendix B.4 for our survey instrument):

Q. *Assume this is a photo of you and your car, and you would like to share it with the license plate of the car being masked by the security chip in the recipient's phone after 24 hours in the recipient's device.* This question varied based on the redaction condition, device TEE condition, and the content type in the photo. Participants were asked to select from a seven-point Likert scale: (1) extremely unlikely (2) moderately unlikely (3) slightly unlikely (4) neither likely nor unlikely (5) slightly likely (6) moderately likely, and (7) extremely likely. The Likert scale is adapted from [88].

3.1.6 Organization of the survey. The survey consisted of 52 questions in the mostly close-ended form. The survey instrument was organized as follows (see Appendix B for the survey instrument):

- CAPTCHA to prevent automated responses
- Pledge to provide authentic and high-quality responses [46]
- Consent form
- Questions about which (if any) social media platforms they use, and how often they visit these sites
- Questions about which (if any) social media platforms they use to share photos online, and how often they share their photos.
- Twenty scenarios with four unique photos along with descriptive text, presented in random order (within-subjects), each with four questions about sharing likelihood with family, friends, colleagues, and other followers. Note that each participant was assigned to a single device condition (whether the target audience's device contains a TEE or not), and these questions were asked in the context of one kind of device.
- Questions about whether they had ever shared a photo containing sensitive information and their most recent experience sharing an image. The questionnaire and scale were adapted from the work by Amon et al. [7] (Appendix B.6)
- Question about their privacy preference. The questionnaire was adapted from the work by Hoyle et al. [41, 42] (Appendix B.7)
- Questions about their interpersonal trust level (Appendix B.8)
- Five demographic questions (age, gender, race or nationality, education, and how long they have lived in the USA).

3.1.7 Recruitment. The survey was conducted on Qualtrics for two months between May and July 2021. The survey was rolled out to small batches of participants on different days of the week and at different times of the day to lower any kind of temporal bias. We recruited over 500 participants from Prolific.[7]

After filtering participants based on our attention check questions, we were left with 385 participants in the final sample. In our final sample, 215 (55.8%) participants identified themselves as male, 163 (42.3%) as female, and six (1.56%) as non-binary. Most of the participants were under 49 years of age, with 149 (38.7%) between 18 and 29 years, 194 (50.4%) between 30 and 49 years, 35 (9.09%) between 50 and 64 years, and six (1.56%) participants 65 years or older. For the highest level of education, 138 (35.8%) participants mentioned having an undergraduate degree, 41 (10.6%) a high school diploma, 87 (22.6%) a Master's degree, and six (1.56%) a professional degree. Most (243, 63.1%) participants were White, 49 (12.7%) were Black or African American, 50 (13.0%) were Asian, and 18 (4.68%) were Hispanic or Latino.

3.1.8 Compensation and ethical considerations . Our protocol was approved by our institution's ethics review board. We performed several online pilot studies to determine the approximate time range required to participate in the study. The median time taken for the study was a little above 16 minutes and the mean time was almost 18 minutes. Regardless of whether or not we used their responses, each participant was paid $3.00 to complete the 15–20 min survey. The payment amount is in line with – and surpasses – the recommendation in Silberman et al [80] to pay workers at least minimum wage in the study's location.

3.1.9 Pilot study. We conducted online pilot studies with small batches of participants (20–30) to determine whether the participants were understanding the scenarios properly and whether the survey structure needed to be modified. We also requested them to suggest improvements to our survey (if any). The pilot studies occurred in five phases, each phase improved from the previous phase with the recommendation of the pilot participants such as improving instruction text or image selection. In the fifth phase, each participant provided satisfactory feedback and thus we finalized our survey.

3.2 Data analysis procedure

We now describe our quantitative analysis procedures.

3.2.1 Quantitative analysis. We collected our data in Likert-scale form. Our data do not meet the assumptions of parametric tests [31], such as normality and equal variance of errors. Hence, We used non-parametric versions for all the statistical tests. We have one dependent variable (sharing likelihood) and several independent variables (content type, redaction techniques, device condition, and target audience). We used linear mixed-effects models [30] with fixed slopes and random intercepts for each participant. We also added interaction terms involving redaction techniques, target audience, and content types. The linear mixed-effect model was chosen because it considers the effects of having participants contribute multiple data points. We used estimated marginal means to compute all the pairwise comparisons, which helped determine the significant effects across the interaction effects. The p-values were adjusted with the Tukey method [89] as it considers multiple comparisons and adjusts the p-value to minimize the risk of Type I errors.

[7] https://prolific.co/

4 FINDINGS

The omnibus test involving the mixed effect model is shown in Table 1. Content, redaction, target, and some interaction terms showed significant effects on photo sharing likelihood. In the following sections, we present our key findings.

	Sum Sq	Mean Sq	DoF	Den DoF	F statistic	η_p^2
Content	108.3	54.1	2	10.1	19.45^{***}	<0.01
Trust Condition	5.2	5.2	1	377.6	1.86	<0.01
Redaction	193.7	48.4	4	29692.3	17.3988^{***}	<0.01
Target	19599	6533	3	29692	2347.25^{***}	0.8
Content: Trust Condition	60.3	30.2	2	29704.7	10.83^{***}	<0.01
Redaction: Target	381.9	31.8	12	29692.3	11.44^{***}	0.016
Content: Redaction	2891.9	361.5	8	29692.3	129.8^{***}	0.12
Trust Condition: Redaction	33.9	8.5	4	29692.3	3.04^{*}	<0.01
Trust Condition: Target	38.7	12.9	3	29692.3	4.63^{**}	<0.01

Table 1. Type III ANOVA Table (with Satterthwaite's method). (* = p < 0.05, ** = p < 0.01, *** = p < 0.001). The effect size η_p^2 (partial η^2) can be interpreted as small if $\eta_p^2 = 0.01$, medium if $\eta_p^2 = 0.06$, and large if $\eta_p^2 = 0.14$ [56]. The viewing platform condition is rephrased as 'Trust condition'.

4.1 Effect of Social Media Platform

Out of 385 participants, 291 participants (75.6%) reported sharing photos online with frequency ranging from multiple times a day to a few times a month, and only one (0.26%) participant reported never posting photos online. A majority (233, 60.5%) of participants indicated that they share most of their photos in Instagram,[8] while 193 (50.1%) use Facebook,[9] 96 (24.9%) participants use Twitter,[10] and 68 (17.6%) use Snapchat[11] as their medium to share photos frequently. Some participants marked multiple platforms as photo-sharing social media.

Frequent Instagram users showed a significant difference in photo sharing likelihood in terms of applications of redactions. They showed a slight preference for sharing ephemeral photos to sharing photos with delayed or gradual applications of redactions. On the other hand, Twitter users had a slightly more sharing likelihood for instant redaction over gradual application of redaction (Table 2).

4.2 Effects of Redaction Applications, Sensitive Content Types, Target Audiences, and Viewing Platform Conditions on Photo Sharing Likelihood

To investigate the main effects of redaction techniques, sensitive content types, target audiences, and viewing platform conditions, we conducted post hoc pairwise tests (Table 3).

We observed slight differences in sharing likelihood involving applications of redactions. For sensitive content types, we found that participants are more likely to share photos containing bystanders than facial expressions. Regarding the target audience, participants were much more likely to share photos with family members and friends compared to other followers. We did not

[8]https://www.instagram.com/
[9]https://www.facebook.com/
[10]https://www.twitter.com/
[11]https://www.snapchat.com/

Redaction technique	Estimate	SE	DF	T-ratio	Effect Size (Cohen's *d*)	P-value
Instagram						
Delayed < Deleted	0.2	0.04	498.53	4.84	0.11 (negligible)	0.001*
Gradual < Deleted	0.22			5.16	0.11 (negligible)	0.0002*
Twitter						
Gradual < Instant	0.27	0.06	49853	4.09	0.15 (negligible)	0.001*

Table 2. Pairwise comparison of sharing likelihood in photos with different redaction techniques for Frequently used Social Media Platforms. Significance after Tukey adjustment is indicated with p-values. The direction of the difference is indicated by '<' and '>'. Pairs without any significant difference are skipped.

Pairwise Comparisons	Estimate	SE	t ratio	Effect Size (Cohen's *d*)	P-value
Redaction Techniques					
Delayed < Instant	0.14		4.36	0.08 (negligible)	0.0001*
Gradual < Instant	0.13		4.21	0.08 (negligible)	0.0002*
Original < Instant	0.1		3.24	0.06 (negligible)	0.01*
Gradual < Deleted	0.13		4.14	0.07 (negligible)	0.0003*
Delayed < Deleted	0.13		4.29	0.08 (negligible)	0.0002*
Deleted < Original	0.1		3.17	0.06 (negligible)	0.013*
Content Types					
Facial expression < Bystander	1.18	0.19	6.28	0.7 (medium)	0.0002*
Information < Bystander	0.5	0.17	2.88	0.28 (small)	0.04*
Information < Facial expression	0.7		4.14	0.41 (small)	0.005*
Target Audiences					
Friends < Family	0.11		4.25	0.07 (negligible)	0.0001*
Other followers < Family	1.93		69.5	1.13 (large)	<0.0001*
CnC < Family	1.3	0.03	45.8	0.75 (medium)	<0.0001*
CnC < Friends	1.15		41.54	0.68 (medium)	<0.0001*
Other followers < Friends	1.81		65.3	1.06 (large)	<0.0001*
Other followers < CnC	0.66		23.76	0.39 (small)	<0.0001*

Table 3. Pairwise comparison of sharing likelihood in photos for redaction techniques, content types, and target audiences. CnC= 'Colleagues and classmates'. Significance after Tukey adjustment is indicated with p-values. The direction of the difference is indicated by '<' and '>'. Pairs without any significant difference are skipped.

find any significant difference between sharing photos in trusted and non-trusted viewing platforms (Figure 4 in Appendix A).

4.3 Interaction effect of Redaction Applications and Sensitive Content Types on Photo Sharing Likelihood

Table 4 shows the estimated mean values of the redaction applications grouped by sensitive content types. To further investigate this interaction effect, we conducted Tukey post hoc pairwise tests. In particular, we tested the following hypothesis.

Content Type	Redaction technique	Estimated mean	SE	DF	CI
Information	Instant	4.82	0.116	20.1	0.48
	Delayed	4.32			0.49
	Gradual	4.32			0.48
	Deleted	4.25			0.48
	Original	4.01			0.49
Facial Expression	Original	4.18	0.144	17.1	0.6
	Deleted	4.14			0.6
	Gradual	3.58			0.6
	Delayed	3.51			0.61
	Instant	2.81			0.61
Bystander	Instant	5.01	0.143	17.1	0.61
	Original	4.85			0.61
	Deleted	4.81			0.61
	Delayed	4.75			0.6
	Gradual	4.71			0.61

Table 4. Sharing Likelihood grouped by content type across redaction techniques, in terms of estimated mean values, standard error, degree of freedom, and confidence intervals. Results are averaged over the levels of the target audience, and condition. Degrees-of-freedom method: Satterthwaite. Confidence level used: 0.95

H1 : Sharing likelihood is affected when a particular temporal redaction technique is applied based on the type of sensitive content.

Instant redaction was preferred for photos containing personally identifiable information or bystanders. On the other hand, sharing the original photo or sharing photos with ephemerality was deemed to be popular when a photo contained facial expressions. We observed statistically significant differences for some of the pairwise comparisons of images with different redaction techniques for each type of sensitive content. The pairwise comparisons in addition to the p-values are shown in Table 5.

Information content: We observed that the participants somewhat preferred the instant application of redaction over all other redaction applications. On the other hand, participants were slightly reluctant to share the original version of the photo compared to sharing ephemeral photos or photos with delayed or gradual application of redaction when the photo contained personally identifiable information such as car registration number or house number in their photos.

Facial expression: Sharing the original photo without any obfuscation applied was slightly preferred when the photo contained facial expressions with an entertaining filter. Overall, the participants slightly preferred delayed or gradual application of redaction compared to the instant application of redaction in this case.

Bystander: When there was a bystander present in the photo, participants were slightly more willing to mask out the face of the bystander right away (instant redaction) while posting the photo on social media compared to applying the obfuscation at a later time.

Overall, these findings indicate varying sharing preferences for photos containing different sensitive contents. Participants were somewhat more comfortable applying redactions instantly when there was a personally identifiable information or a bystander in the photo (Figure 2) compared to the other applications of redactions or the base conditions. For photos with facial expressions,

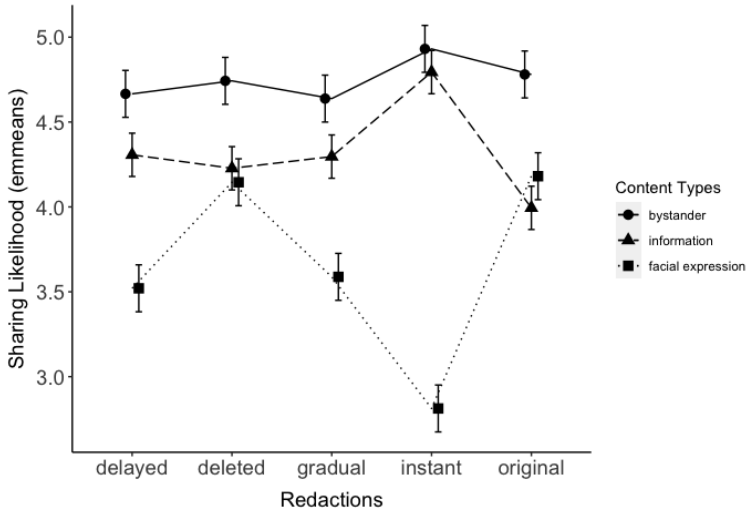


Fig. 2. Differences in sharing likelihood (estimated marginal means) for different content types. Participants somewhat preferred applying redactions instantly for photos containing bystanders or personally identifiable information. In contrast, participants were much more comfortable sharing the photo without any application of redactions or sharing ephemeral photos when the photos contained facial expressions.

participants were much more likely to share photos without any application of redaction or ephemeral photos compared to the application of instant redaction on the face.

4.4 Interaction Effect of Redaction Applications and Target Audience on Photo Sharing Likelihood

The estimated mean of photo sharing likelihood for each redaction application grouped by target audiences can be found in Table 6. We conducted another set of Tukey posthoc pairwise tests to compare redaction applications and target audiences (Table 7). We tested the following hypothesis:

H2: Sharing likelihood varies when different temporal redaction techniques are applied to the photo for the same recipient group.

Our observations are as follows for different target audiences.

Family Members: Participants were somewhat more prone to share original photos compared to applying any redaction techniques (instant, gradual, or delayed) when it came to sharing photos with a family member. Also, deleting photos after 24 hours of sending was slightly preferred over instant, gradual and delayed application of redactions.

Friends: Participants' sharing preference with their friends was similar to their sharing pattern with family. For instance, sharing original photos was slightly preferred over delayed, instant, or gradual application of redaction, and sharing the ephemeral photo was slightly preferred over all other techniques.

Other Followers: Participants slightly preferred instant redaction over sharing the original photo or applying gradual or delayed redaction for the general audience.

Colleagues and Classmates: We did not find any significant effects among redaction applications when photos were being shared with Colleagues and Classmates.

In short, our findings suggest that participants slightly preferred not to apply any kind of temporal redaction on their photos when they share a photo with their friends or family members. In contrast,

Redaction technique	Estimate	SE	t ratio	Effect Size (Cohen's *d*)	P-value
			Information		
Original < Deleted	0.234		5.48	0.14 (negligible)	<0.0001*
Original < Instant	0.81		18.77	0.48 (small)	<0.0001*
Original < Gradual	0.303		7.05	0.18 (negligible)	<0.0001*
Original < Delayed	0.31	0.043	7.22	0.19 (negligible)	<0.0001*
Deleted < Instant	0.57		13.2	0.34 (small)	<0.0001*
Gradual < Instant	0.5		11.72	0.3 (small)	<0.0001*
Delayed < Instant	0.49		11.56	0.3 (small)	<0.0001*
			Facial Expression		
Gradual < Original	0.6		9.87	0.36 (small)	<0.0001*
Delayed < Original	0.66		10.96	0.4 (small)	<0.0001*
Instant < Original	1.37		22.67	0.83 (large)	<0.0001*
Instant < Deleted	1.33	0.06	21.99	0.8 (large)	<0.0001*
Instant < Delayed	0.707		11.7	0.43 (small)	<0.0001*
Instant < Gradual	0.77		12.79	0.47 (small)	<0.0001*
Gradual < Deleted	0.56		9.19	0.34 (small)	<0.0001*
Delayed < Deleted	0.62		10.29	0.38 (small)	<0.0001*
			Bystander		
Gradual < Instant	0.26	0.06	4.83	0.18 (negligible)	0.0014*
Delayed < Instant	0.29		4.27	0.156 (negligible)	0.0001*

Table 5. Pairwise comparison of sharing likelihood in photos with different redaction techniques for each content type. Significance after Tukey adjustment is indicated with p-values. The direction of the difference is indicated by '<' and '>'. Pairs without any significant difference are skipped.

they slightly preferred the instant application of redaction while sharing photos with the general audience (Figure 3).

4.5 Effect of Trusted Device Condition

We analyzed the interaction effect of the trusted device condition with redaction, content type, and target audience separately. In addition, we analyzed the main effect of the trusted device condition. We did not find any significant effect of trusted device condition in the sharing likelihood of photos except in two cases.

4.6 Preference for Temporal Redaction

For each participant, we took the sharing likelihood involving the 16 combinations of content types (four) and target audiences (four) for each redaction application. We counted how many participants had the highest sharing likelihood for delayed or gradual redaction involving each combination of content type–target audience. We considered delayed or gradual redaction having the highest sharing likelihood if they had a tie with other redaction or the base conditions. Table 8 shows that a part of the participants preferred delayed or gradual redactions in particular situations.

5 DISCUSSION

In this section, we discuss a) the effect of sensitive content types on redaction preferences, b) the relation between the target audience and redaction preference, c) the potential of temporal redaction

Target Audience	Redaction technique	Estimated mean	SE	DF	CI
Family	Original	5.38	0.098	43.9	0.39
	Deleted	5.28			0.4
	Delayed	5.01			0.4
	Gradual	5.00			0.4
	Instant	4.84			0.4
Friends	Original	5.17	0.098	43.9	0.4
	Deleted	5.15			0.39
	Delayed	4.91			0.4
	Gradual	4.91			0.4
	Instant	4.78			0.39
Colleagues and Classmates	Deleted	3.94	0.098	43.9	0.4
	Instant	3.88			0.39
	Original	3.79			0.39
	Gradual	3.77			0.4
	Delayed	3.77			0.4
Other Followers	Instant	3.35	0.098	43.9	0.4
	Deleted	3.22			0.4
	Gradual	3.14			0.39
	Delayed	3.09			0.39
	Original	3.05			0.4

Table 6. Sharing Likelihood grouped by Target Audiences across Redaction Techniques, in terms of estimated mean values, standard error, degree of freedom, and confidence intervals. Results are averaged over the levels of content, and condition. Degrees-of-freedom method: Satterthwaite. Confidence level used: 0.95

as a privacy tool, d) the intelligibility of trusted hardware among users, and e) the limitations of our study.

5.1 Redaction decision depends on what type of content the photo contains

Our findings indicate that people tend to choose similar redaction techniques to obscure bystanders and sensitive information content types. Participants preferred applying the instant, gradual, or delayed application of redactions when there was personally identifiable information present in the photo (such as vehicle license plate or house number). This finding is intuitive because of people's general tendency to preserve their private information. A study on photo privacy detection suggests that people prefer that contents with personally identifiable information such as vehicle license plates and driver's licenses should not be revealed on their social media [85]. Protecting the identity of bystanders with instantly applied redactions supports previous claims that people tend to respect other people's privacy while sharing photos [3, 4, 15, 44]. Prior work also showed that users tend to avoid social tension caused by unauthorized sharing of photos containing other people [17].

In contrast, when participants were asked about photos containing facial expressions with 'funny' filters, most of them chose to share photos without any redactions being applied, although several participants showed an inclination towards delayed or gradual applications of redactions along with simply deleting the photo 24 hours after posting. The aversion to using instant redactions on photos with facial expressions with filters can be backed by previous findings. Choi et al. found that

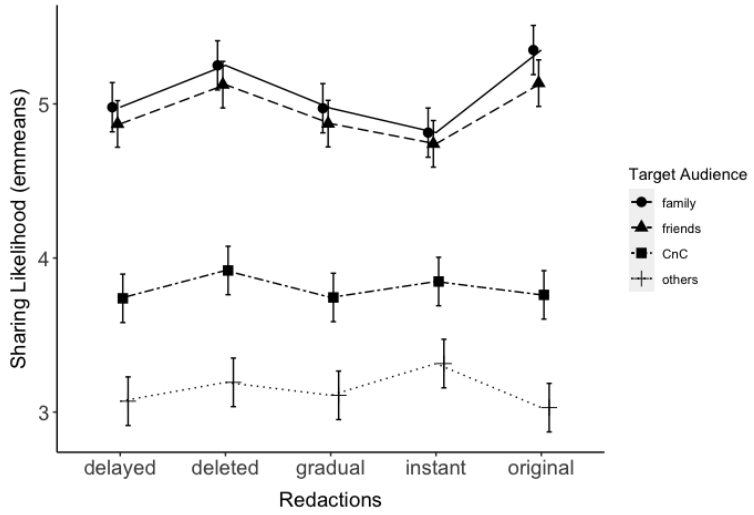


Fig. 3. Differences in sharing likelihood (estimated marginal means) for different target audiences. Participants were comfortable sharing photos without any application of redactions when they were told to share with family members or friends. In contrast, participants showed interest in applying instant redactions to photos before sharing with a general audience.

Redaction Technique	Estimate	SE	T-ratio	Effect size (Cohen's *d*)	P-value
			Family		
Delayed < Original	0.37	0.06	6.1	0.23 (small)	<0.0001*
Instant < Original	0.54		8.86	0.33 (small)	<0.0001*
Gradual < Original	0.39		6.28	0.23 (small)	<0.0001*
Delayed < Deleted	0.27		4.39	0.16 (negligible)	<0.0001*
Instant < Deleted	0.44		7.16	0.26 (small)	< 0.0001*
Gradual < Deleted	0.28		4.57	0.17 (negligible)	0.0008*
			Friends		
Delayed < Original	0.26	0.06	4.2	0.16 (negligible)	0.005*
Gradual < Original	0.26		4.22	0.16 (negligible)	<0.004*
Instant < Original	0.39		6.4	0.24 (small)	<0.004*
Instant < Deleted	0.38		6.15	0.23 (small)	<0.0001*
Gradual < Deleted	0.24		3.97	0.15 (negligible)	0.01*
Delayed < Deleted	0.24		3.95	0.15 (negligible)	0.0115*
			Other Followers		
Original < Instant	0.3	0.06	4.9	0.18 (negligible)	0.0001*
Delayed < Instant	0.26		4.17	0.153 (negligible)	0.004*

Table 7. Pairwise comparison of sharing likelihood in photos with different redaction techniques for each target audience type. Significance after Tukey adjustment is indicated with p-values. The direction of the difference is indicated by '<' and '>'. Pairs without any significant difference are skipped.

users of ephemeral sites are inclined to share their "true and actual self" [21]. Katz et al. found that

	Family	Friends	CnC	Other followers	Redaction
Car	19.9%	18.5%	16.9%	17%	Delayed
	18.7%	17.9%	16.8%	17.1%	Gradual
House	18.9%	19.7%	20.2%	20.8%	Delayed
	18.3%	19%	21.7%	21.8%	Gradual
Bystander	18.4%	19.3%	19.4%	20.2%	Delayed
	19.3%	18%	17.9%	18%	Gradual
Facial expression	17.5%	17.3%	18.8%	18.4%	Delayed
	18.6%	20.3%	19.5%	19.4%	Gradual

Table 8. Percentage of participants having highest sharing likelihood for gradual or delayed redaction involving particular content and target audience. CnC = Colleagues and Classmates.

users consider ephemeral sharing as "fun" with their close network [50]. These findings explain the motive behind posting such photos without any obfuscations or as ephemeral photos that 'disappear' after a short period; any instant redactions would hinder expressing one's true self with their close social contacts.

5.2 Social tie affects redaction preference

When sharing photos with family members or friends, participants were reluctant to apply any obfuscations to their photos. This finding is in line with prior work, which found that users' sharing likelihood is higher when they share a photo with family members or friends [52]. This sharing behavior can also be attributed to the fact that when it comes to sharing photos, people trust family members and friends more than their other social ties [4, 59]. Another reason could be participants wanted to avoid the negative effect on viewers' satisfaction (when redactions are applied to a region of the photo) when the viewer is a family member or a friend [36, 37, 60]. Besides, Holloway et al. investigated how people have adopted Social Media (Facebook[12]) for preserving their family photos [40]. They found that some users archive family photos on Facebook to keep them as memories and to share them with family members and close friends for future reference. However, it was interesting to notice that participants were also inclined toward the ephemeral mode of photo sharing when sharing photos with family members or friends, where the photo will be deleted after 24 hours (Figure 3). Future studies need to be conducted to find out the motivation behind this sharing behavior.

For more distant contacts such as colleagues, our participants did not show any significant tendency toward instant, gradual, or delayed redaction techniques although a prior study found that some users don't share photos with colleagues as the photo may contain sensitive content with the potential to reveal "white lies" or remove "plausible deniability" at work [59]. There is a possibility that the type of sensitive content they don't want to reveal was not included in our study or the participants in general, didn't want to take any risks sharing the photo itself with their co-workers. Future studies need to be conducted to find if there is any particular method that increases sharing of photos with colleagues and classmates.

Finally, for sharing photos with 'other followers', our participants indicated a slight inclination toward applying instant redactions for photos in general which confirms prior finding that users are reserved in terms of privacy concerns while sharing with other followers [52].

[12]https://www.facebook.com/

5.3 Preferences towards temporal redactions

Although we found an overall tendency towards existing redaction techniques such as instant redaction or ephemeral photos, we found that participants often preferred other temporal redactions. For example, we counted the fraction of participants having the highest sharing likelihood for delayed or gradual redaction in each of 16 combinations of content types and target audiences (Table 8). We found that a part of the participants (roughly 17–22%) had the highest sharing likelihood for delayed or gradual application of redactions involving particular target–content type combinations (more details in Section 4.6).

One possible reason for such preference for the gradual or delayed application of redaction can be explained by the temporal protection offered by these obfuscation techniques. For instance, one way of sharing an individual's photo with their friends using an existing redaction technique is to apply instant masking. However, prior studies have shown that using instant masking may reduce the overall appeal of that photo [36, 37, 60]. Given that users trust their friends more while sharing photos on social media [4, 59], some users may prefer to share the original photo with their friends for a certain time retaining the aesthetics of the photo while masking it after a short period to reduce the risk of revealing any sensitive content in the longer term.

In addition, some studies have suggested that users are interested in a 'content decay' feature if they are given the ability to undo the decay or the control over what part of the photo is going to be decayed [67]. It would be interesting to observe how the participants react to temporal redactions if they are provided with more control over the redaction policies.

5.4 Perceived trustworthiness of trusted hardware

Despite our sample size that was chosen to detect medium and large-sized effects, we did not find any significant main effect or an interaction effect with redaction techniques, content type, or target audiences in the sharing likelihood between the trusted device and non-trusted device conditions. Although we found a couple of interaction effects in specific cases, more research is needed to understand the perceptions of trustworthiness in different contexts.

Prior studies have suggested that 'visibility' as a form of 'affective feedback' raises privacy awareness in the use of technology [48, 78]. Because 'secure enclaves' were not directly visible to our participants, it is possible that they were not aware of the implications of trusted hardware. In addition, Davinson et al. suggested that users tend to be unaware when they are not briefed about the potential risks online, and raising awareness of susceptibility increases users' security behaviors [26]. Future work is needed on how to adequately convey the necessity or importance of trusted hardware to users as well as effective communication of risks based on their perceptions and mental models of the risks [8].

Prior studies have identified that users' trust in a social media application largely depends on factors such as the reputation of that organization regarding privacy preservation, privacy policies, and transparency about technology and company policies [55]. Users tend to be unaware of possible breaches in a feature provided by a reputed organization [16, 47]. It is possible that our participants trust the companies behind the applications they use. In addition, some studies have found that users with a security background are more aware of the privacy mechanisms compared to non-security background users [69]. It will be interesting to study users' sharing behaviors focusing on participants' trust in social media and their security background in future studies.

Furthermore, this study has not captured the participants' perceived trustworthiness in the context of institutional privacy threats such as surveillance by intelligence agencies, or generation of behavioral data by the sharing platform itself [35]. There is a possibility that the participants may express different preferences toward trusted hardware if the scenarios involved institutional

privacy threats. For instance, trusted hardware can potentially be an influential tool to regain the trust of users after a social media platform is affected by a data breach. Future studies can look into the importance of trusted hardware in the context of rebuilding trust.

5.5 Limitations and Future Work

We have studied participants' behavior for two concepts – a) temporal photo redactions and b) perceptions of interpersonal privacy threats on trusted viewing platforms while sharing photos on social media. Although the temporal photo redactions were demonstrated to the participants with visual depictions and text based instructions, the trusted device conditions were described succinctly to avoid any bias toward trusted devices. Our descriptions may have been unable to stimulate any privacy benefits in the context of potential risks.

Furthermore, this study focused only on interpersonal privacy threats in the context of photo sharing on social media. If similar experiments are conducted in the context of institutional privacy threats, they may produce different responses from the participants. Future studies can focus on comparing participant preferences towards temporal redaction techniques in the context of institutional and interpersonal privacy threats.

This study did not address how sharing intent might affect the sharing preferences in different scenarios. Sharing intent can potentially affect the redaction preferences in different contexts. For instance, participants wanting to gain more followers might be motivated to share their facial expressions without any redactions [10]. Tosun et al. have suggested that users' Facebook activities vary when they use the platform for entertainment compared to when they use it for organizing social activities [84]. Hence, these different types of intentions may also affect the participants' preference for redaction techniques. Future studies can study redaction preferences by also manipulating or examining sharing intentions.

In this study, the images were selected by the research group and the participants indicated their sharing preferences based on a hypothetical scenario. Although our approach is in line with previous studies [36, 37, 60], the results may not generalize to real-world settings. User preferences may differ when they share their own photos actively on social media and may depend on other factors as discussed earlier. Therefore, future studies should explore more realistic scenarios with interactive sharing options in the context of sharing photos with temporal redaction techniques.

Finally, we conducted the study only with participants from the USA. Future studies are needed to study other populations.

6 CONCLUSION

Photo sharing applications such as Snapchat allow people to share 'ephemeral photos', which disappear soon after the viewer has looked at the photo. This functionality allows people to share photos that are more privacy sensitive by nature. We explore the concept of 'temporal redactions', where viewers can see the original photo, following which redactions are gradually applied to sensitive regions. We conducted a user study (N=385) to investigate the photo sharing behaviors of participants for gradual and delayed applications of redactions on a region of a photo that they share with viewers and compared the result with existing redaction mechanisms such as sharing with an instant application of redaction, sharing without any obfuscation, and sharing with ephemerality (delayed deletion). We also investigated whether the participants' photo sharing preferences vary depending on whether the photo viewers' devices include 'trusted hardware' for stricter enforcement of temporal redactions.

Regarding the temporal application of redactions such as gradual or delayed redactions, we found that participants were comfortable with the existing sharing mechanisms in varying circumstances. However, many participants (17–22%) showed a preference for delayed or gradual application of

redaction when the content of the photo did not pose a significant interpersonal privacy threat or the viewer they shared the photo with was a close social contact. Future studies may provide participants with more control over the decay mechanism of the photo and evaluate their sharing behaviors compared to existing sharing mechanisms. The sharing behavior between the participants presented with the trusted viewing platform vs the non-trusted viewing platform did not show any significant differences. Future studies are needed to better convey the assurances provided by trusted hardware to lay users.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under grants CNS-1703853 and CNS-1704139 as well as Grant Thornton through an equipment grant. We would also like to thank John R. Lange for sharing his insights on this project.

REFERENCES

[1] Esther Addley. Photo-sharing is back: how social media has framed the pandemic. *https://www.theguardian.com/world/2021/apr/16/photo-sharing-is-back-how-social-media-has-framed-the-pandemic*, 2021.

[2] Michael Adorjan and Rosemary Ricciardelli. A new privacy paradox? youth agentic practices of privacy management despite "nothing to hide" online. *Canadian Review of Sociology/Revue canadienne de sociologie*, 56(1):8–29, 2019.

[3] Taslima Akter, Tousif Ahmed, Apu Kapadia, and Manohar Swaminathan. Shared privacy concerns of the visually impaired and sighted bystanders with camera based assistive technologies. *ACM Transactions on Accessible Computing (TACCESS)*, 2021.

[4] Taslima Akter, Bryan Dosono, Tousif Ahmed, Apu Kapadia, and Bryan Semaan. " i am uncomfortable sharing what i can't see": Privacy concerns of the visually impaired with camera based assistive applications. In *29th {USENIX} Security Symposium ({USENIX} Security 20)*, pages 1929–1948, 2020.

[5] Saleem Alhabash and Mengyan Ma. A tale of four platforms: Motivations and uses of facebook, twitter, instagram, and snapchat among college students? *Social media+ society*, 3(1):2056305117691544, 2017.

[6] Maryam S AlOshan. Information privacy violations in ephemeral communications. In *Third International Congress on Information and Communication Technology*, pages 63–77. Springer, 2019.

[7] Mary Jean Amon, Rakibul Hasan, Kurt Hugenberg, Bennett I Bertenthal, and Apu Kapadia. Influencing photo sharing decisions on social media: A case of paradoxical findings. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 1350–1366. IEEE, 2020.

[8] Farzaneh Asgharpour, Debin Liu, and L Jean Camp. Mental models of security risks. In *International Conference on Financial Cryptography and Data Security*. Springer, 2007.

[9] Oshrat Ayalon and Eran Toch. Retrospective privacy: Managing longitudinal privacy in online social networks. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, pages 1–13, 2013.

[10] Saeideh Bakhshi, David A Shamma, and Eric Gilbert. Faces engage us: Photos with faces attract more likes and comments on instagram. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 965–974, 2014.

[11] Debjanee Barua, Judy Kay, Bob Kummerfeld, and Cecile Paris. Theoretical foundations for user-controlled forgetting in scrutable long term user models. In *Proceedings of the 23rd Australian Computer-Human Interaction Conference*, pages 40–49, 2011.

[12] Lujo Bauer, Lorrie Faith Cranor, Saranga Komanduri, Michelle L Mazurek, Michael K Reiter, Manya Sleeper, and Blase Ur. The post anachronism: The temporal dimension of facebook privacy. In *Proceedings of the 12th ACM Workshop on Workshop on Privacy in the Electronic Society*, pages 1–12, 2013.

[13] Joseph B Bayer, Nicole B Ellison, Sarita Y Schoenebeck, and Emily B Falk. Sharing the small moments: ephemeral social interaction on snapchat. *Information, Communication & Society*, 19(7):956–977, 2016.

[14] Michael Bernstein, Andrés Monroy-Hernández, Drew Harry, Paul André, Katrina Panovich, and Greg Vargas. 4chan and/b: An analysis of anonymity and ephemerality in a large online community. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 5, 2011.

[15] Andrew Besmer and Heather Richter Lipford. Moving beyond untagging: photo privacy in a tagged world. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1563–1572, 2010.

[16] Sruti Bhagavatula, Lujo Bauer, and Apu Kapadia. What breach? Measuring online awareness of security incidents by studying real-world browsing behavior. In *2021 European Symposium on Usable Security*. ACM, October 2021.
[17] Jens F Binder, Andrew Howes, and Daniel Smart. Harmony and tension on social network sites: Side-effects of increasing online interconnectivity. *Information, Communication & Society*, 15(9):1279–1297, 2012.
[18] Dmitri Bitouk, Neeraj Kumar, Samreen Dhillon, Peter Belhumeur, and Shree K Nayar. Face swapping: automatically replacing faces in photographs. In *ACM SIGGRAPH 2008 papers*, pages 1–8. 2008.
[19] Rebecca Borison. This hack lets people save your snapchats for all of eternity. *https://www.businessinsider.com/save-your-snapchats-forever-2014-7*, 2014.
[20] Kelly Caine. *Exploring everyday privacy behaviors and misclosures*. Georgia Institute of Technology, 2009.
[21] Tae Rang Choi and Yongjun Sung. Instagram versus snapchat: Self-expression and privacy concern on social media. *Telematics and Informatics*, 35(8):2289–2298, 2018.
[22] Amanda Conway. Face blurring: when footage requires anonymity. *https://blog.youtube/news-and-events/face-blurring-when-footage-requires/*, 2012.
[23] Eric C Cook and Stephanie D Teasley. Beyond promotion and protection: Creators, audiences and common ground in user-generated media. In *Proceedings of the 2011 iConference*, pages 41–47. 2011.
[24] Marciano da Rocha, Dalton Cézane Gomes Valadares, Angelo Perkusich, Kyller Costa Gorgonio, Rodrigo Tomaz Pagno, and Newton Carlos Will. Secure cloud storage with client-side encryption using a trusted execution environment. *arXiv preprint arXiv:2003.04163*, 2020.
[25] Sanchari Das, Tousif Ahmed, Apu Kapadia, and Sameer Patil. Does this photo make me look good? how social media feedback on photos impacts posters, outsiders, and friends. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–32, 2021.
[26] Nicola Davinson and Elizabeth Sillence. It won't happen to me: Promoting secure behaviour among internet users. *Computers in Human Behavior*, 26(6):1739–1747, 2010.
[27] Bernhard Debatin, Jennette P Lovejoy, Ann-Kathrin Horn, and Brittany N Hughes. Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *Journal of computer-mediated communication*, 15(1):83–108, 2009.
[28] Mariella Dimiccoli, Juan Marín, and Edison Thomaz. Mitigating bystander privacy concerns in egocentric activity recognition with deep learning and intentional image degradation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4):1–18, 2018.
[29] Judicael B Djoko, Jack Lange, and Adam J Lee. Nexus: Practical and secure access control on untrusted storage platforms using client-side sgx. In *2019 49th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pages 401–413. IEEE, 2019.
[30] R. A. Fisher. Xv.—the correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52(2):399–433, 1919.
[31] Ronald Aylmer Fisher. Statistical methods for research workers. In *Breakthroughs in statistics*, pages 66–70. Springer, 1992.
[32] Arturo Flores and Serge Belongie. Removing pedestrians from google street view images. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pages 53–58. IEEE, 2010.
[33] Julia Grenberg. Photobucket breach floods web with racy images. *https://www.cnn.com/2012/08/09/tech/photobucket-privacy-breach*, 2012.
[34] Rebecca Gulotta, William Odom, Jodi Forlizzi, and Haakon Faste. Digital artifacts as legacy: exploring the lifespan and value of digital data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1813–1822, 2013.
[35] Seda Gürses and Claudia Diaz. Two tales of privacy in online social networks. *IEEE Security & Privacy*, 11(3):29–37, 2013.
[36] Rakibul Hasan, Eman Hassan, Yifang Li, Kelly Caine, David J Crandall, Roberto Hoyle, and Apu Kapadia. Viewer experience of obscuring scene elements in photos to enhance privacy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018.
[37] Rakibul Hasan, Yifang Li, Eman Hassan, Kelly Caine, David J Crandall, Roberto Hoyle, and Apu Kapadia. Can privacy be satisfying? on improving viewer satisfaction for privacy-enhanced photos using aesthetic transforms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2019.
[38] Rakibul Hasan, Patrick Shaffer, David Crandall, Eman T Apu Kapadia, et al. Cartooning for enhanced privacy in lifelogging and streaming videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 29–38, 2017.
[39] Jianping He, Bin Liu, Deguang Kong, Xuan Bao, Na Wang, Hongxia Jin, and George Kesidis. Puppies: Transformation-supported personalized privacy preserving partial image sharing. In *2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pages 359–370, 2016.

[40] Donell Holloway and Lelia Green. Mediated memory making: The virtual family photograph album. *Communications*, 42(3):351–368, 2017.
[41] Roberto Hoyle. *Privacy considerations in the context of wearable cameras*. PhD thesis, Indiana University, 2016.
[42] Roberto Hoyle, Luke Stark, Qatrunnada Ismail, David Crandall, Apu Kapadia, and Denise Anthony. Privacy norms and preferences for photos posted online. *ACM Transactions on Computer-Human Interaction (ACM TOCHI)*, 27(4), August 2020.
[43] Roberto Hoyle, Robert Templeman, Denise Anthony, David Crandall, and Apu Kapadia. Sensitive lifelogs: A privacy analysis of photos from wearable cameras. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 1645–1648, New York, NY, USA, 2015. Association for Computing Machinery.
[44] Roberto Hoyle, Robert Templeman, Steven Armes, Denise Anthony, David Crandall, and Apu Kapadia. Privacy behaviors of lifeloggers using wearable cameras. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '14, page 571–582, New York, NY, USA, 2014. Association for Computing Machinery.
[45] Panagiotis Ilia, Iasonas Polakis, Elias Athanasopoulos, Federico Maggi, and Sotiris Ioannidis. Face/off: Preventing privacy leakage from photos in social networks. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pages 781–792, 2015.
[46] Nicolas Jacquemet, Alexander G James, Stéphane Luchini, James J Murphy, and Jason F Shogren. Do truth-telling oaths improve honesty in crowd-working? *PloS one*, 16(1):e0244958, 2021.
[47] Yousra Javed, Mohamed Shehab, and Emmanuel Bello-Ogunu. Investigating user comprehension and risk perception of apple's touch id technology. In *Proceedings of the 12th International Conference on Availability, Reliability and Security*, pages 1–6, 2017.
[48] Lukasz Jedrzejczyk, Blaine A Price, Arosha K Bandara, and Bashar Nuseibeh. On the impact of real-time feedback on users' behaviour in mobile location-sharing applications. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, pages 1–12, 2010.
[49] Nathan Jurgenson. Temporary social media. *https://newsroom.snap.com/temporary-social-media/*, 2013.
[50] James E Katz and Elizabeth Thomas Crocker. Selfies| selfies and photo messaging as visual conversation: Reports from the united states, united kingdom and china. *International Journal of Communication*, 9:12, 2015.
[51] Zubeida Casmod Khan, Thulani Mashiane, and Nobubele A Shozi. Snapchat media retrieval for novice device users. In *Proceedings of the 10th International Conference on Cyber Warfare and Security*, pages 162–169, 2015.
[52] Bart Piet Knijnenburg and Alfred Kobsa. Increasing sharing tendency without reducing satisfaction: Finding the best privacy-settings user interface for social networks. In *ICIS*, 2014.
[53] Iryna Korshunova, Wenzhe Shi, Joni Dambre, and Lucas Theis. Fast face-swap using convolutional neural networks. In *Proceedings of the IEEE international conference on computer vision*, pages 3677–3685, 2017.
[54] Balachander Krishnamurthy and Craig E Wills. On the leakage of personally identifiable information via online social networks. In *Proceedings of the 2nd ACM workshop on Online social networks*, pages 7–12, 2009.
[55] Oksana Kulyk, Kristina Milanovic, and Jeremy Pitt. Does my smart device provider care about my privacy? investigating trust factors and user attitudes in iot systems. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, pages 1–12, 2020.
[56] Daniel Lakens. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and anovas. *Frontiers in Psychology*, 4:863, 2013.
[57] Alex Leavitt. "this is a throwaway account": Temporary technical identities and perceptions of anonymity in a massive online community. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work amp; Social Computing*, CSCW '15, page 317–327, New York, NY, USA, 2015. Association for Computing Machinery.
[58] Wenjie Li, Rongrong Ni, and Yao Zhao. Jpeg photo privacy-preserving algorithm based on sparse representation and data hiding. In *International Conference on Image and Graphics*, pages 575–586. Springer, 2017.
[59] Yifang Li, Nishant Vishwamitra, Hongxin Hu, and Kelly Caine. Towards a taxonomy of content sensitivity and sharing preferences for photos. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2020.
[60] Yifang Li, Nishant Vishwamitra, Bart P Knijnenburg, Hongxin Hu, and Kelly Caine. Effectiveness and users' experience of obfuscation as a privacy-enhancing technology for sharing photos. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):1–24, 2017.
[61] Yabing Liu, Krishna P Gummadi, Balachander Krishnamurthy, and Alan Mislove. Analyzing facebook privacy settings: user expectations vs. reality. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 61–70, 2011.
[62] Stine Lomborg and Anja Bechmann. Using apis for data collection on social media. *The Information Society*, 30(4):256–265, 2014.

[63] Andrés Lucero, Jussi Holopainen, and Tero Jokela. Pass-them-around: collaborative use of mobile phones for photo sharing. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1787–1796, 2011.
[64] Michelle Madejski, Maritza Johnson, and Steven M Bellovin. A study of privacy settings errors in an online social network. In *2012 IEEE International Conference on Pervasive Computing and Communications Workshops*, pages 340–345. IEEE, 2012.
[65] Aqdas Malik, Amandeep Dhir, and Marko Nieminen. Uses and gratifications of digital photo sharing on facebook. *Telematics and Informatics*, 33(1):129–138, 2016.
[66] Richard McPherson, Reza Shokri, and Vitaly Shmatikov. Defeating image obfuscation with deep learning. *arXiv preprint arXiv:1609.00408*, 2016.
[67] Reham Ebada Mohamed and Sonia Chiasson. Online privacy and aging of digital artifacts. In *Fourteenth Symposium on Usable Privacy and Security ({SOUPS} 2018)*, pages 177–195, 2018.
[68] Vivian Genaro Motti and Kelly Caine. Users' privacy concerns about wearables. In *International Conference on Financial Cryptography and Data Security*, pages 231–244. Springer, 2015.
[69] Alexios Mylonas, Dimitris Gritzalis, Bill Tsoumas, and Theodore Apostolopoulos. A qualitative metrics vector for the awareness of smartphone security users. In *International Conference on Trust, Privacy and Security in Digital Business*, pages 173–184. Springer, 2013.
[70] Angelo Nodari, Marco Vanetti, and Ignazio Gallo. Digital privacy: Replacing pedestrians from google street view images. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 2889–2893. IEEE, 2012.
[71] Anne Oeldorf-Hirsch and S Shyam Sundar. Social and technological motivations for online photo sharing. *Journal of Broadcasting & Electronic Media*, 60(4):624–642, 2016.
[72] Seong Joon Oh, Rodrigo Benenson, Mario Fritz, and Bernt Schiele. Faceless person recognition: Privacy implications in social media. In *European Conference on Computer Vision*, pages 19–35. Springer, 2016.
[73] Fernando Kaway Carvalho Ota, Jorge Augusto Meira, Cyril Renaud Cassagnes, and Radu State. Mobile app to sgx enclave secure channel. In *2019 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*, pages 258–263. IEEE, 2019.
[74] Global Platform. 'tee system architecture. *Global Platform technical overview*, 2011.
[75] Moo-Ryong Ra, Ramesh Govindan, and Antonio Ortega. P3: Toward privacy-preserving photo sharing. In *10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13)*, pages 515–528, Lombard, IL, April 2013. USENIX Association.
[76] Yasmeen Rashidi, Tousif Ahmed, Felicia Patel, Emily Fath, Apu Kapadia, Christena Nippert-Eng, and Norman Makoto Su. " you don't want to be the next meme": College students' workarounds to manage privacy in the era of pervasive photography. In *Fourteenth Symposium on Usable Privacy and Security ({SOUPS} 2018)*, pages 143–157, 2018.
[77] Mohamed Sabt, Mohammed Achemlal, and Abdelmadjid Bouabdallah. Trusted execution environment: what it is, and what it is not. In *2015 IEEE Trustcom/BigDataSE/ISPA*, volume 1, pages 57–64. IEEE, 2015.
[78] Lynsay A Shepherd, Jacqueline Archibald, and Robert Ian Ferguson. Assessing the impact of affective feedback on end-user security awareness. In *International Conference on Human Aspects of Information Security, Privacy, and Trust*, pages 143–159. Springer, 2017.
[79] Jiayu Shu, Rui Zheng, and Pan Hui. Cardea: Context-aware visual privacy protection for photo taking and sharing. In *Proceedings of the 9th ACM Multimedia Systems Conference*, pages 304–315, 2018.
[80] M Six Silberman, Bill Tomlinson, Rochelle LaPlante, Joel Ross, Lilly Irani, and Andrew Zaldivar. Responsible research with crowds: pay crowdworkers at least minimum wage. *Communications of the ACM*, 61(3):39–41, 2018.
[81] Kit Smith. 53 incredible facebook statistics and facts. *https://www.brandwatch.com/blog/facebook-statistics/*, 2019.
[82] Vidi Sukmayadi and Azizul Halim Yahya. Impression management within a phenomenological study. *The Open Psychology Journal*, 12(1), 2019.
[83] Qianru Sun, Liqian Ma, Seong Joon Oh, Luc Van Gool, Bernt Schiele, and Mario Fritz. Natural and effective obfuscation by head inpainting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5050–5059, 2018.
[84] Leman Pınar Tosun. Motives for facebook use and expressing "true self" on the internet. *Computers in human behavior*, 28(4):1510–1517, 2012.
[85] Lam Tran, Deguang Kong, Hongxia Jin, and Ji Liu. Privacy-cnh: A framework to detect photo privacy with convolutional neural network using hierarchical features. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.
[86] Sonja Utz, Nicole Muscanell, and Cameran Khalid. Snapchat elicits more jealousy than facebook: A comparison of snapchat and facebook use. *Cyberpsychology, Behavior, and Social Networking*, 18(3):141–146, 2015.
[87] T Franklin Waddell. The allure of privacy or the desire for self-expression? identifying users' gratifications for ephemeral, photograph-based communication. *Cyberpsychology, Behavior, and Social Networking*, 19(7):441–445, 2016.

[88] Wenbo Wang, Hean Tat Keh, and Lisa E Bolton. Lay theories of medicine and a healthy lifestyle. *Journal of Consumer Research*, 37(1):80–97, 2010.

[89] Valerie SL Williams, Lyle V Jones, and John W Tukey. Controlling error in multiple comparisons, with examples from state-to-state differences in educational achievement. *Journal of educational and behavioral statistics*, 24(1):42–69, 1999.

[90] Pamela Wisniewski, AKM Islam, Heather Richter Lipford, and David C Wilson. Framing and measuring multidimensional interpersonal privacy preferences of social networking site users. *Communications of the Association for information systems*, 38(1):10, 2016.

[91] Kaihe Xu, Yuanxiong Guo, Linke Guo, Yuguang Fang, and Xiaolin Li. My privacy my decision: Control of photo sharing on online social networks. *IEEE Transactions on Dependable and Secure Computing*, 14(2):199–210, 2015.

[92] Toshio Yamagishi. The provision of a sanctioning system as a public good. *Journal of Personality and social Psychology*, 51(1):110, 1986.

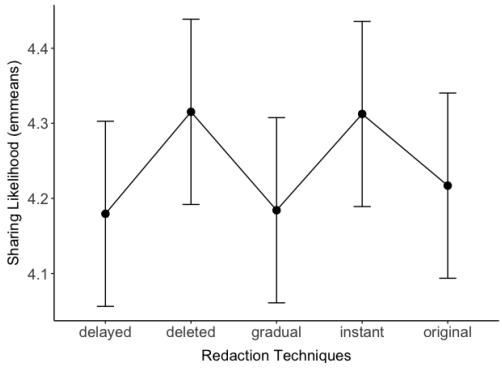
[93] Toshio Yamagishi and Midori Yamagishi. Trust and commitment in the united states and japan. *Motivation and emotion*, 18(2):129–166, 1994.

[94] Sen-Chi Yu and Hong-Ren Chen. Ephemeral but influential? the correlation between facebook stories usage, addiction, narcissism, and positive affect. In *Healthcare*, volume 8, page 435. Multidisciplinary Digital Publishing Institute, 2020.

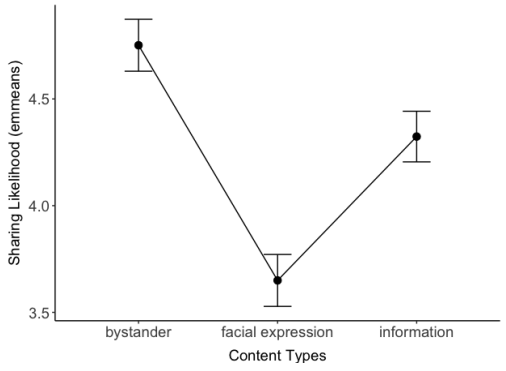
[95] Lin Yuan, Pavel Korshunov, and Touradj Ebrahimi. Secure jpeg scrambling enabling privacy in photo sharing. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, volume 4, pages 1–6. IEEE, 2015.

APPENDIX

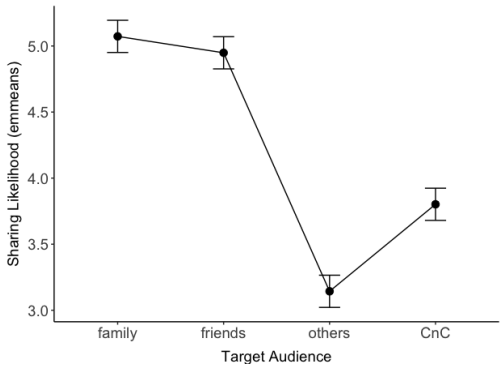
A EXTRA FIGURES



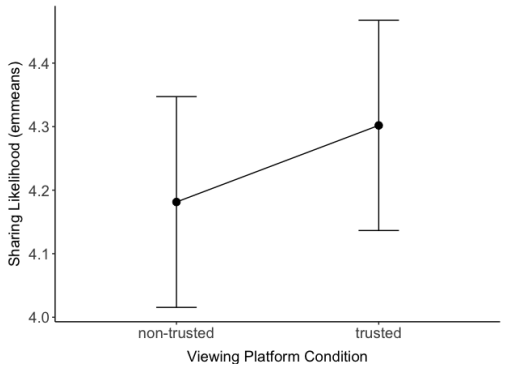
(a) Redaction techniques



(b) Content types



(c) Target audiences. CnC= Colleagues and Classmates.



(d) Viewing platform condition

Fig. 4. Differences in sharing likelihood (emmeans) for different variables

B THE SURVEY

Display the Captcha, pledge, and consent form. Then display screening questions.

B.1 Social Media Usage Questionnaire

- **Q3.1:** Which social media platforms do you have an account for? (Select all that apply.)
  1. Facebook 2. Instagram 3. Pinterest 4. Snapchat 5. Twitter 6. Myspace 7. Flicker 8. Reddit 9. Tiktok 10. LinkedIn 11. Imgur 12. Other (Please describe) 13. I do not use social media
- **Q3.2:** How often you visit social media?
  1. Never, 2. Less than once in a month, 3. Once in a month, 4. Multiple times in a month, 5. Once in a week, 6. Multiple times in a week, 7. Once in a day, 8. Multiple times in a day
- **Q3.3:** What social media platform do you use to share photos online the most? (Select all that apply.)
  1. Facebook 2. Instagram 3. Pinterest 4. Snapchat 5. Twitter 6. Myspace 7. Flicker 8. Reddit 9. Tiktok 10. LinkedIn 11. Imgur 12. Other (Please describe)

Fig. 5. Instant redaction sample (Original Photo by Nima Sarram on Unsplash)

- **Q3.4:** When you share photos online, who do you typically share them with? (Select all that apply.)
  1. Family members 2. Friends 3. Colleagues and classmates 4. Other followers

B.2 Image Sharing Behavior Questionnaire

- **Q4.1:** How often do you share photos on social media?
- **Q4.2:** How often do you share pictures taken by you, your friends, or your family on the most frequently used social media?
- **Q4.3:** How often do you share pictures on social media that you found on the internet or that other people took (not including your friends, family or other people you personally know.)?

B.3 Scenario Description

B.3.1 Trusted and Non-Trusted Device.

Consider a situation where you have taken a photo and would like to share it with others. You are aware that the photo contains some unusual or sensitive content. Some apps like Snapchat allow the viewer to look at a photo posted to one's "story" (a collection of photos for that day) for 24 hours, after which the viewer's app deletes the photo. Imagine a similar social media app that, instead of deleting the photo, obfuscates (hides) parts of the photo from viewers after a certain period (instead of deleting the photo). However, the viewer of the photo might attempt to save the photo or screenshot the original photo before the obfuscations are applied.

Trusted Device scenario: Assume the recipient's phone is equipped with a special security chip that will make it impossible for the viewer to save or take a screenshot of the photo.

Non-Trusted Device scenario: Assume the recipient's social media app does not allow the viewer to save the photo and (like Snapchat) notifies the sender of the photo if the viewer takes a screenshot.

B.3.2 Temporal Redaction Scenario.

Instantaneous Masking: Parts of your photo will be masked before sharing the photo.

Fig. 6. Gradual redaction sample (Original Photo by Nima Sarram on Unsplash)

Fig. 7. Delayed redaction sample (Original Photo by Nima Sarram on Unsplash)

Gradual Redaction: The original photo will be shared, but parts of your photo will be gradually blurred over time by the recipient's phone. Eventually, the content will be completely masked and be totally obscured.

Delayed Masking: The original photo will be shared and then eventually the content will be completely masked and totally obscured in the recipient's phone after 24 hours. Until the content is masked, the entire photo is visible on the recipient's phone.

B.3.3 General Instruction.

For each photo, please do your best to answer each question as honestly and accurately as possible.

Any information you enter is completely anonymous and will not be connected to your identity.

B.4 Experimental Manipulation

How likely are you to share the photo in the following scenario on social media with these groups of people? i) Family members ii) Friends iii) Colleagues and classmates iv) Other followers

1.Extremely unlikely, 2. Moderately unlikely, 3. Slightly unlikely, 4. Neither unlikely nor likely, 5. Slightly likely, 6. Moderately likely, 7. Extremely likely

B.4.1 Temporal Redaction Questions.
These questions were shown in random order. The later part of each question mentioned "by the app" to participants with non-trusted device condition and "by the security chip in the recipient's phone" to participants with trusted device condition.

- Assume this is a photo of you and your car, and you would like to share it with the license plate of the car being masked by the app/ by the security chip in the recipient's phone after 24 hours in the recipient's device.
- Assume this is a photo of the front of your house, and you would like to share it with the house number being obscured by the app/ by the security chip in the recipient's phone.
- Assume this is a photo of your face, and you would like to share it with the face being obscured by the app/ by the security chip in the recipient's phone gradually over 24 hours in the recipient's device.
- Assume this is a photo of you and your car, and you would like to share it with the license plate of the car being obscured by the app/ by the security chip in the recipient's phone gradually over 24 hours in the recipient's device.
- Assume this is a photo of your face, and you would like to share it with the face being obscured by the app/ by the security chip in the recipient's phone.
- Assume this is a photo of you with another person behind you, and you would like to share it with the face of that person being masked by the app/ by the security chip in the recipient's phone after 24 hours in the recipient's device.
- Assume this is a photo of you and your car, and you would like to share it with the license plate of the car being obscured by the app/ by the security chip in the recipient's phone.
- Assume this is a photo of your face, and you would like to share it with the face being masked by the app/ by the security chip in the recipient's phone after 24 hours in the recipient's device.
- Assume this is a photo of you with another person behind you, and you would like to share it with the face of that person being obscured by the app/ by the security chip in the recipient's phone gradually over 24 hours in the recipient's device.
- Assume this is a photo of the front of your house, and you would like to share it with the house number being masked by the app/ by the security chip in the recipient's phone after 24 hours in the recipient's device.
- Assume this is a photo of you with another person behind you, and you would like to share it with the face of that person being obscured by the app/ by the security chip in the recipient's phone.
- Assume this is a photo of the front of your house, and you would like to share it with the house number being obscured by the app/ by the security chip in the recipient's phone gradually over 24 hours in the recipient's device.

B.4.2 Baseline Questions.
These questions were showed in random order. The later part of each question mentioned "by the app" to participants with non-trusted device condition and "by the security chip in the recipient's phone" to participants with trusted device condition.

- Assume this is a photo of you and your car, and the photo will be deleted automatically by the app/ by the security chip in the recipient's phone 24 hours after the recipient views it.
- Assume this is a photo of the front of your house, and you would like to share it without the house number being obscured.
- Assume this is a photo of you with another person behind you, and you would like to share it without the face of that person being obscured.

- Assume this is a photo of the front of your house, and the photo will be deleted automatically by the app/ by the security chip in the recipient's phone 24 hours after the recipient views it.
- Assume this is a photo of you and your car, and you would like to share it without the license plate of the car being obscured.
- Assume this is a photo of your face, and you would like to share it without the face being obscured.
- Assume this is a photo of you with another person behind you, and the photo will be deleted automatically by the app/ by the security chip in the recipient's phone 24 hours after the recipient views it.
- Assume this is a photo of your face, and the photo will be deleted automatically by the app/ by the security chip in the recipient's phone 24 hours after the recipient views it.

B.5 Open Ended Attention Check Question

Q5: Please briefly describe why your sharing preference might have varied for different obfuscations for the last question (minimum 150 characters).

B.6 Privacy Behavior Questionnaire

Answer each of the questions below with options: i) Yes ii) Maybe iii) No

- **Q6.1:** Has anyone ever shared a picture of you online that you did not want them to share?
- **Q6.2:** Has anyone ever shared a picture of you online that you felt violated your privacy?
- **Q6.3:** Have you ever been embarrassed by a picture of yourself that has been posted online?
- **Q6.4:** Have you ever regretted posting a picture of yourself online?
- **Q6.5:** Please select the last option.
- **Q6.6:** Have you ever accidentally posted a picture of yourself online that you did not want to share?
- **Q6.7:** Have you ever shared an embarrassing picture online of someone else you know?
- **Q6.8:** Have you ever regretted posting a picture online of someone else you know?
- **Q6.9:** Have you ever posted a picture online of someone else you know, which may have violated his or her privacy?
- **Q6.10:** Have you ever shared an embarrassing picture online of a stranger (someone that you do not personally know)?
- **Q6.11:** Have you ever regretted posting a picture online of a stranger (i.e., someone you do not personally know)?
- **Q6.12:** Have you ever posted a picture of a stranger (i.e., someone you do not personally know), which may have violated his or her privacy?
- **Q6.13:** Do people you know post pictures that might be embarrassing to other people?
- **Q6.14:** Has anyone you know regretted posting a picture of another person?
- **Q6.15:** Has anyone you know regretted posting a picture of themselves?
- **Q6.16:** Has anyone you know posted a picture that may have violated someone's privacy?

B.7 Privacy Preference Question

Q7: Are you a private person who keeps to yourself or an open person who enjoys sharing with others? 1) Very Private ... 7) Very Open

B.8 Inter Personal Trust Questionnaire

1) Strongly Disagree... 5) Strongly Agree

B.8.1 A 6-item questionnaire by Yamagishi et al. to measure beliefs about honesty and trustworthiness of others [93].

- Most people are basically honest
- Most people are trustworthy
- Most people are basically good and kind
- Most people are trustful of others
- I am trustful
- Most people will respond in kind when they are trusted by others

B.8.2 A 5-item questionnaire by Yamagishi et al. to measure the risk associated with trusting others [92]. One item (Most people are basically honest) is present in the previous scale. So, it has been omitted here.

- Most people tell a lie when they can benefit by doing so
- Those devoted to unselfish causes are often exploited by others
- Some people do not cooperate because they pursue only their own short-term selfinterest. Thus, things that can be done well if people cooperate often fail because of these people
- There will be more people who will not work if the social security system is developed further

B.9 Demographic Questions

- **Q9.1:** Please select your gender i) Male ii)Female iii) Would prefer not to answer iv) Other (text input)
- **Q9.2:** How long have you lived in the United States? i) I don't live in the United States ii) Less than 1 year iii) 1 year and 2 years iv) 2 years- 3 years v) 3 years- 4 years v) 4 years- 5 years vii) 5 years or more.
- **Q9.3:** What's your age? i) Under 18 years ii) 18-29 iii) 30-49 iv) 50-64 v)65 or more.
- **Q9.4:** Please select the highest level of education that you have achieved i) None ii) 1st-4th grade iii) 5th-8th grade iv) 9th-12th grade v) High school graduate or GED vi) Some college, no degree vii) Associate's degree viii) Bachelor's degree ix) Master's degree x) Professional (e.g., MD, JD) degree xi) Doctoral degree.
- **Q9.5:** What is your primary racial or ethnic background? Please select all that apply. i) Hispanic or Latino ii) American Indian or Alaskan Native iii) Asian iv) Black or African American v) Native Hawaiian or Other Paciic Islander vi) White vii) Other (text input).

Received January 2022; Revised April 2022; Accepted May 2022