

Clemson University

**TigerPrints**

---

All Dissertations

Dissertations

---

August 2020

## Investigating Obfuscation as a Tool to Enhance Photo Privacy on Social Networks Sites

Yifang Li

*Clemson University*, [yvonne426@gmail.com](mailto:yvonne426@gmail.com)

Follow this and additional works at: [https://tigerprints.clemson.edu/all\\_dissertations](https://tigerprints.clemson.edu/all_dissertations)

---

### Recommended Citation

Li, Yifang, "Investigating Obfuscation as a Tool to Enhance Photo Privacy on Social Networks Sites" (2020). *All Dissertations*. 2694.

[https://tigerprints.clemson.edu/all\\_dissertations/2694](https://tigerprints.clemson.edu/all_dissertations/2694)

This Dissertation is brought to you for free and open access by the Dissertations at TigerPrints. It has been accepted for inclusion in All Dissertations by an authorized administrator of TigerPrints. For more information, please contact [kokeefe@clemson.edu](mailto:kokeefe@clemson.edu).

# INVESTIGATING OBFUSCATION AS A TOOL TO ENHANCE PHOTO PRIVACY ON SOCIAL NETWORKS SITES

---

A Dissertation  
Presented to  
the Graduate School of  
Clemson University

---

In Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy  
Human-Centered Computing

---

by  
Yifang Li  
August 2020

---

Accepted by:  
Dr. Kelly Caine, Committee Chair  
Dr. Hongxin Hu  
Dr. Apu Kapadia  
Dr. Bart Knijnenburg  
Dr. Nathan McNeese

# Abstract

Photos which contain rich visual information can be a source of privacy issues. Some privacy issues associated with photos include identification of people, inference attacks, location disclosure, and sensitive information leakage. However, photo privacy is often hard to achieve because the content in the photos is both what makes them valuable to viewers, and what causes privacy concerns.

Photo sharing often occurs via Social Network Sites (SNSs). Photo privacy is difficult to achieve via SNSs due to two main reasons: first, SNSs seldom notify users of the sensitive content in their photos that might cause privacy leakage; second, the recipient control tools available on SNSs are not effective.

The only solution that existing SNSs (e.g., Facebook, Flickr) provide is control over who receives a photo. This solution allows users to withhold the entire photo from certain viewers while sharing it with other viewers. The idea is that if viewers cannot see a photo, then privacy risk is minimized. However, withholding or self-censoring photos is not always the solution people want. In some cases, people want to be able to share photos, or parts of photos, even when they have privacy concerns about the photo.

To provide better online photo privacy protection options for users, we leverage a behavioral theory of privacy that identifies and focuses on two key elements that influence privacy – information content and information recipient. This theory provides a vocabulary for discussing key aspects of privacy and helps us organize our research to focus on the two key parameters through a series of studies.

In my thesis, I describe five studies I have conducted. First, I focus on the content parameter to identify what portions of an image are considered sensitive and therefore are candidates to be obscured to increase privacy. I provide a taxonomy of content sensitivity that can help designers of photo-privacy mechanisms understand what categories of content users consider sensitive. Then,

focusing on the recipient parameter, I describe how elements of the taxonomy are associated with users' sharing preferences for different categories of recipients (e.g., colleagues vs. family members).

Second, focusing on controlling photo content disclosure, I invented privacy-enhancing obfuscations and evaluated their effectiveness against human recognition and studied how they affect the viewing experience.

Third, after discovering that avatar and inpainting are two promising obfuscation methods, I studied whether they were robust when de-identifying both familiar and unfamiliar people since viewers are likely to know the people in OSN photos. Additionally, I quantified the prevalence of self-reported photo self-censorship and discovered that privacy-preserving obfuscations might be useful for combating photo self-censorship.

Gaining sufficient knowledge from the studies above, I proposed a privacy-enhanced photo-sharing interface that helps users identify the potential sensitive content and provides obfuscation options. To evaluate the interface, I compared the proposed obfuscation approach with the other two approaches – a control condition that mimics the current Facebook photo-sharing interface and an interface that provides a privacy warning about potentially sensitive content. The results show that our proposed system performs better over the other two in terms of reducing perceived privacy risks, increasing willingness to share, and enhancing usability. Overall, our research will benefit privacy researchers, online social network designers, policymakers, computer vision researchers, and anyone who has or wants to share photos online.

# Acknowledgments

Firstly, I would like to express my sincere gratitude to my advisor Dr. Kelly Caine for the continuous support of my Ph.D. research and my life during the past five years. I really appreciate that she has given me the freedom to choose from various projects based on my interests at the first beginning, and helped me shaped and polished each of the studies. Without her guidance and constant feedback, this Ph.D. would not have been achievable.

I greatly appreciate the support from my committee members. I thank Dr. Bart Knijnenburg for his insightful suggestions that improved the design of all my previous studies and his help on data analysis. My sincere thanks also go to Dr. Hongxin Hu and his student Nishant Vishwamitra for their efforts on paper and rebuttal writing. I am very grateful to Dr. Apu Kapadia for getting me involved in his project about online photo privacy protection which greatly benefits my own research. I thank Dr. Nathan McNeese for all the recommendations he gave during my proposal presentation.

I thank various undergraduate and graduate student research assistants who contributed to my research for all the excellent work that they have done, as well as all of my labmates from HATlab for their great suggestions on my studies during each lab meeting.

Last but not the least, I would like to thank my family – my partner, my parents, my cat Jiejie, and my dog Nash for supporting me spiritually throughout my Ph.D.

Financial support was provided by National Science Foundation under grand no.1527421.

# Table of Contents

<b>Title Page</b> . . . . .	<b>i</b>
<b>Abstract</b> . . . . .	<b>ii</b>
<b>Acknowledgments</b> . . . . .	<b>iv</b>
<b>List of Tables</b> . . . . .	<b>vii</b>
<b>List of Figures</b> . . . . .	<b>ix</b>
<b>1 Introduction</b> . . . . .	<b>1</b>
1.1 Problem Motivation . . . . .	1
1.2 Research Motivation . . . . .	2
1.3 Research Objectives . . . . .	3
1.4 Overview & Summary of Studies . . . . .	4
<b>2 Background and Related Work</b> . . . . .	<b>10</b>
2.1 Introduction to Online Photo Privacy . . . . .	10
2.2 Theories of Privacy in HCI . . . . .	16
2.3 In The Context of Online Photo Privacy . . . . .	17
2.4 Summary of the Literature Review . . . . .	23
<b>3 Study 1: Identifying Sensitive Content and Users' Sharing Preference</b> . . . . .	<b>25</b>
3.1 Introduction . . . . .	25
3.2 Method . . . . .	27
3.3 Results . . . . .	31
3.4 Discussion . . . . .	39
3.5 Limitation and Future Work . . . . .	44
3.6 Chapter Conclusion . . . . .	46
<b>4 Study 2: Identifying Effective and Usable Obfuscations</b> . . . . .	<b>47</b>
4.1 Introduction . . . . .	47
4.2 Method . . . . .	48
4.3 Results . . . . .	53
4.4 Discussion . . . . .	64
4.5 Limitations . . . . .	71
4.6 Chapter Conclusion . . . . .	72
<b>5 Study 3: De-identifying Familiar and Unfamiliar People</b> . . . . .	<b>74</b>
5.1 Introduction . . . . .	74
5.2 Method . . . . .	76
5.3 Results . . . . .	83

5.4	Discussion . . . . .	88
5.5	Limitations and Future Work . . . . .	94
5.6	Chapter Conclusion . . . . .	95
<b>6</b>	<b>Study 4: Obfuscation May Combat Self-Censorship . . . . .</b>	<b>96</b>
6.1	Introduction . . . . .	96
6.2	Method . . . . .	98
6.3	Results . . . . .	100
6.4	Discussion . . . . .	102
6.5	Limitations and Future Work . . . . .	105
6.6	Chapter Conclusion . . . . .	105
<b>7</b>	<b>Study 5: An Experiment to Determine Whether Obfuscation Reduces Privacy Concerns and Increases Willingness to Share . . . . .</b>	<b>107</b>
7.1	Introduction . . . . .	107
7.2	Photo Privacy Protection Interface Design . . . . .	108
7.3	Method . . . . .	109
7.4	Results . . . . .	120
7.5	Discussion . . . . .	130
7.6	Limitation . . . . .	133
7.7	Chapter Conclusion . . . . .	133
<b>8</b>	<b>Discussion of all Five Studies of the Dissertation . . . . .</b>	<b>135</b>
8.1	Contributions . . . . .	136
8.2	Impact on Privacy Research . . . . .	137
8.3	Impact on Industry . . . . .	139
8.4	Practical Considerations for System Implementation . . . . .	140
8.5	Ethical Issues Related to Obfuscation . . . . .	141
	<b>Appendices . . . . .</b>	<b>143</b>
A	Study 1 Online Experiment Questions . . . . .	144
B	Study 2 Online Experiment Questions . . . . .	156
C	Study 3 and 4 Online Experience Questions . . . . .	173
D	Study 5 Online Experience Questions . . . . .	190
	<b>Bibliography . . . . .</b>	<b>213</b>

# List of Tables

1.1	Blumenfeld chart summarizing the five studies in this dissertation . . . . .	5
1.2	Blumenfeld chart summarizing the five studies in this dissertation (coninued). . . . .	6
2.1	Summary of systematic literature review . . . . .	11
2.2	Summary of previously identified sensitive categories in photos . . . . .	21
2.3	Eight obfuscation methods. . . . .	22
2.4	Summary of recipients from prior literature . . . . .	23
3.1	Recipient groups used in our study . . . . .	27
3.2	Twenty-eight sensitive categories with the number of data points in each category and their examples. Each word or phrase in the example column represents a unique piece of sensitive content, as identified and named by participants, in response to the open ended question, “What content in this photo do you consider sensitive?”. . . . .	36
4.1	Identification success rate, odds ratio, 95% confidence interval, and p-value by region and obfuscation for all cases where the <i>as is</i> is the baseline. The obfuscations are ordered by identification success of body region from lowest (most effective) to highest (least effective). . . . .	54
4.2	Identification confidence for Hit, Miss, Correct Rejection, False Alarm, Total Correct (Hit + Correct Rejection), and Total Wrong (Miss + False Alarm) on a scale from 1 - 7 where 7 is most confident. Standard deviations appear in parentheses beside the means. Within face and body categories, the order of the obfuscations is from most to least effective. . . . .	57
4.3	Obfuscation preference, willingness to use, and preference given privacy concerns. Standard deviations appear in parentheses beside the means. Obfuscations are ordered from most to least effective. . . . .	63
4.4	Summary of photo obfuscation methods (body-obfuscations only because they are more effective; see “Effectiveness: Face-obscuring vs. Body-obscuring.”) Effectiveness is defined by the difference in the identification success percentage of <i>as is</i> and each body obfuscation (see Table 4.1). The misidentification of <i>as is</i> is 23% (100% minus 77%). An obfuscation that achieves at least twice of <i>as is</i> misidentification (46%) is defined as “Somewhat effective”, so the identification success should be no more than 54%. An obfuscation that achieves at least three times of <i>as is</i> misidentification (69%) is considered “Effective”, so the identification success should be at most 31%. Obfuscations are ordered from most to least effective. . . . .	65
5.1	Six obfuscation methods. In the example figures, we applied the methods on familiar people. They were also applied on unfamiliar people in the study, yielding in 14 conditions (We added <i>as is</i> as the baseline condition). . . . .	76
5.2	Participants’ familiarity with the famous people in our stimuli. The three columns show the famous people’s names, percentage of being named, and means of familiarity with standard deviations. . . . .	79



5.3	Identification rate in all cases (including both target present and absent), odds ratio and p-value between familiar and unfamiliar cases for each obfuscation, and odds ratio and p-value between each obfuscation and the baseline <i>as is</i> regardless of familiarity. The obfuscations are ordered by identification rate of total cases (familiarity + unfamiliar) from lowest (most effective) to highest (least effective). . . . .	82
5.4	Obfuscation preference, willingness to use, and preference given privacy concerns. Standard deviations appear in parentheses beside the means. Obfuscations are ordered from most to least effective. . . . .	87
6.1	Six obfuscation methods. Study participants were shown images, but not provided name, definition, or related work citations. . . . .	99
6.2	Coefficient, standard error, and odds ratio of photo censorship model (if participants have declined to share a photo due to privacy concern) . . . . .	101
6.3	Coefficient, standard error, and odds ratio of the model of willingness to share the photo again with obfuscations applied . . . . .	102
7.1	Participants' demographics in step one. . . . .	113
7.2	Participants' demographics in step two. . . . .	117
7.3	Pre and post-test measurements. . . . .	119
7.4	Percentages of each recipient group and whether or not participants posted photos. .	127
7.5	Percentages of obfuscation options selected. . . . .	128

# List of Figures

1.1	Illustration of different phases of this dissertation. . . . .	7
2.1	Content by recipient interaction described in the behavioral privacy model [50]. . . .	19
3.1	A part of the dendrogram. All items in this sort are listed vertically. Items placed next to each other vertically are more similar. The horizontal line from each item joins other items vertically, showing where items are grouped at higher levels of relationship [21].	32
3.2	Participants' likelihood to share each sensitive content category across all recipient groups. . . . .	37
3.3	Participants' likelihood to share with each recipient across all sensitive content categories. . . . .	38
3.4	Example interface of content detection and obfuscation. . . . .	45
4.1	Experiment interface with one stimuli and ID photo examples . . . . .	51
4.2	Means and standard errors of identification confidence of Total Correct (Hit + Correct Rejection) and Total Wrong (Miss + False Alarm). . . . .	56
4.3	Photo satisfaction rating ( $M$ and $SE$ ). Obfuscations ordered from most to least effective.	58
4.4	Information sufficiency rating ( $M$ and $SE$ ). Obfuscations ordered from most to least effective. . . . .	59
4.5	Enjoyment rating ( $M$ and $SE$ ). Obfuscations ordered from most to least effective. .	60
4.6	Social presence rating ( $M$ and $SE$ ). Obfuscations ordered from most to least effective.	61
4.7	Obfuscation likability ( $M$ and $SE$ ) from most to least effective. . . . .	62
4.8	Scatterplot of Likability (X axis) against Identification Success (Y axis). This plot shows the general trade-off between effectiveness and user experience. However, body avatar and body inpainting are outliers. They are both effective and provide a good user experience. . . . .	68
5.1	An example of a blurred familiar person . . . . .	75
5.2	Experiment interface with one stimuli and ID photo examples . . . . .	80
5.3	Means and standard errors of identification confidence of Total Correct and Total Wrong separated by familiarity. . . . .	83
5.4	Photo satisfaction rating ( $M$ and $SE$ ). Obfuscations ordered from most to least effective.	85
5.5	Information sufficiency rating ( $M$ and $SE$ ). Obfuscations ordered from most to least effective. . . . .	86
5.6	Enjoyment rating ( $M$ and $SE$ ). Obfuscations ordered from most to least effective. .	87
5.7	Social presence rating ( $M$ and $SE$ ). Obfuscations ordered from most to least effective.	88
5.8	Obfuscation likability ( $M$ and $SE$ ) from most to least effective. . . . .	89
5.9	Scatterplot of Satisfaction (X axis) against Identification Rate (Y axis) which shows the general trade-off between effectiveness and satisfaction. <i>Morphing</i> , <i>inpainting</i> and <i>avatar</i> are below the regression line, which means they are relatively effective and satisfying. . . . .	90

5.10	Scatterplot of Likability (X axis) against Identification Rate (Y axis) which shows the trade-off between effectiveness and obfuscation likability. <i>Inpainting</i> and <i>avatar</i> are below the regression line, and <i>morphing</i> is effective but not likable. . . . .	91
5.11	Example of awkward <i>inpainting</i> . . . . .	92
6.1	The percentage of people who have censored photos and who have not. . . . .	100
6.2	The percentage of people who are willing to share photos which they previously had privacy concern using obfuscations and who are still withholding even with obfuscations. . . . .	101
7.1	Photo protection interface— content detection and obfuscation options. . . . .	109
7.2	Photo protection system interface—obfuscation options. . . . .	110
7.3	A screenshot of the control condition. The control condition mimics Facebook current photo sharing interface. . . . .	114
7.4	A screenshot of the control condition. The control condition mimics Facebook current photo sharing interface. . . . .	115
7.5	Privacy warning condition. . . . .	115
7.6	Results of the power analysis. . . . .	116
7.7	Marginal effects of between-subjects independent variable interface condition on perceived privacy risk, usage effort, and system satisfaction. All three metrics were measured in post-test. . . . .	121
7.8	Marginal effects of privacy enhancing conditions on willingness to share (measured in post-test) . . . . .	121
7.9	Interaction effects between experimental conditions and trust in SNS on ease of use (measured in post-test). . . . .	123
7.10	Interaction effects between experimental conditions and trust in SNS on system satisfaction (measured in post-test). . . . .	123
7.11	The path model shows that experimental conditions has effect on perceived privacy risks, and the increased privacy risks decreases willingness to share. . . . .	124
7.12	Pre and post tests results of perceived privacy risks. . . . .	125
7.13	Pre and post tests results of willingness to share. . . . .	126

# Chapter 1

## Introduction

### 1.1 Problem Motivation

Sharing photos on SNSs helps users manage the impression others have of them, maintain an off-line relationship with their family and friends, and gain attention from a wider audience than their existing friend circle [211]. While users enjoy these benefits, the rich visual information that photos contain may lead to privacy leakages, such as identification, location leakage, sensitive information leakage, and social activity leakage [6]. Privacy leakage could harm users' impression management and even influence their career [6].

A social recruiting survey of Jobvite shows that recruiters take candidates' social media profiles seriously when evaluating them [129]. Ashley Payne, a high school English teacher, was fired because of a parent's complaint on her Facebook photo of her holding wine and beer at a bar, though this photo was taken during her private vacation and she did not friend any of her students [68]. Similarly, an employee working at a nonprofit was nearly fired because she posted a photo of a donation card and revealed the donor's name [231].

Privacy leakage could also harm users and their family members' safety. Parents often innocently share their children's photos to record precious life moments. However, a Children's eSafety Commissioner warned that over half of the photos found on pedophilia websites were downloaded from parents' social media. Even photos depicting children's regular activities such as swimming or doing homework could make them vulnerable or become victims offline [236].

In another story, sensitive content leakage even led to a teenager's suicide. A topless photo

of a 15-year-old girl spread quickly among her classmates who bullied her on Facebook and in person, and she then hung herself after years of being bullied [116].

The above stories show us the possible serious consequences of photo privacy leakage. If failing to protect users' privacy, SNSs, with tens of millions of users over the world, might become an irresponsible and possibly criminal environment.

## 1.2 Research Motivation

A large number of research approaches have been explored to alleviate privacy leakage from different perspectives. These approaches can be categorized to controlling photo content disclosure and controlling photo recipients.

For example, in terms of the first approach, controlling photo content disclosure, Face/Off system applies facial recognition in a photo and obfuscates the detected faces [124]. However, obfuscation methods are very limited in these studies and blurring is the most commonly used method (e.g., [124, 251, 269, 297]) though it is ineffective. Moreover, these systems made incomplete assumptions about what types of content raise users' privacy concerns. For example, Google StreetView only identifies and blurs faces and license plates [103].

Most SNSs (e.g., Facebook) adopt the second approach—controlling photo recipients to protect photo privacy, such as selectively sharing with a group of recipients [86] or untagging themselves from others' photos [85]. Nonetheless, untagging is ineffective since unwanted viewers can still trace back to the identity of the person who reveals sensitive content. Besides SNSs, some researchers have explored this approach and developed privacy protection systems. For example, Cryptagram allows users to encrypt photos then upload them to SNSs [275]. Only the recipients that users specify can have the right credential to retrieve the original photo. Yet these systems all have drawbacks. First, it is time-consuming to choose the right recipients every time when sharing a photo. Next, due to a large number of SNS friends, users may accidentally select or exclude certain recipients [34]. Under such threats, users consider censoring/withholding photos to be safer option [191]. However, self-censorship hampers the communicative ability of SNSs [219]. Social media risks becoming an asocial environment if content sharing is limited [218].

More importantly, knowing that a number of privacy frameworks could inform privacy research, such as the behavioral privacy model, contextual integrity, and networked privacy, we find

that little work on online photo privacy protection leverages these privacy frameworks into their study design and mechanism development.

To address these research gaps, in my doctoral thesis, I aim to leverage the behavioral privacy model to inform my research, gain an understanding of privacy parameters, and examine privacy protection methods to inform the building for an effective and usable photo privacy protection system on SNSs.

### 1.3 Research Objectives

The behavioral privacy model identifies two core elements that could affect privacy – information content and information recipient [50]. Adjusting either one would affect privacy. Aiming at protecting online photo privacy, I decided to leverage this model with the focus on the two elements. My research aims to optimize the behavioral privacy model in the context of photo privacy; study the two parameters, content and recipient; offer an alternative photo privacy protection strategy other than self-censorship and SNS recipient control; and design an effective and usable photo privacy system. To achieve this goal, I plan to answer the following questions.

- Understand what to obscure (*content*) and prevent from whom (*recipient*):
  - **RQ1:** What is the sensitive content in photos to be obscured? (**Chapter 3**)
  - **RQ2:** What are users' preferences with different groups of recipients? (**Chapter 3**)
- Study and select promising obfuscation methods (can be viewed as a part of the *content* parameter):
  - **RQ3:** What are the effective and usable obfuscations? (**Chapter 4**)
  - **RQ4:** Are the obfuscations robust in terms of de-identifying both unfamiliar and familiar people? (**Chapter 5**)
- **RQ5:** As an extreme privacy protection scheme, is photo self-censorship prevalent? Can obfuscations combat it and encourage photo sharing? (**Chapter 6**)
- The outcome of my dissertation is a photo privacy protection system that enables sensitive content detection and offers obfuscation options to protect the identified sensitive content. We conducted an evaluation:

- **RQ6:** Can our system reduce users' privacy concerns when sharing photos? (**Chapter 7**)
- **RQ7:** Can our system encourage sharing? (**Chapter 7**)
- **RQ8:** Is our system usable? (**Chapter 7**)

## 1.4 Overview & Summary of Studies

To answer the research questions, we conducted five studies as Figure 1.1, Table 1.1, and Table 1.2 show.

Study	Research Questions	Method	Sample Size	Procedure	Analysis
1	<p>What content is sensitive in photos from a human-centered perspective?</p> <p>How sharing preferences for content differs across recipients?</p>	<p>Study 1: Online experiment</p> <p>Study 2: In-person card sorting</p>	<p>Study 1: 116 Recruited on MTurk</p> <p>Study 2: 14 Recruited on campus</p>	<p>Study 1: We asked participants to look at their photos on their phone and find one that they considered “private (means not share with anyone)” (photo 1). Once they found such a photo, we offered them three choices: 1) share the photo with us, 2) look for a photo online which has similar sensitive content and share it with us, or/and 3) describe the photo in detailed text. After the identified the photo and either uploaded it, a similar photo or described the photo they answered 20 questions which measured their likelihood to share the photo with the 20 recipient groups.</p> <p>After photo 1, participants then repeated this procedure four additional times with the following variations: we asked them to look for a photo they would NOT want to share with their family (photo 2), friends (photo 3), colleagues/classmates (photo 4), and acquaintances (photo 5).</p> <p>Study 2: Each participant first saw digital cards in XSort. All 181 cards of sensitive content elicited from Study 1 were placed randomly on the computer desktop. Next, we instructed participants to “place cards into groups in a way that makes the most sense to you, but please make sure the cards in the same group have a similar sensitivity level and content.” Once they were satisfied with a group, they labeled it with a name they generated. They could regroup and relabel until they were happy with the groups and names.</p> <p>In total, we had 14 obfuscation conditions. Participants completed 14 trials where they saw photos with semi-randomly assigned obfuscation conditions and target people, and identified the target person. Participants saw all 14 conditions and 14 target people. Afterwards, they rated their confidence, and rated the four statements about their feeling. After finishing all trials, participants were shown 14 conditions individually, and rated their preference towards each condition.</p>	<p>Study 1: Linear mixed effect model</p> <p>Study 2: XSort generated dendrogram</p>
2	<p>Which obfuscations are effective against human recognition and provide a good viewer experience?</p>	<p>Online experiment</p>	<p>271 Recruited on MTurk</p>	<p>In total, we had 14 obfuscation conditions. Participants completed 14 trials where they saw photos with semi-randomly assigned obfuscation conditions and target people, and identified the target person. Participants saw all 14 conditions and 14 target people. Afterwards, they rated their confidence, and rated the four statements about their feeling. After finishing all trials, participants were shown 14 conditions individually, and rated their preference towards each condition.</p>	<p>Generalized linear mixed effect model</p> <p>Linear mixed effect model</p>

Table 1.1: Blumenfeld chart summarizing the five studies in this dissertation



Study	Research Questions	Method	Sample Size	Procedure	Analysis
3	Which obfuscations are effective and provide a good viewer experience when applied on both familiar and unfamiliar people?	Online experiment	230 Recruited on MTurk	The procedure is the same as study 2. The only difference is that the 14 targets participants saw included 7 familiar people and 7 unfamiliar people.	Generalized linear mixed effect model
4	Is photo self-censorship prevalent? How is photo self-censorship related to gender, age, Internet and SNS usage frequency, and privacy consciousness? May obfuscation help combat photo self-censorship?	Online experiment	230 Recruited on MTurk	We introduced the study by showing participants six example photos reflecting each type of obfuscation presented in randomized order. We then asked participants "Have you ever declined to upload a photo to an online social network for privacy reasons?" There were three options: "Yes," "No," and "I don't know." we asked participants who answered "Yes" to the previous question: "In the last question, you said you had declined to upload a photo to an online social network for privacy reasons. If you had access to one of the privacy filters here, would you be willing to upload this photo using one of the filters?" Again, they responded "Yes," "No," and "I don't know," and then immediately provided qualitative feedback by answering the open-ended question "Please tell us the reason."	Logistic regression model Open coding
5	Can obfuscation reduce users' privacy concerns on their photos and increase their willingness to share while still maintaining a good usability?	Online experiment	159 Recruited via Qualtrics sourcing platform	This is a two-step study. In the first step, participants provided photos that they would like to share on Facebook but have not due to privacy concerns, and identified the sensitive content in their photos. In the second step, participants were split into three groups. Each group of participants first answered pre-test questionnaire in which we measured the perceived privacy risks and willingness to share their sensitive photos. They then evaluated a version of interfaces that assigned to them. After they performed the photo-sharing tasks, they answered post-test questionnaire in which we had the same two measurements as in the pre-test questionnaire and three usability measurements.	Path analysis

Table 1.2: Blumenfeld chart summarizing the five studies in this dissertation (continued).



Figure 1.1: Illustration of different phases of this dissertation.

**Study 1 Identifying sensitive content and users' sharing preference.** Focusing on the first element in the behavioral theory of privacy, controlling content disclosure, we must know what portions are considered sensitive and should be obscured. Although machine learning methods exist that can identify content in photos, we currently do not have a taxonomy that describes what content is considered sensitive, and how sharing preferences for content differs across potential photo recipients. To fill this gap, we collected photos that contain sensitive content from 116 participants and recorded their sharing preferences for these photos with 20 recipient groups. Next, we conducted a card sort study on the 181 unique pieces of sensitive content identified in this study to surface user-defined categories of sensitive content. Using data from these studies we generated a taxonomy that identifies 28 categories of sensitive content. We also establish how sharing preferences for content differs across groups of potential photo recipients. This taxonomy can serve as a framework for understanding photo privacy, which can in turn inform new photo privacy protection mechanisms.

**Study 2 Identifying effective and usable obfuscations.** With the focus on the content element, the second study introduces privacy-enhancing obfuscations for photos and conducts an online experiment with 271 participants to evaluate their effectiveness against human recognition and how they affect the viewing experience. Results indicate the two most common obfuscations, *blurring* and *pixelating*, are ineffective. On the other hand, *inpainting*, which removes an object or person entirely, and *avatar*, which replaces content with graphical representation are effective. From a viewer experience perspective, *blurring*, *pixelating*, *inpainting*, and *avatar* are preferable. Based

on these results, we suggest *inpainting* and *avatar* may be useful as privacy-enhancing technologies for photos because they are both effective at increasing privacy for elements of a photo and provide a good viewer experience.

**Study 3 Obfuscation effectiveness of de-identifying both unfamiliar and familiar people.** In Study 2, one limitation is that we only explored obfuscations’ effectiveness for de-identifying unfamiliar people (people in stimuli were unknown to the participants). Hence in Study 3, we conducted another online experiment with 230 participants where we investigated the effectiveness of enhancing photo privacy using obfuscations, which hide part of the photo content. Results indicate that obfuscations reduce privacy concerns associated with online photo sharing and encourage people to share more photos. Furthermore, we find that obfuscations’ effectiveness as a privacy-enhancement is differentially affected by familiarity, or whether viewers know people in a photo. We identify obfuscations that are robust to the increased likelihood of recognition associated with familiarity and provide a good viewer experience. We suggest these obfuscations would be useful tools for photo privacy enhancement, especially in cases where viewers are familiar with the people who are in the photos, such as SNSs.

**Study 4 Obfuscation may combat photo self-censorship.** SNS users self-censor or withhold content to achieve privacy. Due to the lack of useful photo privacy-protection tools, photos are a likely, but unexplored, target for self-censorship. We reported results from the survey we conducted with 230 participants in Study 3 which also elicited data about photo self-censorship on SNSs. We quantified the prevalence of self-reported photo self-censorship and associated this with gender, age, privacy preference, Internet and SNS usage, and interrogated whether privacy-preserving obfuscations, such as blurring, may be useful for combating photo self-censorship. Our results indicate that over half of the participants have self-censored photos on SNSs and privacy-conscious people are more inclined to censor photos. We also find that women are more likely to report they would share a photo they had previously self-censored if they were able to obfuscate portions of the photo to enhance privacy.

**Study 5 An experiment to determine whether obfuscation reduces privacy concerns and increases willingness to share.** Based on the formative research we have conducted, we developed a prototype of a photo privacy protection system. To evaluate this prototype, we conducted a three-group between-subject pretest-posttest experiment. We focused our assessment on the system’s ability to reduce privacy concerns, the ability to encourage photo sharing, and the

overall usability.

## Chapter 2

# Background and Related Work

To gain a better understanding of the current state of research about photo privacy protection on SNSs, I conducted a systematic literature review of research published in the ACM Digital Library and IEEE Xplore. I used multiple terms and combinations to search both titles and abstracts in each database and limited the search to manuscripts published between 2000 to 2020 (see Table 2.1). After the initial search, I manually excluded articles that are duplicated and/or not directly relevant to the scope of my dissertation. For example, I excluded papers about developing algorithms to recognize objects in photos.

The papers can be grouped into four categories: (1) investigating users' strategies or behaviors when faced with privacy risks (2.1.1), (2) developing online photo privacy protection systems (2.1.2), (3) identifying potentially sensitive content that leads to privacy concerns (2.3.1.1), and (4) implementing obfuscations to protect photo privacy (2.3.1.2). In the following sections, I first give an overview of online photo privacy and then discuss articles in each category.

### 2.1 Introduction to Online Photo Privacy

Hundreds of millions of Online Social Network (SNS) users present themselves, communicate, and share thoughts and pictures every day [82]. In 2018, 68% of U.S. adults used Facebook, and three-quarters of those users accessed Facebook on a daily basis. Additional social media platforms such as Snapchat and Instagram are popular among young adults (e.g., 78% of 18- to 24-year-olds are Snapchat users) [256]. On Facebook alone, 350 million photos are uploaded every day. In total,

Terms	ACM Digital Library		IEEE Xplore	
	Initial Search	After Screening	Initial Search	After Screening
photo privacy	182	41	153	16
social media privacy photo	17	2	18	0
image photo face de-identification	18	2	3	0
photo image obfuscation redaction privacy	27	0	29	4
privacy blur/pixelate/inpaint	51	5	112	5
self-censorship social network	64	3	26	0
Total	359	53	341	25

Table 2.1: Summary of systematic literature review

more than 250 billion photos have been uploaded by Facebook users [257], while on Snapchat about 20,000 photos are shared every second [14].

The massive amount of data shared on SNSs sometimes includes sensitive details, which generates a number of privacy issues. For example, people often reveal information such as date of birth, gender, location which can lead to physical privacy risks [126]. Other privacy issues include unintentional facial recognition, inference attacks, location leakage, identity theft, relation privacy leakage, phishing, and profiling risk [92, 153, 170, 283].

Unlike textual information shared on SNSs, the rich visual information that photos shared on SNSs may pose an even higher level of privacy risk. One primary privacy concern for SNS users is impression management within their social circles [25]. For example, users expressed concerns about unflattering photos and incriminating evidence. However, despite these concerns, people fail to predict the short-term or long-term consequences of the information flow, at least in part due to the limitations of human working memory [187].

Other serious privacy invasions may occur without users’ awareness. For example, identification, location leakage, sensitive information leakage, and social activity leakage are all possible consequences of photo sharing on SNSs [6]. Photo privacy leakage not only harms users’ impression management, but may also lead to financial and social embarrassment [8] and even threatens users physical and property security (e.g., stalking) [6, 8].

### 2.1.1 Users’ Privacy-Protective Practices

With increasing concerns, users actively take steps to alleviate privacy issues. A behavioral theory of privacy examines Privacy-Enhancing Behaviors (PEBs) and clusters them into three cate-

gories: avoidance, modification, and alleviatory [50]. Though this model is not limited to the online environment, it does capture, predict, and help organize users’ photo privacy behaviors. Many online photo PEBs fall within the avoidance and alleviation categories. Modification refers to users modifying their behaviors during the act of photo capture which often occurs in the physical environment (e.g., being careful, not in front of others, quietly). While we see the potential for work on this category in the future, for the work proposed here, we focus on avoidance and alleviation.

#### **2.1.1.1 Avoidance**

Most photo privacy protection behaviors fall within the avoidance category. Avoidance behaviors refer to the actions that people take to avoid privacy leakage before it occurs [50]. A very common behavior is photo self-censorship—self-censoring potentially problematic content before sharing. Self-censorship, described as “the act of intentionally and voluntarily withholding information from others in the absence of formal obstacles [18]”, is adopted by SNS users to maintain a consistent self-presentation among different audience groups [191]. Since SNSs contain few visual cues about the audience [48], users are unlikely to have an accurate understanding of their social graph considering the large number of SNS “friends,” or are underestimating their photo recipients [23]. In such cases, they consider censoring/withholding photos to be a safer option [191]. For instance, a prior study alludes to photo self-censorship by suggesting teens refuse to share sensitive photos online [61]. Yet photo self-censorship hampers the communicative ability of SNSs, which is an important feature of social networks [219]. Social media risks becoming an asocial environment if content sharing is limited [218].

Another avoidance solution, which is often adopted by SNS users, is selective sharing, defined as “refrain from sharing content or by selecting recipients.” Regarding recipient selection, users utilize privacy settings or recipient controlling tools offered by SNSs to disclose their personal information selectively, hence this approach is also named “technical strategy” by Stutzman et al. [268]. Based on a survey of undergraduate Facebook users’ privacy-enhancing practices, 83% of participants stated using Facebook privacy settings and 58% indicated they had made their profile friends-only [268]. In terms of photo sharing, the mainstream SNSs such as Facebook and Snapchat allow users to choose public, friends only, or a specific group of friends to share a photo. However, “friends” on SNSs is very ambiguous which may include anyone from a significant other to a complete stranger [63] since a small portion of users is likely to accept friend requests from unknown

people [132]. As users might believe that they have done an adequate job in protecting privacy by only allowing “friends” to access their uploaded content, they are very likely to share their personal information which brings them risks [63]. In terms of the other aspect of selective sharing—content restriction, users would share a photo with potential sensitive content but hide the most sensitive part by, for example, cropping it or covering it with a sticker.

#### **2.1.1.2 Alleviation**

Alleviation are the actions taken to reduce the consequences of the spread of information after a photo has been captured [50]. For example, users might delete photos when they notice potential privacy risks after posting photos. Though users believe their photos are deleted immediately, most existing SNSs have the deletion delayed up to 30 days which lets others re-access those deleted photos via photo links [181].

Privacy on SNSs is not only about one’s disclosures, but also the disclosures about one’s self by other users [268]. For example, a user may be tagged and auto-identified in another user’s photo without knowledge and consent. Thus, users often leverage alleviatory strategies to address conflicting sharing decisions by others. Specifically, these strategies involve asking others to make a photo containing him/herself private, asking others to completely remove a photo [107], and untagging him/herself from a photo [24, 268]. However, this approach may cause social tension. Furthermore, even if a user untags him/herself, viewers are still able to access that photo in the photo uploader’s profile.

### **2.1.2 System Solutions for Photo Privacy Protection**

Many studies have been conducted from different perspectives to protect online photo privacy. Intervention timing is an important consideration when developing these photo privacy control mechanisms which align with the behavioral privacy model. The model elucidates temporal aspects of privacy: avoidance, which occurs prior to an act; and alleviation, which occurs after an act [50]. The review of research approaches is summarized based on their intervention timing.

#### **2.1.2.1 Avoidance**

Most of the research fits within the avoidance temporal category.



**At the time of capture.** A photo could be protected before upload, and an SNS would never gain access to the raw photo, which addresses organizational threats related to SNS providers and various third parties [151]. The novel approach “Offlinetags” can be applied at the time of capture to avoid privacy leakage. Offlinetags consist of four symbol stickers which represent people’s privacy preferences—no photos, blur me, upload me, tag me. They are designed to be easily recognizable by algorithms, so that cameras either do not take photos or automatically blur the people who wear “no photo” or “blur me” stickers during capture. Unlike cameras or phones that people can effortlessly manage the stream of data, wearable lifelogging camera devices such as Google Glass continuously capture large numbers of photos without users’ actions. The results of in situ user studies and interviews suggest that instead of reviewing and deleting problematic photos later, the wearable cameras should enable obscuring certain content or pausing instantaneously once detected [66, 122]. In general, this approach controls a portion of the photo content (e.g., people’s faces) or the entire photo.

**Before sharing.** Despite the possible computational resource-intensity, photo privacy can be protected before photos are uploaded online. For example, a photo privacy-preserving tool consists of two parts – a client-side application for applying scrambling obfuscation on people’s faces and a server for hosting photos [306]. The client-side application obfuscates a photo then uploads it to the private server. Viewers who have the key are able to view the obfuscated-free version. Researchers also developed a system that encrypts photos prior to sharing on Facebook [270]. Similarly, PrivacyJPEG encrypts several sensitive areas of photos and binds the access control policies before uploading photos [169]. Another tool is integrated in camera application [248]. After a photo is taken, the tool automatically detects all the faces in this photo and notifies the user, therefore he/she will beware of other people’s privacy when making the sharing decision.

**At the time of sharing.** Most techniques in previous research are designed to protect photo privacy at the time of sharing (e.g., after uploading). These techniques can be categorized into three approaches – controlling photo content, controlling photo viewers, and controlling photo metadata.

In terms of controlling photo content, face blurring is the most widely adopted method. For example, Face/Off determines which viewer is not permitted to view which face in a photo and consequently blurs that face [124]. A recent study on multi-party photo privacy proposes a built-in option in SNSs which enables photo content modification such as blurring faces and cropping people

out when a sharing conflict is detected [269]. Similarly, another system utilizes facial recognition, notifies other stakeholders in a group photo, and provides the face blurring option [297]. A tool named “Cardea” predicts users’ privacy preferences based on four elements in photos – location, scene, other people’s presences, and hand gestures, then applies blurring on certain people’s faces [251]. Apart from face blurring, researchers propose an automated method to replace people unintentionally captured in photos with synthesized people from a dataset. Though the application context in this study is Google Street View, this method can also be applied to any online photos [208]. Additionally, cartoon stickers have been employed to hide people’s faces [173].

Controlling photo viewers is another common approach. “Cryptagram” enables converting photos to encrypted images [275]. Users upload the encrypted images to SNSs and only the viewers with the right credentials can retrieve the original photo. It guarantees that both unwanted viewers and SNSs cannot infer the photo. Several other systems protect photo privacy by predicting and optimizing photo access control settings using a set of elements of photos. For example, a study identifies photo content and aesthetics as elements that influence users’ sharing preferences [136], while two studies find photo content inferred by tags and the tie strength of relationship may decide the preferred photo recipients [95, 140]. In addition, a context-based personalized privacy settings recommender system provides users recommendations on photo access settings [264]. With the focus on multi-party privacy conflicts, a system analyzes the social intimacies between the visitor and co-owners and restricts certain viewers’ access to an entire photo [87].

Metadata associated with photos include cameras’ identifier numbers and the GPS coordinates of the location where a photo was taken. Besides the photo content, the metadata can cause privacy leakage as well, especially since people may not be aware of it when sharing photos [101]. To address this concern, researchers developed a metadata protection application to enable users to change metadata and protecting posting location [101]. Metadata leakage is often caused by users’ unawareness of the existence of metadata. Hence, in another study, an SNS extension first shows users the photo metadata and provides options to modify or remove it [113]. Additionally, researchers propose the use of external metadata storage services which allow users to easily manage metadata for photos on SNSs [113]. On the other hand, metadata can be used to protect privacy. For example, SnapMe watchdog utilizes metadata and keeps track of photos taken in the users environment [114].

### 2.1.2.2 Alleviation

**After sharing.** “Restrict Others” tool was developed to protect personal information in others’ photos [25]. Besides untagging, it allows each stakeholder depicted in a photo to negotiate with the owner to hide the photo from certain viewers. PRIMO notifies users privacy violations in photos that other people upload, for example, photos showing a user with other people not having his/her gender [42]. However, since these tools intervene after the act, it is likely that the unwanted recipients have viewed the photo.

Though the previous work has sought to protect online photo privacy using various approaches, little of the work leverages privacy theories. Even if a little work invokes privacy theories, most of them only mention theories in the background section without deep engagement [16]. In the section below, I introduce common theories of privacy and a behavioral privacy model that informs my research.

## 2.2 Theories of Privacy in HCI

Privacy, as a multifaceted concept, has been studied from the perspective of many different disciplines including philosophy, law, communication, social psychology, and Human-Computer Interaction (HCI). The concept of privacy was first brought up in philosophical discussions. Aristotle identified two distinct spheres of life—the public sphere of political activity and the private and domestic sphere of the family [64]. In the discipline of law, Warren and Brandeis defined privacy as the legal right to be let alone [292]. In the physical world, letting alone is not a tough task since the boundary between public and private can simply be the walls of one’s house. However, besides physical privacy, researchers started to concern themselves with the informational privacy invasions, such as the unauthorized dissemination of portrait photos and the news that contain personal information [292]. In the discipline of communication, Burgoon et al. characterized information privacy as the “ability to control who gathers and disseminates information about one’s self or group and under what circumstances [43].”

Privacy has been investigated extensively in social psychology as well. The widely held definition of privacy in social psychology is the right to control information about oneself. For example, Westin defined privacy as the right to “control, edit, manage and delete information about themselves and decide when, how, and to what extent information is communicated to others [293].”

Stone et al. described privacy as “the ability of the individual to control personal information about one’s self [265].” Beyond the individual privacy, Altman framed privacy as “an interpersonal boundary process by which a person or a group regulates interaction with others [11].” He argued that the boundary regulation process is dynamic. An individual alters the degree of openness to others to achieve his/her desired level of privacy over time. One’s disclosing and stimuli from others are both involved in boundary regulation.

Communication Privacy Management Theory (CPM), a framework in the discipline of communication, further expands Altman’s theory by defining privacy as “the feeling that one has the right to own private information, either personally or collectively [226].” CPM introduces two types of boundaries: personal boundaries that manage private information about oneself and collective boundaries which involve the private information shared with others. Collective boundaries are the overlap of two circles which represent two persons’ private information, hence the lines of ownership is often ambiguous.

In the field of HCI, an important privacy framework is contextual integrity (CI) developed by Nissenbaum [207]. It describes privacy as an appropriate flow and emphasizes that privacy is not about not collecting data nor data minimization. In contrast, individuals have the need to share information, but the information should be shared appropriately. Individuals interact within a context and the flow should conform with the legitimate contextual informational privacy norms in the context. CI introduces five independent parameters in privacy – actors which include subject, sender, and recipient (e.g., physician, teacher, friend), information type (e.g., demographics, transaction history, photo), and transmission principle (e.g., consent, compel, buy). For example, a physician (sender) can legally release a patient’s medical information (attribute) to designated people (recipient) with his/her consent (principle), while releasing information without consent is considered an invasion of the patient’s privacy. An appropriate flow is hard to regulate in a networked online environment [31, 192]. For example, the information we share about ourselves may unintendedly disclose other people’s personal information.

## **2.3 In The Context of Online Photo Privacy**

Unlike other types of data such as text, photos that contain rich visual information are more likely to cause privacy issues when shared on SNSs. Much photo privacy leakage is due to

the information disclosure which occurs when “privacy is violated in that information not intended for a specific person(s) is nevertheless revealed to that person [50].” For example, the intended content “drinking in a bar” can be intended to be disclosed to a close friend recipient; however, accidentally sharing this content with unintended recipients such as colleagues and supervisors may cause trouble (recipient disclosure). On the other hand, people may be willing to share photos that depict themselves with the appropriate clothes; however, accidentally sharing partial nude photos are considered content disclosure.

To my knowledge, there is no framework that is specifically developed for the context of online photo sharing. To ground my research within a privacy theory, I first delved into CI. CI has been useful in privacy studies such as investigating the threat raised by vehicle safety communication technology [308] and identifying privacy issues on SNSs [183]. However, this model may not be very suitable in a photo privacy setting. For example, the attributes can only be the visual content in photos and in most cases, the subject is the user him/herself. Moreover, CI lacks another important element in photo privacy – photo content (e.g., identity, object).

Hence, I looked into the behavioral theory of privacy [50]. Firstly, this theory elucidates temporal aspects of privacy: avoidance, which occurs prior to an act; modification, which occurs during an act; and alleviation, which occurs after an act. In the photo setting, users’ privacy behaviors and privacy protection techniques I reviewed above all fit within the avoidance and alleviation categories. Recent work supports the usefulness of a temporal perspective. For example, Rashidi and colleagues [232] describe how users manage privacy in a collaborative environment via the life span of photos which includes four temporal stages: whether the user is ready to be captured in a photo, whether the photo should be taken, whether a photo can be shared, and whether the sharing should be mitigated. Second, this behavioral privacy model identifies two core elements that could affect privacy – information content and information recipient [50]. Adjusting either recipient or content would affect privacy (Fig. 2.1). Aiming at protecting online photo privacy, I decided to leverage this behavioral privacy model with the focus on the two elements – content and recipient.

### **2.3.1 Controlling Content**

Controlling information content has been implemented in visual content protection extensively. For example, since most YouTube videos are public by default, it is hard to regulate the information recipients. Instead, YouTube uses a blurring obfuscation to hide faces and objects in

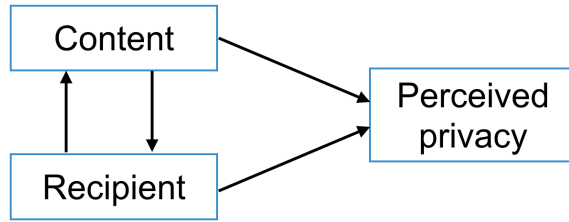


Figure 2.1: Content by recipient interaction described in the behavioral privacy model [50].

certain videos [301]. In the home environment, two privacy-preserving mechanisms (blob tracker and point-light) have been used to protect elements of users’ identities [49]. Specifically focusing on photo privacy protection, I have summarized the research that aligns with the approach of controlling photo content in the paragraph of “At the time of sharing” in “System Solutions for Photo Privacy Protection.” However, most of the previous work made incomplete assumptions that people’s faces were the only sensitive content to be protected and applied blurring obfuscation on faces.

### 2.3.1.1 Sensitive Content

We have some hints from prior work about the kinds of content in a photo that people consider sensitive (summarized in Table 2.2). For example, interview studies reveal that people are very cautious when sharing photos which illustrate their own faces or family members’ faces on SNSs either because they want to project a perfect image to manage others’ impression of them or they want to avoid others misusing these photos [3, 25, 155]. Sensitive features extracted from participants’ photos and users’ comments via machine learning indicate that people, landscape, and certain places and events are sensitive [45]. Certain objects, backgrounds [6], and phone screens [121] are also common concerns. When cameras are ubiquitous, such as in life-logging, monitor screens, and irrelevant persons in photos lead to privacy concerns [121]. People are also concerned about revealing photos that contain text such as their address, organizational affiliation, and email address [15, 100].

The most comprehensive study to date using an ML approach examining content sensitivity is work that claims to identify 268 privacy-sensitive object classes [303, 304]. The privacy-sensitive object classes include sensitive people, sensitive locations, toilet, discrimination texts, home shrines, and visual attributes for personal hobbies. However, there are a number of limitations to this work

that makes it difficult to apply towards the goal of understanding sensitive content.

First, researchers identified the privacy-sensitive object classes using a set of photos that people had uploaded to an SNS. While the sets were labeled as “private,” it is not clear what “private” meant to the people who uploaded the photos. Obviously, the photos they uploaded were shared with the organization hosting the photos (in this case, probably Flickr, but see further limitations below), so not “private” if we mean that the photos were not shared with anyone. It is unlikely that people would have shared to Flickr their most sensitive photos. Instead, they may have chosen not to upload the most sensitive photos at all [255]. Hence, the photos in this dataset may not represent the most sensitive photos. For example, they would not contain any photos that participants chose not to upload to Flickr.

Additionally, current machine vision approaches are only able to detect object classes present in existing photo datasets, such as ImageNet [241] and MS COCO [182], which are not privacy-specific. MS COCO, for example, focuses on objects that “would be easily recognizable by a 4-year-old.” Given the general-purpose goal, those datasets do not contain private images with sensitive objects. As a result, the machine learning approaches based on those datasets are limited to detecting general-purpose objects, rather than sensitive content. Because sensitive objects are not a part of these object sets they, therefore, cannot currently be detected reliably.

Finally, the paper fails to provide critical methodological details and detailed results which makes judging the rigor and implications of the work impossible. For example, while we think the SNS the researchers drew from was probably Flickr, this information is not presented in the paper, and requests for this information to the authors were not answered. Furthermore, there is no information about whether the privacy setting was fine-grained, whether the sample size is sufficient, and whether the sample of participants was representative. All of these factors taken together make it impossible to use this prior work to understand sensitive content in photos.

To our knowledge, no work systemically identifies and summarizes sensitive content in photos. Without an instructive framework, many photo obfuscation systems do not refer to any studies that examine sensitive content, but rather make untested or incomplete assumptions about what types of content raise users’ privacy concerns. For example, Google Street View considers people’s faces and vehicle license plates to be the highest priority sensitive content but neglect other content, such as private houses or objects in yards [103].

Category	Sensitive content	Research method
Identity	Photo owner [25, 124]	Focus group; N/A
	Family members [3]	Interview
Nudity	Children [3, 136]	Interview
	[215]	EU Data Protection Directive 95/46/EC, EUS Privacy Act of 1974, SNSs rules
Factors that harm impression management	Unflattering/embarrassing shots [3]	Interview
	Activity that may be misinterpreted [3]	Interview
	Presentation management [121]	In situ study
	Environment [6]	Interview
	Event [6, 25]	Interview; focus group
Factors that reveal personal information	Monitor screen [121, 238]	In situ study
	Location [121]	In situ study
	Written information [121]	In situ study
	Bedroom [238]	Online experiment
Illegal	Illegal activity [25]	Focus group
	Copyright [3]	Interview
Photo quality	Technically flawed photo [3, 121, 136]	Interview; in situ study
No need to share	Irrelevant to viewers [3]	Interview

Table 2.2: Summary of previously identified sensitive categories in photos

### 2.3.1.2 Obfuscation methods

Blurring is the most commonly used obfuscation to control information content disclosure both in research and in practice. However, blurring may not provide sufficient privacy protection [180]. Both in the photo and video surveillance, blurring is less effective than solid masking in terms of preventing people from recognizing the obscured individual [180, 148, 179]. In addition to human recognition, blurring is also susceptible to reversal by machine identification; researchers have used generative adversarial networks to refine image details [162], trained artificial neural networks to perform re-identification [195], and automated “faceless recognition” using clothes and/or pose [212]. Redaction tools from video surveillance can be applied to privacy-preserving online photo sharing, since both of these applications attempt to protect a subject’s identity by hiding the subject’s visual information (see Table 2.3 for an overview of the obfuscation methods). However, most of these focus on the effectiveness of these redaction tools against automatic recognition software. To the best of our knowledge, none of these tools have been used to enhance privacy for photos shared via SNSs. In particular, I have seen no work that investigates their effectiveness against human re-identification, especially in relation to users’ attitudes towards these tools.











Example	Name & Definition	Related Work	Example	Name & Definition	Related Work
	<b>Blurring.</b> Reduces image detail by generating a weighted average of each pixel and its surrounding pixels.	[24, 69, 80, 98, 138, 149, 158, 301]		<b>Pixelating.</b> Replaces original small pixels, which are single-colored square display elements that compose the bitmap, with larger pixels.	[65, 146, 148, 158, 286, 302]
	<b>Silhouette.</b> Replaces content with a monochrome visual object that mirrors the extracted shape of the original content.	[49, 149, 214, 216, 305]		<b>Avatar.</b> Replaces content with a graphical representation that preserves some elements of the underlying content. For example, a human avatar can preserve facial expression and gesture, but hide biometrically unique elements (e.g., face) of identity.	[216, 235, 262]
	<b>Point-light.</b> Replaces content with several moving dots that preserves some elements of the underlying content. For example, a point-light image of a human can preserve a person's activity, but hide many biometrically unique elements.	[49]		<b>Bar.</b> Replaces content with a monochrome visual object that is the shape of a small, thin rectangle.	[305]
	<b>Masking.</b> Replaces content with a monochrome solid box that covers the content to be protected and surrounding image content.	[148, 149, 305]		<b>Inpainting.</b> Completely removes content, fills in the missing part of the image in a visually consistent manner.	[149, 216, 272, 282, 305]

Table 2.3: Eight obfuscation methods.

Category	Recipients
Private	Only me
Family	Spouse/significant others [47, 55]
	Household members [47, 56, 55, 102, 224]
	Relatives [47, 224]
Friends	Close friends [47, 102, 224, 285]
	Normal friends [47, 56, 55, 224, 300]
Colleagues & Classmates	Colleagues, co-workers [47, 55]
	Classmates [285]
Acquaintances	Supervisors [47, 55]
	SNS friends that haven't met offline [300]
	Acquaintances [224]
	Loose acquaintances [102]

Table 2.4: Summary of recipients from prior literature

### 2.3.2 Controlling Recipients

People have different levels of privacy preference for various groups of photo recipients [56, 213]. We summarize different recipient groups from prior literature in Table 2.4, including private, family, friends, colleagues or classmates, and acquaintances. Most SNSs (e.g., Facebook) leverage the recipient control approach (access control list model) [192] which enables users to select a subset of friends to share their photos or posts with; or if they are unsatisfied with a photo, they may self-censor the photo. This approach addresses privacy concerns by preventing unwanted others from viewing their photos [25, 266]. However, in practice, SNS users may not fully understand their social graph due to a large number of SNS friends [23], thus may include inappropriate recipients (recipient disclosure). Choosing recipients from SNS friends is also a cumbersome task. Furthermore, their intended recipients may change considering different types of sensitive content. Hence, there lacks a system that provides recommendations to users on which recipients they could safely share a photo with.

## 2.4 Summary of the Literature Review

According to the behavioral privacy model [50], we know that both of the photo content and recipients in combination influence privacy (Figure 1.1). Hence, in my research, I combined controlling content and controlling recipients to provide better photo privacy management. From the above literature review, we understand the limitations of photo privacy protection tools that

implemented in practice and developed in research:

- Insufficient knowledge of user-defined sensitive content in online photos
- Insufficient knowledge on appropriate photo recipients considering different sensitive content depicted in photos
- Lack of effective and satisfying obfuscation methods that can be applied on SNSs
- Lack of knowledge on obfuscation adoption willingness and its potential to combat self-censorship

My research aims to address these limitations and propose an effective and usable photo privacy protection on SNSs which considers both content control and recipient control.

## Chapter 3

# Study 1: Identifying Sensitive Content and Users' Sharing Preference

Note: This work was published at CHI 2020 [176].

### 3.1 Introduction

To protect online photo privacy, researchers have developed photo obfuscation systems which make part of the photo content invisible to viewers, such as masking a person's face [250]. However, these systems make incomplete assumptions about what types of content raise privacy concerns. For example, the Face/Off system assumes that faces are the only sensitive content that needs to be protected [124]. Researchers have tried to use machine learning to understand what content is sensitive, but this work has severe methodological limitations limiting its usefulness. Therefore, there is a need for a user-defined taxonomy of sensitive content in photos. This taxonomy should be based on content users identify as sensitive. Moreover, because people have different levels of privacy preference for various groups of photo recipients [56, 213], we do not yet understand the variations in sharing preferences by recipient group. To bridge the gap, we propose a taxonomy that systemically identifies and summarizes sensitive content in photos and facilitates an understanding

of people’s sharing preferences for sensitive content categories with different recipients.

We also introduce a new method for sensitive content elicitation which overcomes the limitations of prior machine learning approaches. Using this approach, we collected 181 unique pieces of sensitive content from 116 participants. We then further grouped the content into 28 categories via a card sort with a different set of 14 participants. We not only report what content is considered sensitive but also summarize why participants are unwilling to share various types of sensitive content, for example, to avoid getting into trouble or harming impression management. In terms of recipients, we observed a four-level sharing preference pattern (i.e., private, significant others, close relatives and friends, colleagues). We also found several cases that did not align with this pattern when we compared recipient groups in the subset of each sensitive category. Finally, we describe how our work might be applied to Social Network Sites (SNSs) and how it might benefit relevant machine learning studies.

The contributions of this paper are sixfold. We:

- Introduce a novel method to elicit sensitive content from participants. It removes many of the barriers in collecting private content by providing participants with alternative ways to identify sensitive data that preserve their privacy.
- Integrate prior work from across disciplines, test it, and extend it. We collected a much larger data set (563 total items including 181 unique pieces of sensitive content) from a larger sample size compared to prior work (see Table 2.2).
- Provide a more granular level of detail about sensitive content categories which may be more practical for privacy researchers, computer vision researchers and practitioners.
- Connect granular sensitive content categories to potential recipient categories, surfacing both consistencies in terms of sharing preferences and exceptions to these consistencies.
- Describe, based on qualitative data, reasons people might not want to share sensitive content in photos.
- Provide design implications for building new photo privacy protection systems.

Category	Recipient groups
Private	Private, not share with anyone
Family	Significant others Household members Close relatives Distant relatives
Friends	Close friends Distant friends Ex-girl/boyfriends
Colleagues & classmates	Close colleagues/classmates Distant colleagues/classmates Close supervisor Distant supervisor
Acquaintances	Friends of friends People you've only met online People you've only met once or twice
Age	People of your age People younger than you People older than you
Gender	People of the same gender as you People of different gender

Table 3.1: Recipient groups used in our study

## 3.2 Method

### 3.2.1 Study One: Photo Elicitation

We collected two types of data via the photo elicitation: first, we gathered photos and/or descriptions of photos with sensitive content to understand what content is sensitive. To collect a purposefully diverse set of sensitive content, we defined private as photos that participants keep 1) private, and are unwilling to share with 2) family, 3) friends, 4) colleagues/classmates, and 5) acquaintances, asked them to upload corresponding photos for each category and then to identify sensitive content. Second, for each photo, they answered a question about their likelihood to share that photo with the 20 different recipient groups shown in Table 3.1.

#### 3.2.1.1 Participants

Our goal was to obtain a sample whose demographic and technology experience characteristics mirrored and reflected the variations among U.S. Internet users. In particular, our goal was to recruit a sample that was reflective of the target population in terms of age, gender, race, Internet

usage, and SNS usage. We use the Pew Research Center’s [228, 225] data on Internet usage and demographics for comparison.

To determine the necessary sample size for our study, first, we ran a pilot study to understand how the data points (photos and text descriptions) were distributed in each sensitivity category. We recruited 20 participants via MTurk and asked them to complete the procedure in the ‘Procedure’ subsection. Next, we conducted a power analysis based on the pilot study to calculate the necessary sample size. Specifically, if we want to find an effect at 0.85 power level between different recipient groups within the smallest sensitive content category which has only five data points in our pilot study, the power analysis revealed we would need 84 participants. To allow for a larger margin of error, we decided to increase the number of participants to 120 for the full-scale study. We recruited 120 participants via MTurk. MTurk meets one of our criteria for our target sample in that MTurkers are Internet users [240]. Additionally, MTurk recruitment results in a more diverse sample compared to standard Internet sampling and college sampling [39]. The data in studies using MTurk are as reliable as those obtained via other recruitment methods [51]. Moreover, MTurk is commonly used successfully for conducting privacy research [53, 234, 296]. We paid participants \$4.00 to complete the 30-minute session which is in line with the recommendation in [253] to pay workers at least minimum wage in the study’s location. To ensure high data quality, we set restrictions to only include US-based MTurk workers with a high reputation (above 97% approval ratings), and with the number of HIT approved being greater than 500 [223]. Additionally, we included three attention check questions throughout the survey to detect inattentive respondents [1] (e.g., “How likely is that you are paying attention, please do not select anything”).

Excluding the data of participants who failed two or more attention check questions, the final sample size is 116 (56 men, 59 women, and one participant preferring not to disclose gender). Fifteen percent ranged in age from 18 to 24; forty-eight percent ranged from 25 to 34; twenty-three percent ranged from 35 to 44; fourteen percent were 45+. Seventy-eight percent were White. Seventy-two percent visited SNSs most of the day or several times a day and 48% uploaded photos at least a few times a week. This sample mirrors and reflects the variations [152] among the demographic characteristics of the population of U.S. adults who use the Internet in terms of age, gender, race, Internet usage, and SNS usage as compared to samples obtained by Pew. The Pew samples, in turn, are representative of the population of U.S. Internet users as a whole [228, 225]. In other words, our sample has similar demographic characteristics in terms of age, gender, race, Internet usage, and

SNS usage to the population of U.S. Internet users.

### 3.2.1.2 Measurements

**Sensitive photo.** First, participants identified one personal photo that they considered sensitive. Next, they had one of three options: 1) upload the photo (we reminded them that only researchers would have access to this photo and would not share it), 2) find a photo online that contained similar sensitive content and upload that photo, or 3) or describe the photo in words.

**Identify sensitive content.** After providing a photo or description, we asked participants to answer an open-ended question “What content in this photo do you consider sensitive?”

**Sharing Likelihood.** After identifying the sensitive content in a photo, participants rated the sharing likelihood with each of the 20 recipient groups (Table 3.1). These recipients were developed based on prior work (Table 2.3) with additional granularity in the form of close and not close as suggested by [144]. Additionally, we included two more dimensions: age and gender. Participants answered “How likely are you to share this photo with \_\_\_?” on a Likert-type scale from 1-very unlikely to 7-very likely. This likelihood scale is adapted from [289].

### 3.2.1.3 Procedure

The entire study was IRB approved. Before the actual test, we conducted a pilot study to check for bugs and to assure that the data collection worked well. During the actual test, participants accessed our experiment website, hosted by Qualtrics, via the link posted on MTurk. After they consented, they answered six demographic questions, two social network familiarity questions, and a social network photo uploading frequency question. Next, we asked participants to look at their photos on their phone and find one that they considered “private (means not share with anyone)” (photo 1). Once they found such a photo, we offered them three choices: 1) share the photo with us, 2) look for a photo online which has similar sensitive content and share it with us, or/and 3) describe the photo in detailed text. After the identified the photo and either uploaded it, a similar photo or described the photo they answered 20 questions which measured their likelihood to share the photo with the 20 recipient groups listed in Table 2.4.

After they completed all 20 questions for the first photo they identified, participants then repeated this procedure four additional times with the following variations: we asked them to look for a photo they would NOT want to share with their family (photo 2), friends (photo 3), col-



leagues/classmates (photo 4), and acquaintances (photo 5). For each variation, we gave them examples of each recipient group. For example, the examples for family are significant others, household member, close relatives, distant relatives. After finishing the five photo collection tasks, participants received a code and pasted it to MTurk to receive remuneration.

## **3.2.2 Study Two: Open Card Sort**

The photo elicitation study resulted in 181 unique pieces of raw sensitive content (see further details in the results section). To group the sensitive content items into categories, we conducted an open card sort. A card sort is a method to discover how people think content should be organized and named [260, 21]. Because there was not a predetermined number of categories required, and because we were interested in having participants generate names for categories, we conducted an “open” (vs. closed) card sort. In an open card sort, participants can create as many categories as they want and generate a name for each category they create [21].

### **3.2.2.1 Participants**

We recruited 14 participants (in line with the sample size recommended in [279]) to take part in the in-person study via posting flyers on campus. Five participants were male, and nine were female. They ranged in age from 18 to 38. We offered them \$10 Amazon gift cards for their participation in the 40-minute session. As is standard for card sort studies (e.g., [259]) there was no overlap in participants between study one, where participants provided content and study two, where participants sorted content.

### **3.2.2.2 Procedure**

Each participant first saw digital cards in XSort, a computer program designed to collect card sort data. All 181 cards were placed randomly on the computer desktop. Next, we instructed participants to “place cards into groups in a way that makes the most sense to you, but please make sure the cards in the same group have a similar sensitivity level and content.” Once they were satisfied with a group, they labeled it with a name they generated. They could regroup and relabel until they were happy with the groups and names.

### 3.3 Results

From the photo elicitation we collected 563 data points, of which 545 were photos uploaded by participants. Of these, 329 were personal photos and the remaining 216 were photos that participants found online which had similar sensitive content to the personal photos they identified on their phones. For each photo or text description that they provided, we used an open-ended question to ask them to identify and describe the sensitive content. Across the 563 data points, we identified 181 unique pieces of sensitive content (see the Example column in Table 3.2). The answers to this question also revealed some reasons that people don't want to share certain sensitive photo content, which we discuss in the "Why Don't People Share?" section of the Discussion.

#### 3.3.1 Sensitive content categories

The primary purpose of the card sort study was to group the 181 pieces of content into categories. To generate categories based on the card sort data we performed a hierarchical cluster analysis [21]. Hierarchical cluster analysis progressively groups items based on their tendency to co-occur in participants' card sorting groups. This analysis allows us to answer the question "which items are often grouped together and therefore perceived to be similar, and which items are rarely grouped together and therefore perceived to be dissimilar [21]?" The results are visualized in a dendrogram. Due to space limitations, Figure 3.1 only shows a portion of the complete dendrogram, but see the supplemental document titled "dendrogram" for the version containing the entire dendrogram. Upon deliberation, we selected the 0.8 breakpoint. Selecting a breakpoint (or level in the hierarchy) impacts the number of clusters. Choosing a smaller breakpoint would result in more categories, whereas a larger breakpoint would result in fewer categories (with lower granularity). The 0.8 breakpoint resulted in 28 categories of sensitive content (Table 3.2). These categories roughly align with the sensitive content categories which were derived from previous literature and therefore provide support for these prior findings. However, due to our much larger data set compared to any of the prior work in Table 2.2, our results are much more granular in detail, and therefore simultaneously expand and refine those categories. For example, whereas prior work [136, 215] found that nudity was a category of sensitive content, our work revealed nuances such as that *breastfeeding* is not in the same category as other types of *nudity*.

Our results regarding photos of children are similarly notable as compared to prior work:

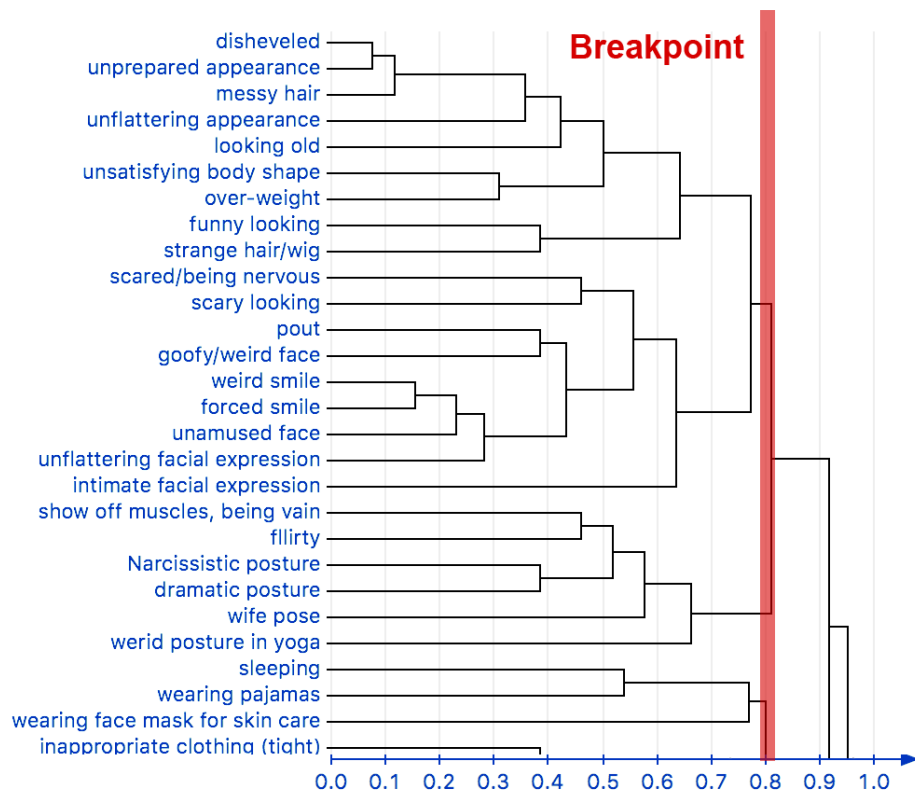


Figure 3.1: A part of the dendrogram. All items in this sort are listed vertically. Items placed next to each other vertically are more similar. The horizontal line from each item joins other items vertically, showing where items are grouped at higher levels of relationship [21].

while prior work [3, 136] identified “children” as a sensitive category, it is unclear that what makes this category sensitive. Many people share photos of their children on SNS regularly. Are all images including children sensitive? Our work revealed that specific types of photos of children are considered sensitive, such as when the child is nude, is wearing inappropriate clothes, or in a dangerous situation. We are not aware of prior work that has reported this type of nuance about the sensitivity of photos of children. It is not just that the photo contains a child, it matters what the child is doing or wearing. Similarly, [3] identifies the category *unflattering/embarrassing shots* which by itself may be too vague to guide any automated sensitive content detection. However, our results unpack this category in great detail, with subcategories such as *messy hair*, *looking old*, *strange hair/wig*, and *pout*, which may be more easily detected automatically, and furthermore, help us understand what types of occurrences in photos make people feel like a photo is unflattering or embarrassing.

Arguably, the additional detail provided by our taxonomy makes it more practical for privacy researchers, computer vision researchers, social scientists, and practitioners to apply in their work. For example, if computer vision researchers would like to identify sensitive content in photos, using prior work, they would not have known to train their systems to separate breastfeeding from other types of nudity or all photos of children, from photos of children in dangerous situations.

Category	Example
Nudity/Sexual (113)	- Genitals; naked person; butt crack; naked buttock; breasts; naked same-sex; cleavage; bare back; shirtless; masturbation; sexual action; erotic online photo; sexualized objects; sexual motion with statue; suggestive posture
Mitigated (10)	- Breastfeeding; bent over showing behind; kissing
Close up (6)	- Close up
Irresponsible to child/pet (8)	- Child in dangerous situation; child in inappropriate clothes; naked child; delinquent pet owner
Bad characters/ unlawful/ criminal (27)	- Infidelity/cheating; photo owner in dangerous situation; illegal drug; being physical abused; mug shot/get arrested; incriminating evidence

Appearance/facial expression (59)	- Ungroomed; messy hair; unflattering appearance; looking old; unsatisfying body shape; overweight; funny looking; strange hair/wig; scared/being nervous; scary looking; pout; goofy face; weird smile; forced smile; unamused face; unflattering face; intimate expression
Pose (8)	- Show off muscles, being vain; flirty; Narcissistic posture; dramatic posture; wife pose (no sexual meaning); weird posture in yoga
Not professional at work (9)	- Activities that break work rules; negative attitude towards work; look unprofessional at work; co-workers kissing
Sleep and grooming (5) Clothing (33)	- Sleeping; wearing pajamas; wearing face mask for skin care - Tight clothing; revealing clothing; wearing body-shaping corset; changing clothes; not fashion outfit; tacky outfit; wearing bib for dining; cross-dressing; wearing disposable gown
Drinking/party (30)	- Drinking; drinking a body shot; drunk; hang out with friends; at a party
Food/smoking (8)	- Diet/food; unhealthy eating; smoking
Medical condition/visible blood (40)	- Black eye; swollen eyes; abscess; peeling skin; blister; rash; bad teeth; bad skin condition; acne; moles; stretch marks; gore; bloody person; bloody animal; dog bite; body injury; eye removal; surgery wound; baby waste; period blood
Medical treatment (7)	- In hospital with doctors; on a stretcher; with hospital ward mates; wearing oxygen mask; in medical treatment; family member medical accident
LGBTQ/Religion (6)	- Lgbtq event; being gay; same-sex partner; spiritual inclinations; religious clothing; people in different races
Political and vulgar text (13)	- Negative texts/meme; vulgar/explicit texts/meme; politically incorrect texts/meme; racist texts/meme; violation of religious dogma

Other people (74)	- Grandparents; family member; significant other; step-parents; step-children; young family member; older children; friends; family member who passed away; photo owner's children; estranged people; ex-significant other; people who is unacceptable by photo owner's family
Personal moment (14)	- Affectionate moment with significant other; affectionate moment with friends
Event (5)	- Family event/party; children beauty pageant; funeral
Photo owner (18)	- Photo owner non-sensitive body parts; photo owner him/herself; selfie
Bad quality of photo (2)	- Unclear photo; old photo
Objects/personal assets (11)	- Pumpkin pie; video game; cat; kitten; dog; boyfriend's cat; car; PC; money; expensive necklace
Unorganized home (9)	- Nasty toilet; dirty bedding; uncleaned swimming pool; messy room
Gun (7)	- Gun; fake gun; hunting
Space/relaxed phase at home (8)	- In bed; bedroom; in bathroom; leisure at home; living room; house
Toilet (3)	- Using toilet; head in toilet
Other people's information (9)	- Screenshot of other's baby registry; friend's to do list; brother's diploma; person in the photo considers it private; save others photos without permission
Personal identifiable information (24)	- Vehicle license plate; driver license; order history; bank account; debit/credit card; online password; private project; only for job purpose; home address; to do list; body weight number; confidential work photo; vacation location

---

Table 3.2: Twenty-eight sensitive categories with the number of data points in each category and their examples. Each word or phrase in the example column represents a unique piece of sensitive content, as identified and named by participants, in response to the open ended question, “What content in this photo do you consider sensitive?”.

### 3.3.2 Sharing Preference by Sensitive Content

We analyzed people’s sharing preference of sensitive content via a linear mixed-effects model with fixed slopes and random intercepts set for each participant, where the outcome variable was the likelihood to share and the predictor was the sensitive content category. We conducted Tukey posthoc tests to compare all possible category pairs since it accounts for multiple comparisons and adjusts p-values accordingly. Note that we reversed the rating of recipient “only me,” because a higher likelihood to keep the photo private means a lower likelihood to share with others which may bias the results. We then conducted a chi-square test to evaluate the significance of fixed effects. The overall  $\chi^2$  shows significant variation among 28 categories,  $\chi^2(27) = 139.65$ ,  $p < .0001$ , indicating that sensitive categories affected sharing likelihood differently. Though the categories are all considered sensitive, we know, from Figure 3.2 which illustrates the overall likelihood to share a category across all recipients, some categories are even more sensitive than others.

People are least likely to share *other people’s information*. We found differences between *other people’s information* ( $M = 1.07$ ,  $SD = 0.82$ ) and *personal identifiable information, not professional at work, photo owner, drinking/party, political/vulgar text, other people, objects/personal assets*, and *bad quality of photos* (all  $d^1 \geq 0.75$ , all  $p < .05$ ). *Nudity and partial nudity* ( $M = 1.65$ ,  $SD = 1.46$ ) is less likely to be shared compared to *personal identifiable information, photo owner, drinking/party, other people, objects/personal assets*, and *bad quality of photos* (all  $d \geq 0.45$ , all  $p < .05$ ). Though the means of *medical treatment* ( $M = 1.38$ ,  $SD = 2.04$ ) and *sleep/grooming* ( $M = 1.53$ ,  $SD = 1.06$ ) are low in Figure 3.2, the variation in the data and fewer data points lead to non-significant comparisons with other categories, except for the difference between *sleep/grooming*

---

<sup>1</sup> $d$  represents Cohen’s  $d$ .

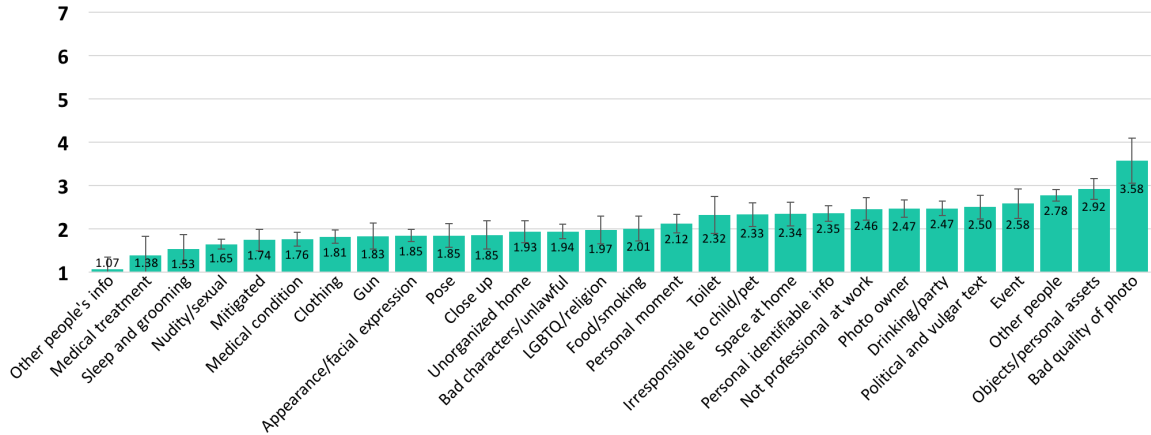


Figure 3.2: Participants' likelihood to share each sensitive content category across all recipient groups.

and *other people* ( $d = 0.63$ ,  $p < .05$ ).

### 3.3.3 Sharing Preference by Recipient

We created another mixed-effect model to look at the sharing preference by recipient. Again, there is a variation among all recipient groups,  $\chi^2(19) = 3112.25$ ,  $p < .0001$ . Unlike the similar likelihood rating between categories in the last section, the blue bars in Figure 3.3 clearly show a four-level pattern: only myself, significant other, people who are close to the photo owner, and people who are not close or work-related.

Again, we did Tukey post-hoc tests to compare recipient groups. As we expected, besides keeping photos *private*, people are most likely to share sensitive content with their *significant others* ( $M = 4.33$ ,  $SD = 2.37$ ) compared to all other recipients (all  $d \geq 0.79$ , all  $p < .001$ ). On the other hand, people are less likely to share with people who are not close to them (e.g., *people only met once or online*, *ex significant others*, *friends of friends*, *distant friends and relatives*) and people in their work no matter how close they are (e.g., *colleagues*, *supervisors*) when comparing with *close relatives* ( $M = 2.56$ ,  $SD = 1.68$ , all  $d \geq 0.31$ , all  $p < .01$ ), *close friends* ( $M = 2.65$ ,  $SD = 1.68$ , all  $d \geq 0.36$ , all  $p < .01$ ), and *household members* ( $M = 2.71$ ,  $SD = 1.68$ , all  $d \geq 0.38$ , all  $p < .01$ ). In terms of age (three red bars in Figure 3.3), people are more likely to share sensitive content with *people in their age group* ( $M = 2.24$ ,  $SD = 1.68$ ) than *younger people* ( $M = 1.86$ ,  $SD = 1.68$ ,  $d = 0.23$ ,  $p < .001$ ). However, we did not find evidence for a difference between *people in their age group*



and *older people*. The two yellow bars in Figure 3.3 show the means of recipient in *different gender* and *same gender* with photo owners, but we did not find evidence for a difference between these.

Besides the overall plots (Figure 3.2 and 3.3), we explored if there were interactions between the sensitive content categories and recipients. We did individual plotting by subsetting each sensitive category, then compared the overall plot with the subset plots to see if there were abnormal higher or lower bars. We also plotted the subset of each recipient. Most plots followed the pattern in the overall plot.

For plots which did not align with the overall plots, we conducted follow-up Tukey post-hoc tests within each subset. In the subset of *nudity* category, besides keeping the photo *private* and excluding the age and gender groups, the likelihood of sharing with *significant others* ( $M = 4.12$ ,  $SD = 2.62$ ) are much higher than any other recipients (all  $d \geq 1.48$ , all  $p < .001$ ), while there is no difference among other recipients. The trend of *personal moment* is the same as *nudity*. Though the sharing likelihood among *close friend*, *household member*, and *close relative* is somewhat similar in the overall plot, we noticed that people are more likely to share photos that depict when they are *unprofessional at work* with their *close friends* ( $M = 4.96$ ,  $SD = 2.24$ ) and *significant others* ( $M = 5.65$ ,  $SD = 1.22$ ) compared with all other recipients (all  $d \geq 0.97$ , all  $p < .05$ ), except for *close colleagues*. In the *event* subset, since the content is mostly family-related, there is no difference in the likelihood to share with *significant others*, *household members*, and *close relatives* (all  $p > .05$ ). For *personal assets*, except for the comparison with *significant others*, there is no difference among the combinations of *household members*, *relatives*, *friends*, *ex*, *colleagues*, *supervisors*, *friends of friends*, and *friends only met online* or *met once* (all  $p > .05$ ).

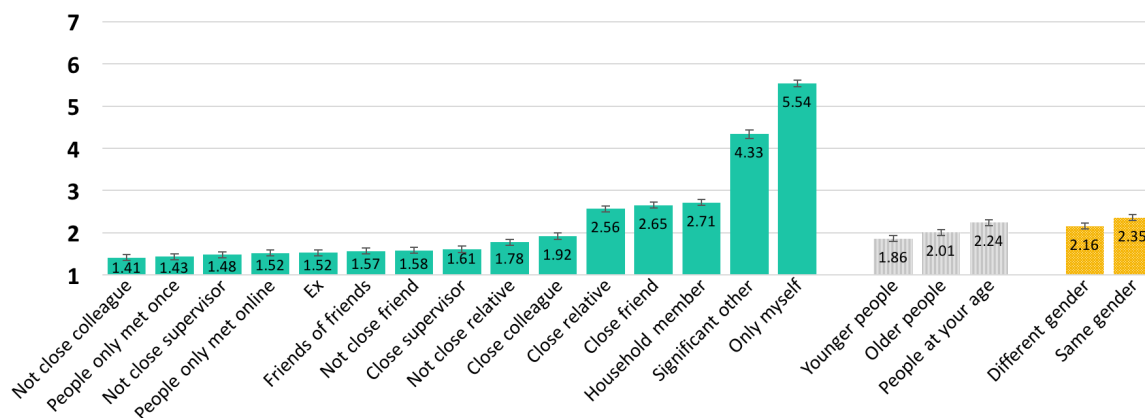


Figure 3.3: Participants' likelihood to share with each recipient across all sensitive content categories.

## 3.4 Discussion

### 3.4.1 With Whom Do People Share Or Not Share?

Previous work identifies several clusters of recipients treated similarly when sharing information, in which *significant other* is treated differently than any other recipients [213]. Indeed, our result suggests that in general, *significant other* is the group that people are most likely to share a sensitive photo with. However, this pattern is reversed in situations where the photo’s content shows the photo owner cheating. Participants reported qualitatively that they would not share these photos with a spouse because “it creates problems in my marriage.”

Following significant others, people are similarly likely to share sensitive photos with people who are emotionally or biologically close to them: *household members*, *close friends*, and *close relatives*. Kairam et al’s study on selective sharing in Google+ suggests the same pattern in which this cluster of recipients is categorized as ‘strong ties’ recipients [135]. However, we found an exception that people do not mind sharing photos in which they look unprofessional at work with *close friends*, but they prefer not to disclose them with *household members* and *close relatives*. The reason behind this could be that the content is mostly “inappropriate” humor and joking (e.g., give the middle finger with a goofy face) in the workplace which could be fun when sharing with friends; however, household members might worry about their attitudes towards the work and possible negative judgments from supervisors [54]

Though sharing information with work-related recipients on SNSs is prevalent because of the specific sharing needs for workplace SNS use [28, 254], the likelihood of sharing sensitive content is generally very low. First, people share very little sensitive information with their *colleagues* and *supervisors*, no matter whether they are close or not. This result may reflect the phenomenon described in a longitudinal study about social isolation in the workplace suggesting that people find it difficult to establish friendships with their colleagues or supervisors [134]. Moreover, some content may be sensitive because it has the potential to reveal “white lies” or remove “plausible deniability” at work [249]. For example, one participant reported she would refuse to share a photo with her colleagues and supervisor because she “took this when I had called in sick to work one day and was instead hanging out with my boyfriend.” Her supervisor might consider her behavior an irresponsible abuse of the company’s sick leave policy.

Next, as one participant said, “I wouldn’t want to share to people I don’t know well,”

suggesting people are hesitant to share sensitive content with acquaintances or ‘weak-ties’ [135], such as *distant friends or relatives, friends of friends, or people they only met once or online.*

### 3.4.2 Why Don’t People Share?

When asked to identify the sensitive content, participants’ responses revealed many of the reasons they don’t want to share certain sensitive photo content. We can summarize the reasons behind the desire not to share sensitive photo content as follows: first, avoiding getting into trouble (e.g., social tension, losing job, law violation); second, avoiding harming their impression management (e.g., appearance); third, avoiding content leakage that may harm themselves, family, and property safety (e.g., home address); last, maintaining a comfortable social distance with others (e.g., not being monitored when relaxing at home).

Interestingly, *other people’s information* is rated as the content least likely to be shared even if the underlying content itself would otherwise be less sensitive (e.g., friend’s todo list, brother’s diploma). In this way, the ownership of the photo clearly affects perceived sensitivity. This result is in line with prior work by Eiband et al. who found that people do not like being shoulder surfed even when content (e.g., third persons’ information) is not sensitive. Their work suggested a reason behind this is that the content may reveal relationships [79]. It also suggests that people try to avoid social tension caused by unauthorized sharing and saving of others’ photos [29]. People also generally respect others’ privacy concerns [25]. For example, in this study, we found that people are unlikely to share a photo if a person in the photo considers it private. However, multi-party sharing conflicts may occur if the photo uploader is not aware of others’ concerns [273]. Existing privacy controls on SNS are unable to protect a user from content leakage by their friends [273], hence emerging work has developed multi-party privacy control mechanisms to alleviate this problem [123]. On the other hand, unauthorized saving of others’ photos may not only cause social tension but may also harm impression management. For example, one participant in our study noted: “They’d think it’s creepy that I have it [a female friends’ bikini photo that he had saved].”

*Nudity, sexual or mitigated* content is another common concern that has been identified in prior work [215, 304] and is substantiated by our study. Three reasons for this concern were revealed in our qualitative data. First, photos with nudity or sexual content are mostly sent only between significant others to maintain a romantic relationship [274]. However, leakage of these photos damages people’s impression management and reputation, and even leads to social ostracism,

depression, and suicide [237]. Second, sharing sexually suggestive photos may become a potential threat to physical safety via off-line contact [199]. Last, disseminating other people’s nude photos violates the law and may get photo uploaders into legal trouble.

Aligning with previous work [210], *medical treatment* and *medical condition* are both rated as very unlikely to be shared with others. People express concerns that employers may change hiring decisions or limit job opportunities based on seeing their medical information [105]. This type of content could also harm their impression management since it indicates an unhealthy condition that may show the person’s weakness to photo viewers. People tend to share photos that depict socially desirable characteristics [74, 119], but avoid sharing photos which are not socially desirable such as photos showing a *disorganized home, food and smoking, or a toilet*.

Besides managing impression, SNS users selectively share photos because they want to maintain their personal space free from intrusion, which is similar to maintaining a comfortable social distance in the off-line world [2]. Hence, people are not likely to share content about their *sleep and grooming, personal moment, space or relaxed phase at home*.

Other types of content that may get photo uploaders into trouble are *bad characters, unlawful and criminal evidence*, and content showing that they are *irresponsible in regards to children or pets*. Regarding a photo that depicts a water pipe with cannabis, one participant stated: “I could lose my job and friends if this photo were posted to my Facebook. It is sensitive because it could nuke my life.”

Though *personal identifiable information* and *personal assets* are not the top sensitive content in Figure 1.1, their leakage could lead to personal, family, and property safety issues. For example, online fraud and identity theft attacks can be perpetrated by collecting information such as a user’s name, online password, SSN, or bank account information from multiple sources [26, 201].

### 3.4.3 Privacy is Subjective Except for the Consistency

There is a debate in the literature about the extent to which privacy is subjective. While privacy is a universal necessity for the proper functioning of human society [200], it may be subjective and dependent on complex social, cultural, and historical factors [62, 93, 200]. At an individual level, privacy could vary among people based on the environment and prior experience which could encourage them to reveal more or less information [62]. What some people are comfortable sharing others might consider a threat to the privacy [287]. On the other hand, prior work on people’s privacy

concerns suggests at least some consistency. For example, a study on photo privacy detection suggests that people generally agree that certain types of content should not be shared, such as photos of a driver’s license, a legal document, and a pornographic photo [278]. Another study situated in an online context found that there is a consensus about certain privacy concerns such as personally identifiable information (e.g., credit card number, SSN, fingerprints) and sensitive content (e.g., religion, sexual preference, wage) [10]. Some other commonly identified categories of private items in personal photos include human faces, sensitive text, and objects such as cars and animals) [111].

The categories of sensitive content suggested by prior work are consistent with our findings, suggesting the taxonomy we report here is not merely a reflection of the subjective privacy preferences of the participants in our study. Instead, taken together, our taxonomy and the prior work we describe here suggest that there is consistency in some aspects of privacy, such as what people consider sensitive content in photos. Furthermore, even assuming that privacy is subjective would not challenge our taxonomy of sensitive content. Though people may have different privacy concerns about their personal photos, there is consistency in the types of content that people feel is sensitive and potentially privacy-invasive. Even if an individual does not feel that their own photo containing some of this content is sensitive to *them*, there is usefulness in helping that person understand that *others* may consider it sensitive, because we know that people tend to avoid sharing photos they know may offend others [255]. Moreover, we know that there is a desire to use machine learning approaches to find consistencies regarding sensitive content [303, 304]. In our study, we also find a consistent pattern of privacy concerns from participants’ personal photos. Our goal was to identify consistencies in people’s perception of content sensitivity. People’s consensus can address the reported subjective nature of aspects of privacy [298], and this consensus is obtained through our study. We collected 563 data points of which only 181 are unique that again suggests that there is some agreement about content sensitivity which may be useful to understand.

#### **3.4.4 A New Method for Sensitive Content Elicitation**

As we described in the background section, existing methods for identifying sensitive content in photos are severely limited. However, the method we introduce in this paper is not subject to the limitations we outlined for ML approaches, for example, because we do not rely on existing general purpose databases and we provide participants with alternative, privacy-preserving, ways to identify sensitive data while. Our method gives participants the option to find a photo - similar to

their own sensitive photo - and share that one instead, or just describe the photo. We can see the success of our method and the biases of previous methods by comparing it to the categories elicited using a ML approach applied to the categories from [303, 304]. Whereas we found that people are unlikely to upload photos depicting that they are irresponsible to children or pets, this category was not present in the categories generated by [303, 304]. Moreover, from our study, we learned that other people’s information is a top concern even if the content itself seems less sensitive (e.g., friend’s to-do list, brother’s diploma). On the other hand, a ML approach is unable to distinguish between a person’s own information and other’s information, which results in an inaccurate, or at least incomplete, classification of sensitivity.

One straightforward way our work could work in concert with ML approaches is by introducing our photo elicitation method as a way to supplement existing datasets or to create a new dataset of sensitive photos from scratch. This method could be used to gather and add new images with important private content to existing general-purpose image datasets which would then make them useful for image privacy tasks. An important question that arises is whether and how private content collected using our elicitation method may be made ethically available to ML practitioners. One potential solution we propose is to use the taxonomy in combination with advanced privacy-preserving ML approaches, such as transfer learning [217, 277]. In transfer learning, a model can be first pre-trained with sensitive content and then shared along with the trained model parameters for further use without directly sharing sensitive content. Such models can also be fine-tuned according to the requirements of different ML approaches.

Another way our work could benefit ML for privacy tasks is by using the taxonomy itself as a point of comparison. For example, we could compare the categories in our taxonomy to the categories in the Flickr dataset [303, 304]. Doing this, we see that while we found that people are unlikely to upload images depicting their medical condition or treatment, this category was not present in the categories generated by [303, 304]. In this way, our taxonomy can serve as one form of ground truth for categories generated via ML, that could be further triangulated with other sources of ground truth.

### **3.4.5 Implication: A Usage Scenario for SNSs**

The only photo privacy protection technique currently provided by most SNSs (e.g., Facebook) is choosing or excluding certain recipient groups [84]. Even when sensitive content is just a

small part of a photo, uploaders' only options are to either share the sensitive content as part of the photo or withhold the entire photo from some or all recipients which leads to a large sharing loss [255]. Furthermore, it can be overwhelming for users to have to make privacy decisions about every photo they share. Uploaders may have a large number of connections (e.g., friends on facebook) making it difficult for them to sort through all potential recipients and make decisions about desirable recipients every time they upload a photo [34]. Current privacy management options that allow users to choose or exclude certain recipient groups only target one side of the photo-sharing equation (recipients, but NOT content). Our work lays the foundation for new solutions that could help people to make decisions about photo sharing easily. The taxonomy can be used to inform an automatic photo privacy protection system that combines existing recipient control mechanisms with our proposed solution addressing controlling content. For example, a new system could help automatically identify content that the uploader may find sensitive or that may be offensive to others so that it can be highlighted for additional scrutiny by users, who can then make sharing (or not sharing) decisions based on additional aspects of context. The taxonomy may also be useful for solutions aimed at reducing users' effort toward recipient selection. We uncovered which recipient groups would be most likely targets for exclusion when sharing certain content. These recipient groups could be highlighted for additional scrutiny or become part of user-tailored privacy solutions which provide guidance based on users prior behaviors and preferences [141].

A usage scenario could be the following: upon uploading a photo, the system detects possible sensitive content in the photo based on our categories and highlights the content for review by the person who uploaded the photo; next, depending on the sensitive content, the system could suggest applicable obfuscations (e.g., cartooning, inpainting [180]) that when applied, would prevent some viewers from seeing the sensitive content as shown in Figure 3.4 (e.g., removing/inpainting the beer can). Afterward, the system gives the photo uploader recommendations about viewers who the uploader may wish to exclude from the recipient list. Together, these approaches could dramatically improve the privacy and sharing options available to people who share photos online.

### 3.5 Limitation and Future Work

One limitation of our work is that we only focus on U.S. Internet users, and therefore the results of our study only inform us about this population. The sensitive content elicited from U.S. par-

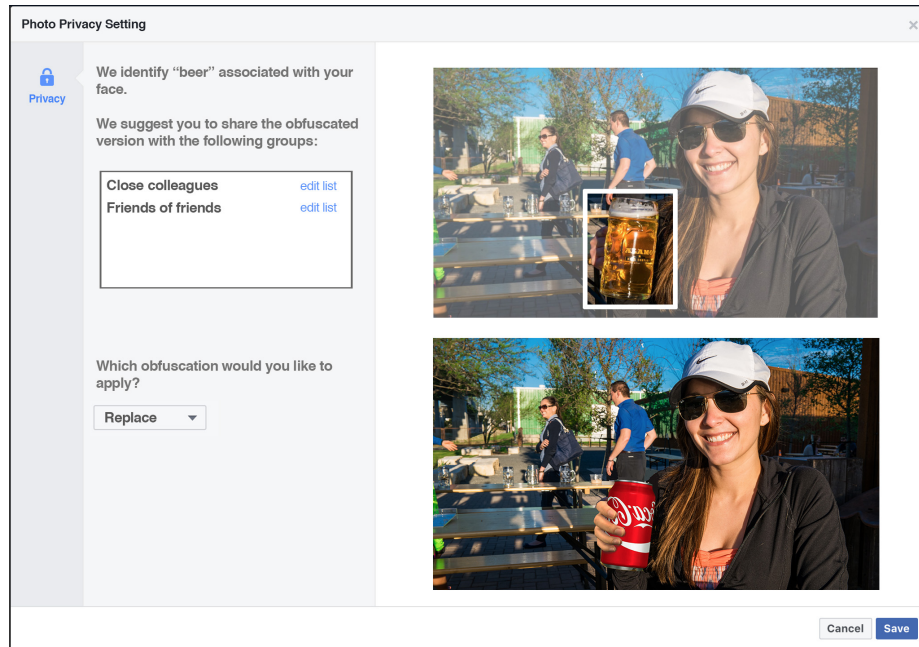


Figure 3.4: Example interface of content detection and obfuscation.

ticipants could be useful to designers and practitioners interested in designing for a U.S. population. Furthermore, researchers who have the resources to study cross-cultural privacy (e.g., [174, 291]) may be able to use the methods we describe here to determine whether different sensitive categories emerge across cultures. The sample for our card-sorting study is also limited. The participants for the card sort study were all members of the university community. It is possible that other participants in a replication could group and/or name categories differently. For example, instead of grouping the item “acne” into “medical condition” other participants could have grouped it into “grooming.” However, it is less likely that other participants would have different perspectives about the category for many items such as “blister” or “surgery wound”; there is not another category besides “medical condition” where they could reside. However, reviewing Table 3.1, most items seem intuitively to fit within each category. Future work could replicate the card sort using the items we introduce here (listed in Table 3.2).

Another limitation is that we were only able to collect sensitive content that participants would identify in one of three ways: 1) by uploading a photo from their own phone, 2) by uploading a photo similar to a sensitive photo from their phone, or 3) by describing a photo and the sensitive content in it. It is possible that people are unwilling to identify content that is so sensitive



that they do not want to reveal it to researchers in any form. Despite this limitation, we see the methodological innovation we report in this paper as a step in the direction of getting closer to the ideal of understanding sensitive content categories. Notably, we see it as an improvement over complementary approaches such as those that rely on applying machine learning to photos posted on Flickr [303, 304].

Last, while we did investigate some individual differences (e.g., age and gender), our results mainly represent general sharing preferences. Future work should investigate individual differences in photo sharing preferences across different demographic variables. Finally, since this work demonstrated that the photo elicitation method can help elicit content that would otherwise be missing from datasets of sensitive photos, future work could investigate how the method could be adapted to other types of data such as video.

### **3.6 Chapter Conclusion**

We report a taxonomy for photo privacy that describes what content is considered sensitive and how sharing preferences differ across potential photo recipients. We derived the taxonomy by synthesizing existing literature, collecting photos that contain sensitive content from 116 participants and recording their sharing preferences with 20 recipient groups and then conducting a card sort to surface 28 user-defined categories of sensitive content. This taxonomy can serve as a framework for understanding photo privacy, which can, in turn, inform new photo privacy protection mechanisms. Moreover, we introduce a new sensitive content elicitation method which overcomes many of the limitations of prior approaches. Understanding the sensitive content that needs to be protected, in the next chapter, I investigate effective and usable obfuscations that can be applied to the sensitive content.

## Chapter 4

# Study 2: Identifying Effective and Usable Obfuscations

Note: This work was published at CSCW 2018 [180].

### 4.1 Introduction

From the last study I learned that what content could be obfuscated, hence, for the next step, I need to identify some obfuscations that can be applied to the sensitive content. While some prior work has investigated methods to hide elements of photos to be shared online, it has been limited to a few approaches, most notably blurring and pixelating [124, 284], which are ineffective at preventing human and machine identification [161, 195]. Moreover, researchers did not make any arguments about why and how the obfuscation choices were made. In another study that I collaborated, we applied obfuscations to scene elements in photos, however, the obfuscation options were limited to blurring, pixelating, masking, and silhouette [109]. Hence, I aimed to explore other alternatives that are both effective and have a good viewer experience.

Recognizing the need for more effective photo privacy-enhancing obfuscations, we identified silhouette [216], box masking [305], avatar [235], point-light [49], bar [305], and inpainting [216, 305]. Moreover, recognizing that the audience for online photos is human beings, we investigated human (vs. machine) perception of these options (in terms of photo satisfaction, perceived photo information

sufficiency, photo enjoyment, social presence, obfuscation likability and preference), as well as their ability to identify content in the photos. To our knowledge, there is no existing research that addresses both effectiveness against human recognition and users’ perceptions of obfuscations as privacy-enhancing tools.

Our results show that even though people are generally satisfied with *blurring* and *pixelating*—the two most investigated and widely-adopted obfuscation methods—these methods do not enhance privacy. When developing collaborative privacy management systems, researchers should consider alternative privacy-enhancing obfuscations, such as *inpainting* and *avatar*, which are both effective and likable. Though our findings focus on face and body, they can be further applied to other PII, such as object and location.

## 4.2 Method

### 4.2.1 Overview

We conducted an experiment with 271 participants to understand how type of obfuscation and region the obfuscation is applied to influenced effectiveness and users’ perceptions (satisfaction, perceived information sufficiency, photo enjoyment, social presence, and likability).

### 4.2.2 Participants

Three hundred and forty seven participants from United States were recruited via the Amazon Mechanical Turk. While imperfect, it is considered one of the better sampling strategies because Turkers are relatively more diverse than the samples collected by other means (e.g., U.S. college samples) [39]. We paid participants \$1.50 to complete the study [239]. To ensure the data quality, we set restrictions to only include MTurk workers with high reputation (above 95% approval ratings), and with the number of HIT approved being greater than 1000 [223]. Excluding the data of participants who failed more than one attention check questions, the final sample size is 271 (131 men and 140 women). Fifty-seven participants were from the Midwest region; 99 were from the South; 66 were from the West, and 49 were from the Northeast [40]. Forty-three percent ranged in age from 25 to 34; twenty percent ranged from 35 to 44; and forty-eight percent is was from 45 to 54. Seventy-six percent was White. Ninety-eight percent of participants used Internet most of the day or several

times a day; and 72% visited SNSs most of the day or several times a day.

### 4.2.3 Experimental Design

We used a 8 (privacy-enhancing obfuscation) by 2 (body region) experimental design. The eight obfuscation methods were: *blurring*, *pixelating*, *silhouette*, *avatar*, *point-light*, *masking*, *bar*, and *inpainting*. The two regions were face and body. Please see below sections for descriptions of why we chose them.

#### 4.2.3.1 Regions

Different elements in an image can be considered sensitive: for example, people, personal belongings, affiliation, and privacy information [122, 284]. For this study, we decided to study the recognizability of people for a variety of reasons, but one very compelling reason was because of prior work on human recognition of faces [27, 65, 146], machine recognition of faces [161, 195], and obfuscation of humans in video [33, 149]. We chose two regions to obfuscate: face and body (which includes the face). Masking the face is the most common strategy for hiding the identity of a person in photos or videos [301, 98, 124]. The face also has special meaning and significance in the human visual system. The perceptual process in facial recognition is different from the process in recognizing non-facial stimuli, that faces are recognized at the individual level [89]. However, prior work on video surveillance suggests that masking only the face is ineffective: obscuring the entire body is more effective than obscuring only the face [52].

#### 4.2.3.2 Obfuscation Methods

We chose eight redaction tools from previous work on online photo privacy, video surveillance, and video monitoring [49, 216, 262, 305]. We did not investigate some tools that were less applicable to photo redaction. For example, “see-through (translucent) [305]” and “monotone [305]” are excluded because the semitransparent or monochrome subjects may not be identified accurately in videos where they are dynamic, while in a photo, people can easily identify a static subject. Generally, we excluded the tools that were either ineffective or overlapping. The eight redaction tools we studied are listed in Table 2.3, including blurring, pixelating, silhouette, avatar, point-light, bar, masking, and inpainting.

We also tested the baseline condition of no obfuscation (*as is*). We left out three region by obfuscation combinations resulting in 14 total conditions (including *as is*; see obfuscation methods Table 2.3 and its caption). We did not test the following obfuscations for face: *inpainting*, *point-light* and *bar* because we anticipated these might make viewers uncomfortable. For example, *inpainting* just the face would have resulted in what appeared to be a headless person, which we anticipated might be jarring to view.

## 4.2.4 Stimuli

### 4.2.4.1 Targets

Target is the person in a photo who needs to be identified. We selected targets from racial categories broadly representing the racial makeup of the United States [41] including white, African American, Asian, and Hispanic and Latino, who were unknown to participants. The target photos were taken by our lab and researchers. We applied each of the 14 obfuscations to all targets with 2 different backgrounds resulting in 392 unique images.

### 4.2.4.2 Backgrounds

We selected the backgrounds and background people photos online which had licenses that allowed for reuse and modify, and photos taken by our lab and researchers, cut them out, and reassembled them to include the target person (Figure 4.1 ). Each photo has the same number of background people (three people) and similar background (campus building etc.).

### 4.2.4.3 Photo Creation

We used Photoshop to create photos so they would be consistent, except for elements we intentionally varied (e.g., target or obfuscation). Each photo consists of the target with an obfuscation applied, three non-target people, and a background (Figure 4.1 ). To generate a complete set of experimental stimuli, we created an image of each target/obfuscation pair and overlaid these on each background. The experiment platform randomly selected the combination of target, obfuscation condition and background. In total, the stimuli set has 392 unique photos (14 targets \* 14 obfuscations \* 2 backgrounds).

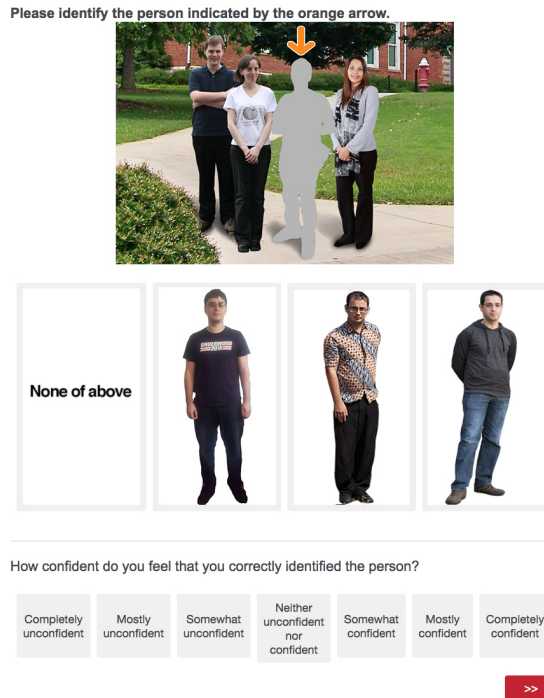


Figure 4.1: Experiment interface with one stimuli and ID photo examples

#### 4.2.4.4 ID Photo

For each target we collected one ID photo (e.g., the second person from the left in Figure 4.1), and three ID photos of similar looking people. A similar looking person might be the same gender, have a similar hair style, skin color, body shape and/or height (see the right two people in Figure 4.1).

### 4.2.5 Measurements

We measured obfuscation effectiveness using identification success and confidence, and users' experience via existing, psychometrically validated Likert scales.

#### 4.2.5.1 Obfuscation effectiveness

- Identification Success.** We measured identification success by asking “Please identify the person indicated by the orange arrow.” Four answer choices included three ID photos and “None of above.”

- **Identification Confidence.** After each identification, we measured confidence using the question “How confident do you feel that you correctly identified the person?” Participants rated their response on a scale from 1 ‘Completely unconfident’ to 7 ‘Completely confident,’ where the higher score meant more confident [230].

#### 4.2.5.2 Users’ experience

Next, we measured the following four aspects of the privacy-enhancing obfuscation in the photo. All the responses used 7-point Likert scale from 1 ‘Strongly disagree’ to 7 ‘Strongly agree.’ For participants’ ease of use, we adapted all scales to 7-point. Additionally, 7-point scales are more suitable for electronic distribution [91] and the data collected is more accurate than other point scales [130].

- **Photo Satisfaction.** We measured perceived photo satisfaction using the item “The photo is satisfying” derived from the image appeal scale [59].
- **Perceived Photo Information Sufficiency.** We selected a single item “The photo provides sufficient information” from the photo information quality scale to measure the perceived information sufficiency [247].
- **Photo Enjoyment.** We measured perceived photo enjoyment using the single-item photo enjoyment scale [233].
- **Perceived Social Presence.** We measured perceived social presence using the item “There was a sense of human contact when I saw the photo” from perceived social presence scale [154].
- **Obfuscation Likability.** We measured likability of each obfuscation using the item “I like the \_\_\_\_\_ obfuscation” which was derived from the interface preference scale [203].
- **Obfuscation Preference.** We asked participants’ preference for each obfuscation with the question “If you could use any of the obfuscations for photos you post on online social networks, which one, if any, would you like to use?” We followed up this question by asking an open-ended question about the reason, and queried participants’ willingness to use the obfuscation they selected. We also asked participants, “Have you ever declined to upload a photo to an online social network for privacy reasons?” If yes, they additionally answered which obfuscation they might use in such a scenario, and their reasons.

Note that we also captured the time participants spent on each question so that we could exclude participants who used automatic survey response software.

#### 4.2.6 Procedure

Prior to the study, we conducted three pilot tests to check for bugs, gather data about the length of the study and ensure that the data collection worked well.

In the actual testing, first, participants accessed the experiment website (Qualtrics) via the publically distributed link through MTurk. After consenting, they answered six demographic questions and two social network familiarity questions. Next, they tested the browser and monitor size and followed resizing instructions to make sure they all viewed stimuli in a similar visual environment. Afterwards, they saw 14 obfuscation conditions examples with the descriptions as an overview.

Next, we trained participants about the tasks. During training, participants learned about the tasks they would perform, and completed two training trials. Participants then completed 14 trials where they saw photos with semi-randomly assigned obfuscation conditions and targets, and identified the target person. Participants saw all 14 conditions and 14 targets. There were no repeating conditions or targets. For example, in the first trial, if the photo includes condition 1-target 3, photos including condition 1 and target 3 will be excluded in future trials. Note that in most cases, the target was among the four choices offered, but there was around 21% chance that the target was NOT present. Afterwards, they rated their confidence, and rated the four statements about their feeling.

After finishing all trials, participants were shown 14 conditions individually, and rated their preference towards each condition. Then they answered a set of obfuscation preference questions. After all tasks, a random code was generated. Participants copied this code to MTurk to receive remuneration.

### 4.3 Results

The experiment was completed by 347 participants. We excluded the data of 76 participants who either failed more than one attention check questions, or answered some questions instantly (reaction time = 0), indicating the potential use of automatic responding software [223]. The final



		% of success	Odds ratio	95% CI	p-value
Face	Masking	41%	0.20	[0.14, 0.29]	<.001***
	Silhouette	45%	0.23	[0.16, 0.34]	<.001***
	Avatar	47%	0.26	[0.18, 0.37]	<.001***
	Blurring	64%	0.52	[0.36, 0.76]	.05*
	Pixelating	72%	0.73	[0.50, 1.08]	.97
Body	Inpainting	19%	0.07	[0.05, 0.10]	<.001***
	Masking	20%	0.07	[0.05, 0.11]	<.001***
	Bar	27%	0.11	[0.07, 0.16]	<.001***
	Point-light	28%	0.12	[0.08, 0.17]	<.001***
	Avatar	33%	0.14	[0.10, 0.21]	<.001***
	Silhouette	40%	0.20	[0.13, 0.28]	<.001***
	Blurring	67%	0.59	[0.41, 0.87]	.33
	Pixelating	67%	0.58	[0.40, 0.86]	.27
Baseline	As is	77%	NA	NA	NA

Table 4.1: Identification success rate, odds ratio, 95% confidence interval, and p-value by region and obfuscation for all cases where the *as is* is the baseline. The obfuscations are ordered by identification success of body region from lowest (most effective) to highest (least effective).

sample size is 271, which provides sufficient power for the statistical tests we planned (i.e., 271 is more than the required 225 suggested by our a priori power analysis to achieve a power of 0.85).

### 4.3.1 Obfuscation Effectiveness

The primary measures of obfuscation effectiveness are identification success and identification confidence. Identification success is the percentage of trials in which a participant correctly identified a target. If we were to recast this as obfuscation success, or the percentage of trials in which a participant was unable to correctly identify a target, we would subtract the identification success percentage from 100%. For example, if a participant achieved a 60% identification rate, the corresponding obfuscation rate would be 40%. Identification confidence is a self-reported rating of how confident the participant was that their identification was correct.

#### 4.3.1.1 Identification Success

We analyzed the identification results using signal detection [271]: hit (the target is present, and the response is correct), miss (the target is present, but the response is incorrect, such as selecting the wrong person, or “None of above”), correct rejection (the target is absent, and the response is “None of above”), and false alarm (the target is absent, but participants do not select “None of above”). Using this approach, we can classify identification success using three categories:

among all cases, among trials where the target is present, and among trials where the target is absent. In next paragraph, we focus on identification success among all cases, as shown in Table 4.1

As expected, the identification success of as is is the highest across categories (*all cases* (77%), *target present* (80%), and *target absent* (70%)). A Tukey post-hoc test based on a logistic mixed-effects model of *all cases* shows that the identification success of as is (77%) is higher than all obfuscations (all  $p < .05$ ) except for *body blurring* (67%), *body pixelating* (67%), and *face pixelating* (72%). In addition, the identification success of *blurring* (face: 64%; body: 67%) and *pixelating* (face: 72%; body: 67%) are similar to each other (all  $p > .05$ ), and much higher than other obfuscations (all  $p < .001$ ; see Table 4.1). The success percentage difference between *blurring/pixelating* and other obfuscations ranges between 17 and 48%, which suggests that, in addition to being stastically less effective, they are also practically less effective. The lack of a difference between *blurring* and *pixelating*, two of the most common obfuscations [98, 158], and as is (5-13%) on the other hand, indicates that they are ineffective protections against human recognition, regardless of whether they are applied to the face or the entire body.

#### 4.3.1.2 Body vs. Face

Overall, body-obfuscations were more difficult to identify ( $M = 45\%$ ) than face-obscuring obfuscations ( $M = 54\%$ ;  $p < .001$ ), indicating body-obscuring obfuscations are generally more effective than face-obscuring obfuscations. Looking at individual obfuscation methods, though, there was not always a difference between face and body. Face-obscuring obfuscations were about as effective as body-obfuscations for many of the less effective obfuscations including *blurring* (face: 64%; body: 67%), *pixelating* (face: 72%; body: 67%), and *silhouette* (face: 45%; body: 40%; all  $p > .05$ ).

Obfuscations that protect more details of the target including body avatar, body point-light, body masking, body bar, and body inpainting, tend to be more effective. While body inpainting performs the best, there is no difference among these obfuscations, except between *body inpainting* ( $M = 19\%$ ) and *body avatar* ( $M = 33\%$ ,  $p < .05$ ). It is also worth noting that for *target absent* cases, the correct rejection rate is higher either when the obfuscation transformation level is low (e.g., *as is*, *face blurring*) or when the the obfuscation shows no sign of a visible body (e.g., *body masking*, *body bar*, *body inpainting*). The reasons for these higher rates are different, though: in the

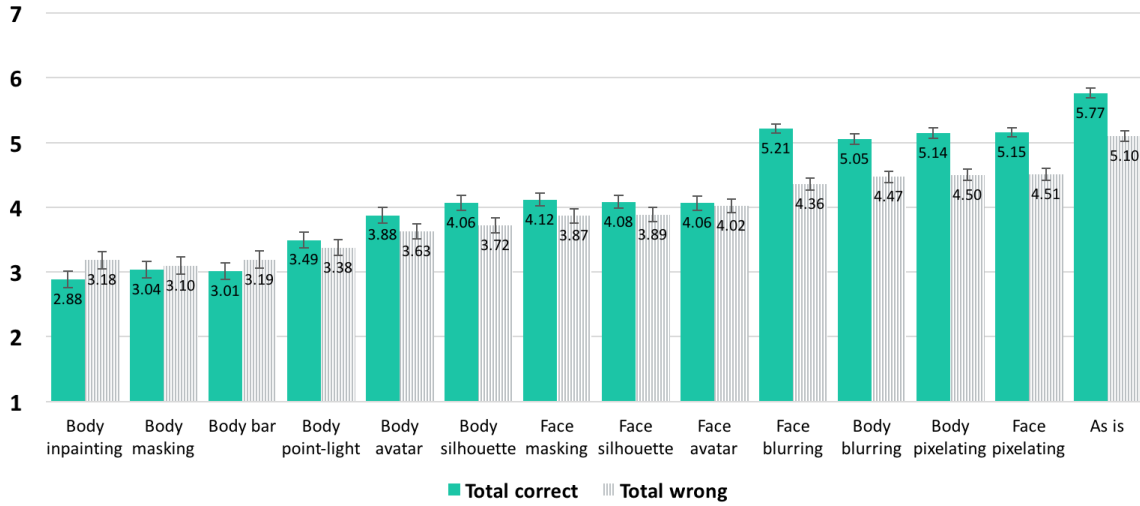


Figure 4.2: Means and standard errors of identification confidence of Total Correct (Hit + Correct Rejection) and Total Wrong (Miss + False Alarm).

former participants easily found out the target was not in three full-body ID photos, while in the latter cases there was no hint at all to identify the target, hence participants tended to choose “None of above” as a last resort.

#### 4.3.1.3 Identification Confidence

Identification confidence increases as obfuscation effectiveness decreases, or, in other words, people are more confident with answers when they can correctly identify the target. In Figure 4.2, for each obfuscation, the left bar represents identification confidence of total correct (hit and correct rejection) and the right bar represents total wrong (miss and false alarm).

We conducted a linear mixed effects model for total correct and total wrong, and compared obfuscation conditions using a Tukey post-hoc test. When the participant is correct, the identification confidence of *as is* is much higher than any other obfuscation methods, as expected (all  $d \geq 0.36$ , all  $p < .001$ ). Notably though, confidence when viewing *blurring* and *pixelating* obfuscations—while lower than *as is*—is higher than other obfuscations, with medium to large effects (all  $d \geq 0.58$ , all  $p < .001$ ). The means are above five (somewhat confident), providing further evidence that *blurring* and *pixelating* are ineffective. Moreover, identification confidence of total correct is higher than total wrong for these four methods (all  $d \geq 0.41$ , all  $p < .05$ ), and larger than the differences for any of the other obfuscations, indicating that it is also easier for participants to detect when

		Hit	Miss	Correct rejection	False alarm	Total correct	Total wrong
	As is	5.93 (1.08)	5.29 (1.38)	5.14 (1.52)	4.68 (1.34)	5.77 (1.23)	5.10 (1.39)
Face	Masking	4.07 (1.61)	3.89 (1.86)	4.28 (1.93)	3.74 (1.66)	4.12 (1.68)	3.87 (1.83)
	Silhouette	4.21 (1.68)	3.90 (1.84)	3.45 (1.57)	3.81 (1.62)	4.08 (1.68)	3.89 (1.79)
	Avatar	4.06 (1.73)	4.06 (1.80)	4.09 (1.95)	3.88 (1.36)	4.06 (1.76)	4.02 (1.71)
	Blurring	5.19 (1.22)	4.27 (1.55)	5.36 (0.95)	4.50 (1.56)	5.21 (1.19)	4.36 (1.55)
	Pixelating	5.18 (1.17)	4.64 (1.42)	5.03 (1.25)	4.31 (1.67)	5.15 (1.18)	4.51 (1.53)
Body	Inpainting	3.29 (1.90)	3.25 (2.24)	2.74 (2.18)	2.47 (1.78)	2.88 (2.10)	3.18 (2.21)
	Masking	3.68 (1.92)	3.08 (2.18)	2.69 (2.15)	3.25 (1.77)	3.04 (2.11)	3.10 (2.15)
	Bar	2.93 (1.64)	3.18 (2.18)	3.07 (2.29)	3.35 (2.09)	3.01 (2.06)	3.19 (2.16)
	Point-light	3.82 (1.81)	3.34 (2.05)	3.03 (2.24)	3.63 (1.86)	3.49 (2.02)	3.38 (2.03)
	Avatar	3.70 (1.79)	3.56 (1.97)	4.24 (2.20)	3.94 (1.63)	3.88 (1.94)	3.63 (1.91)
	Silhouette	4.34 (1.71)	3.73 (2.05)	3.31 (2.11)	3.71 (1.62)	4.06 (1.87)	3.72 (1.96)
	Blurring	5.09 (1.39)	4.76 (1.48)	4.84 (1.18)	4.03 (1.34)	5.05 (1.37)	4.47 (1.46)
	Pixelating	5.15 (1.39)	4.62 (1.50)	5.12 (1.20)	4.35 (1.29)	5.14 (1.36)	4.50 (1.41)

Table 4.2: Identification confidence for Hit, Miss, Correct Rejection, False Alarm, Total Correct (Hit + Correct Rejection), and Total Wrong (Miss + False Alarm) on a scale from 1 - 7 where 7 is most confident. Standard deviations appear in parentheses beside the means. Within face and body categories, the order of the obfuscations is from most to least effective.

they incorrectly identified the target (see Table 4.2 for means and standard deviations).

Conversely, as we see in Figure 4.2, mean identification of the five most effective obfuscation methods (those on the left side of Figure 4.2: body inpainting, body masking, body bar, body point-light, and body avatar) are all below four (neither unconfident nor confident) for both total correct and total wrong, indicating that participants were not confident about their identification, regardless of whether they correctly or incorrectly identified the target.

### 4.3.2 Users' Experience of Obfuscations

We analyzed users' experience of the obfuscations via five linear mixed-effect models, where the outcome variables were photo satisfaction, information sufficiency, enjoyment, social presence, and obfuscation likability, and the predictor was the obfuscation condition. We conducted Tukey post-hoc tests to compare all possible obfuscation pairs.

#### 4.3.2.1 Photo Satisfaction

We now know that some obfuscation filters are more effective than others, but how do they influence users' satisfaction with the photos? From the results of our linear mixed-effects model, the overall  $\chi^2$  shows significant variation among 14 obfuscation conditions,  $\chi^2(13) = 986.62$ ,  $p <$

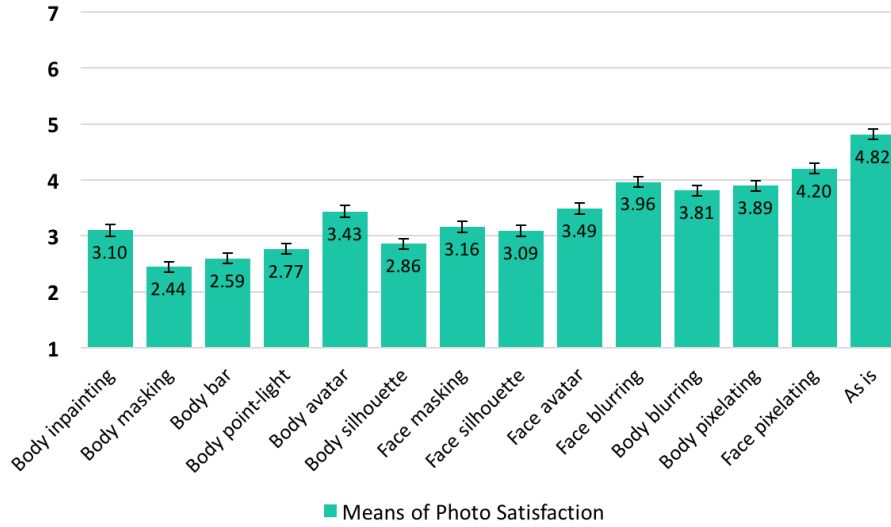


Figure 4.3: Photo satisfaction rating ( $M$  and  $SE$ ). Obfuscations ordered from most to least effective.

.0001, indicating that obfuscations affected satisfaction differently. Indeed, Figure 4.3 shows that participants are generally less satisfied with the more effective obfuscations (see Table 4.3 ). From the Tukey post hoc test on this model, the results show that participants are most satisfied with *as is* ( $M = 4.82$ ,  $SD = 1.62$ ) compared to any other obfuscations (all  $d \geq 0.37$ , all  $p < .001$ . As the smallest difference, the difference between *as is* (4.82) and *face pixelating* (4.20) has an effect size of  $d = 0.37$ ; while other effect sizes are all above 0.5, which represent medium or large effects). Participants are also satisfied with *face pixelating*, *face blurring*, *body pixelating*, and *body blurring*, but as mentioned before, these methods are not particularly effective.

Among the more effective obfuscations, participants are most satisfied with *face avatar*(1) ( $M = 3.49$ ,  $SD = 1.71$ ) and *body avatar*(2) ( $M = 3.43$ ,  $SD = 1.76$ ) with both scores higher than body masking ( $M = 2.44$ ,  $SD = 1.51$ ,  $d_1 = 0.59$ ,  $p_1 < .001$ ,  $d_2 = 0.56$ ,  $p_2 < .001$ ), body bar ( $M = 2.59$ ,  $SD = 1.55$ ,  $d_1 = 0.54$ ,  $p_1 < .001$ ,  $d_2 = 0.48$ ,  $p_2 < .001$ ), body point-light ( $M = 2.77$ ,  $SD = 1.54$ ,  $d_1 = 0.44$ ,  $p_1 < .001$ ,  $d_2 = 0.41$ ,  $p_2 < .001$ ), and body silhouette ( $M = 2.86$ ,  $SD = 1.49$ ,  $d_1 = 0.39$ ,  $p_1 < .001$ ,  $d_2 = 0.36$ ,  $p_2 < .001$ ). Moreover, the most effective obfuscation among all 14 conditions, *body inpainting* ( $M = 3.10$ ,  $SD = 1.73$ ), scores are higher than *body masking* ( $d = 0.39$ ,  $p < .001$ ) and *body bar* ( $d = 0.32$ ,  $p < .001$ ), and is also slightly (but not significantly) more satisfying than *body point-light*, *body silhouette*, and *face silhouette*.

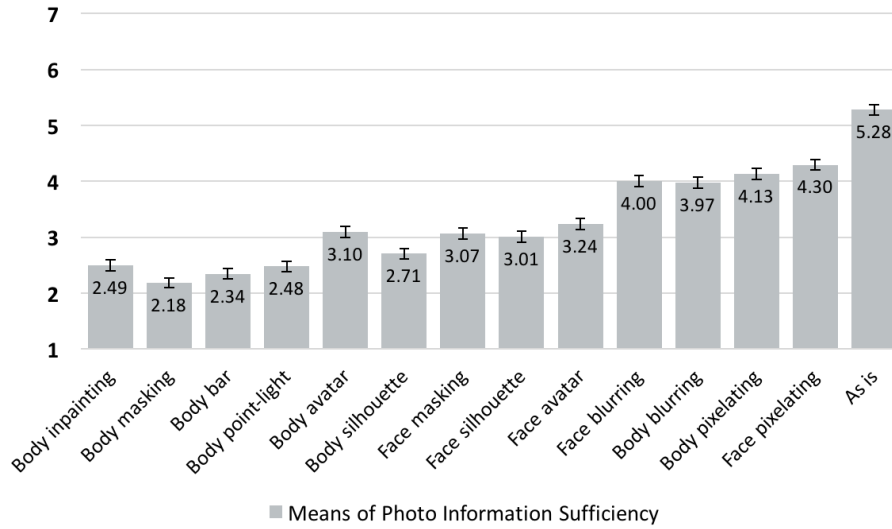


Figure 4.4: Information sufficiency rating ( $M$  and  $SE$ ). Obfuscations ordered from most to least effective.

#### 4.3.2.2 Photo Information Sufficiency

Do users believe that obscured photos still provide sufficient information? It seems that this also depends on the obfuscation method. Similar to photo satisfaction, from the linear mixed-effects model, the overall  $\chi^2$  on information sufficiency shows a variation among the 14 obfuscations,  $\chi^2(13) = 1555.11$ ,  $p < .0001$ , with more effective obfuscations generally provide less information (Figure 4.4). As expected, the information sufficiency of *as is* ( $M = 5.28$ ,  $SD = 1.54$ ) is higher than all other obfuscations (all  $d \geq 0.59$ , all  $p < .001$ ). Participants also give higher information sufficiency ratings to *face pixelating* ( $M = 4.30$ ,  $SD = 1.52$ ), *body pixelating* ( $M = 4.13$ ,  $SD = 1.59$ ), *body blurring* ( $M = 3.97$ ,  $SD = 1.62$ ), and *face blurring* ( $M = 4.00$ ,  $SD = 1.59$ ) compared to the remaining 9 obfuscation methods (all  $d \geq 0.41$ , all  $p < .01$ ), which means that *blurring* and *pixelating* preserve more information in photos. Among the more effective obfuscation methods, *body avatar* ( $M = 3.10$ ,  $SD = 1.65$ ) provides more information than *body inpainting* ( $M = 2.49$ ,  $SD = 1.64$ ,  $d = 0.33$ ,  $p < .001$ ), *body masking* ( $M = 2.18$ ,  $SD = 1.43$ ,  $d = 0.55$ ,  $p < .001$ ), *body bar* ( $M = 2.34$ ,  $SD = 1.55$ ,  $d = 0.45$ ,  $p < .001$ ), *body point-light* ( $M = 2.48$ ,  $SD = 1.56$ ,  $d = 0.38$ ,  $p < .001$ ), and *body silhouette* ( $M = 2.71$ ,  $SD = 1.52$ ,  $d = 0.24$ ,  $p = .01$ ).

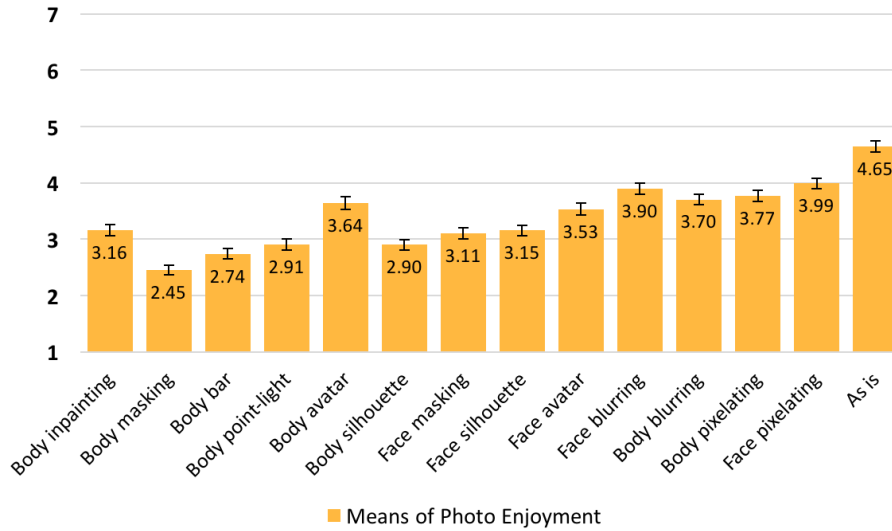


Figure 4.5: Enjoyment rating ( $M$  and  $SE$ ). Obfuscations ordered from most to least effective.

### 4.3.2.3 Photo Enjoyment

From the linear mixed-effects model of enjoyment, a similar pattern occurs where there is again a variation among the 14 conditions,  $\chi^2(13) = 795.09$ ,  $p < .0001$ , with that more effective obfuscations are less enjoyable (Figure 4.5 ). The mean enjoyment of *as is* photos ( $M = 4.65$ ,  $SD = 1.64$ ) is higher than all others (all  $d \geq 0.39$ , all  $p < .001$ ). Participants felt that photos with the *body avatar* obfuscation ( $M = 3.64$ ,  $SD = 1.81$ ) were about equally enjoyable with *body pixelating* ( $M = 3.77$ ,  $SD = 1.61$ ,  $d = 0.07$ ,  $p = .99$ ), *body blurring* ( $M = 3.70$ ,  $SD = 1.61$ ,  $d = 0.03$ ,  $p = 1.00$ ), and *face blurring* ( $M = 3.90$ ,  $SD = 1.58$ ,  $d = 0.15$ ,  $p = .38$ ), though they create the most enjoyable photos (aside from *as is* and *face pixelating*). In addition, as our most effective obfuscation method, *body inpainting* ( $M = 3.16$ ,  $SD = 1.70$ ) is more enjoyable than *body masking* ( $M = 2.45$ ,  $SD = 1.46$ ,  $d = 0.43$ ,  $p < .001$ ) and *body bar* ( $M = 2.74$ ,  $SD = 1.54$ ,  $d = 0.27$ ,  $p < .01$ ).

### 4.3.2.4 Social Presence

Do the obscured photos still provide a sense of human contact? From the results of the linear mixed-effects model, the overall  $\chi^2$  of social presence scores demonstrated significant variation among the 14 conditions,  $\chi^2(13) = 754.27$ ,  $p < .0001$ . The social presence score in the *as is* condition ( $M = 4.81$ ,  $SD = 1.69$ ) is higher than in all other obfuscation conditions (Figure 4.6 ) (all  $d \geq 0.30$ ,

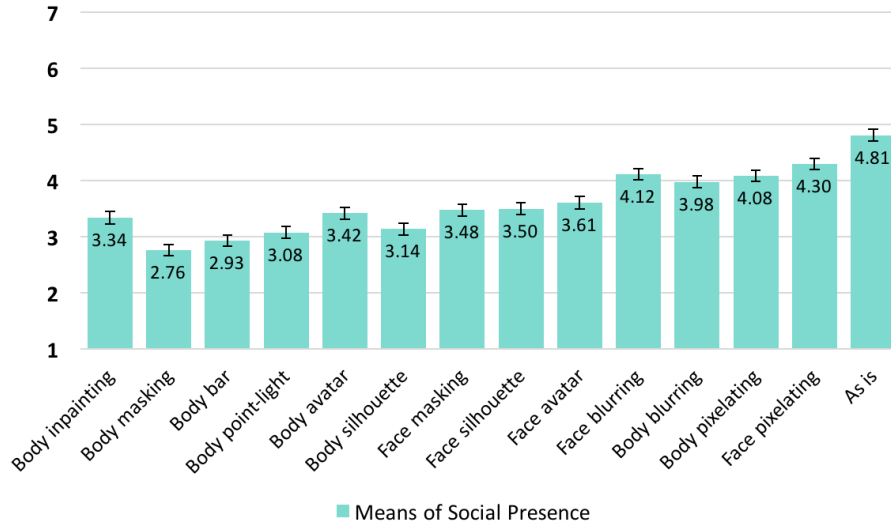


Figure 4.6: Social presence rating ( $M$  and  $SE$ ). Obfuscations ordered from most to least effective.

all  $p < .001$ ). Beyond *as is*, the scores of other obfuscations are less spread out between conditions than the other scores, with most social presence scores around 3 or 4. *Body masking* has the lowest social presence score ( $M = 2.76$ ,  $SD = 1.65$ ). Again, *body inpainting*(1) ( $M = 3.34$ ,  $SD = 1.85$ ) and *body avatar*(2) ( $M = 3.42$ ,  $SD = 1.75$ ) provide a better sense of human contact than *body masking* ( $M = 2.76$ ,  $SD = 1.65$ ,  $d_1 = 0.34$ ,  $p_1 < .001$ ,  $d_2 = 0.40$ ,  $p_2 < .01$ ) and *body bar* ( $M = 2.93$ ,  $SD = 1.64$ ,  $d_1 = 0.24$ ,  $p_1 < .001$ ,  $d_2 = 0.29$ ,  $p_2 < .001$ ). While not significant, their social presence ratings are slightly higher than *body point-light* ( $M = 3.08$ ,  $SD = 1.71$ ) and *body silhouette* ( $M = 3.14$ ,  $SD = 1.67$ ).

#### 4.3.2.5 Obfuscation Likability

Moving from participants' attitudes towards the photos to their attitudes towards the obfuscations themselves, we ask how much they like (or dislike) the obfuscations. From the results of the linear mixed-effects model, there is a variation among obfuscation conditions,  $\chi^2(13) = 963.46$ ,  $p < .0001$ , but There is no difference between *as is* ( $M = 4.76$ ,  $SD = 2.02$ ), *face pixelating* ( $M = 4.58$ ,  $SD = 1.74$ ), *body pixelating* ( $M = 4.31$ ,  $SD = 1.75$ ), *body blurring* ( $M = 4.52$ ,  $SD = 1.68$ ), and *face blurring* ( $M = 4.71$ ,  $SD = 1.70$ ) (all  $d \leq 0.19$ , all  $p > .05$ ). Generally, the rightmost five conditions in Figure 4.7 are similarly likable. Among the remaining nine obfuscation methods, participants like *body avatar* ( $M = 4.02$ ,  $SD = 2.08$ ), *face avatar* ( $M = 3.82$ ,  $SD = 1.99$ ), and *body*



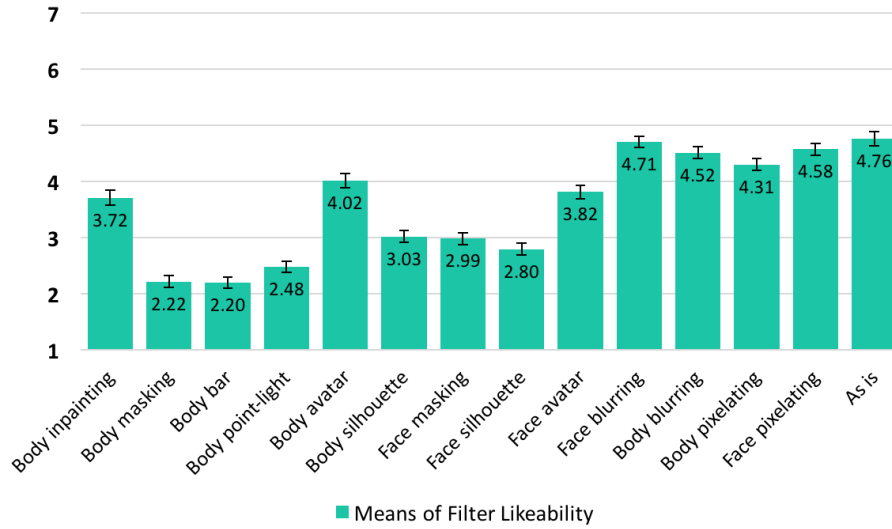


Figure 4.7: Obfuscation likability ( $M$  and  $SE$ ) from most to least effective.

*inpainting* ( $M = 3.72$ ,  $SD = 2.13$ ) more than the other six obfuscations (all  $d \geq 0.26$ , all  $p < .05$ ).

#### 4.3.2.6 Obfuscation Preference

At the end, we asked participants which obfuscation method they would most like to use to obfuscate their own online photos (Table 4.3 ). Participants reported they would most like to use *as is* (23%), *face blurring* (15%), *body avatar* (12%), *body inpainting* (11%), and *face avatar* (9%). In contrast, very few people chose *body bar* (1%), *body masking* (2%), *body point-light* (2%), *body silhouette* (2%), or *face masking* (2%), and there was only one participant who preferred *face silhouette* (resulting in a rounded percentage of zero). Leaving out ineffective *blurring* and *pixelating*, the preferences for *body avatar* and *inpainting* (around or larger than 10%) are about five times as high as for other obfuscations which are mostly below 2%. Asking participants how willing they would be to use their preferred obfuscation method, we found that they generally had a positive attitude, with all obfuscations scoring at or above 4 on a 7-point scale (Table 4.3 ). Aside from *as is* ( $M = 6.15$ ,  $SD = 1.21$ ), participants who preferred *body avatar* reported the highest willingness to use the obfuscation of their choice ( $M = 5.94$ ,  $SD = 1.03$ ). In the open-ended question, participants stated that *body avatar* “as least give some context to the photo if someone saw it online and looks kind of fun, ” “most pleasing to the eyes, cute, ” “protects someones privacy but it makes it lighthearted,” “the avatar keeps the person’s identity semi-private while not taking away from the

		General preference	Willingness to use	Preference given privacy concern
	As is	23%	6.15 (1.21)	1%
Face	Masking	2%	5.40 (0.89)	2%
	Silhouette	0%	4.00 (0.00)	0%
	Avatar	9%	5.80 (0.96)	17%
	Blurring	15%	5.60 (1.40)	26%
	Pixelating	7%	5.74 (0.73)	9%
Body	Inpainting	11%	5.29 (1.53)	15%
	Masking	2%	4.50 (2.74)	0%
	Bar	1%	5.25 (1.50)	4%
	Point-light	2%	5.00 (2.35)	1%
	Avatar	12%	5.94 (1.03)	16%
	Silhouette	2%	4.67 (0.52)	1%
	Blurring	5%	5.23 (1.30)	4%
	Pixelating	7%	5.75 (0.97)	4%

Table 4.3: Obfuscation preference, willingness to use, and preference given privacy concerns. Standard deviations appear in parentheses beside the means. Obfuscations are ordered from most to least effective.

composition of the photo with a line, block, or blur,” and “privacy does not have to be so bland, the avatar is creative.” For *inpainting*, they considered it “looks best to fully remove the person from the picture if it can be done in a way that isn’t fully obvious,” “it is like they are not there at all,” “just removes the person so that the photo isn’t ruined,” and “provides the true privacy.” All above results introduce that *avatar* and *inpainting* are practically more preferable and create a better user experience than other effective obfuscations.

#### 4.3.2.7 Would Privacy Obfuscations Change Privacy Behaviors?

As a follow up question, participants answered whether they had ever decided not to upload a photo to a SNS for privacy reasons. Fifty three percent of participants reported they had indeed done so. Over half (56%) of those who had declined to upload a photo for privacy reasons reported that they would upload the photo they previously declined to share, if having access to an obfuscation. We asked the 81 participants who had privacy reasons that prevented them from sharing a photo in the past but reported they would upload a photo using one of the obfuscations which which obfuscation they would choose. Twenty-six percent selected *face blurring*, 17% selected *face avatar*, 16% preferred *body avatar*, and 16% would like to use *body inpainting* (Table 4.3).

## 4.4 Discussion

Our overall goal is to increase the privacy options people have when sharing photos by discovering obfuscations that are both effective against re-identification and preferred/likable by users. First, we discuss the effectiveness of the obfuscations. We find that body obfuscations are generally more effective than face obfuscations (see the "body vs. face" part of the Results section), and there is no practical difference in user experience between face and body (for example, the likability difference between *face avatar* and *body avatar* is just 0.2, which means participants have almost the same attitude towards these two obfuscations). Hence in the following discussions, we only discuss body obfuscations which are relatively more effective and without user experience decreasing. Next, we discuss the user experience of the obfuscations. Finally, we integrate these, along with prior work on machine re-identification (vs. human re-identification) to generate recommendations about the most effective and likable obfuscations for photo privacy (Table 4.4).

### 4.4.1 Effectiveness: Face vs. Body

We found that body-obfuscations were more effective than face-obscuring obfuscations with a 9% success difference (see section *Body vs. Face*). From the practical perspective, obscuring the body is also supposed to be more effective against human recognition than obscuring the face because it can conceal more details such as clothes, gestures, gender, race, and height that may reveal a person's identity [5, 242]. Similarly, machines may be able to infer a person's identity from a photo with only the face obscured based on body information or the same clothes appearing in different photos over time [212]. In the case of SNSs, this effect would be exacerbated because we would expect people to primarily view familiar faces; *blurring* and *pixelating* are even less effective for familiar vs. unfamiliar faces [65]. Because they are more effective overall both in our work and in previous work, in the following discussion we only consider body obfuscations.

### 4.4.2 Moving Beyond Blurring and Pixelating

Although *blurring* and *pixelating* are commonly used both in research and in practice [24, 124, 286], our results suggest that they are two of the least effective obfuscation methods against human recognition (identification rates as high as 67%, Table 4.1; with above average confidence, Figure 4.2 and Table 4.2).

	<b>Prior Use for Privacy Protection</b>	<b>Preference</b>	<b>Effectiveness Against Human Recognition</b>	<b>Effectiveness Against Machine Recognition</b>
Inpainting	Less common. Used for photo [272] and video [149, 216, 305].	Preferred	Effective	Unknown, suspected highly effective
Masking	Less common. Used for photo [148] and video [149, 305].	Not preferred	Effective	Unknown, suspected highly effective
Bar	Rare. Used for video [305].	Not preferred	Effective	Unknown, suspected highly effective
Point-light	Rare. Used for video [49].	Not preferred	Effective	Unknown, suspected effective
Avatar	Rare. Used for photo [235] and video [216, 262].	Preferred	Somewhat effective	Unknown, suspected effective
Silhouette	Less common. Used for photo [216, 305] and video [149].	Not preferred	Somewhat effective	Unknown, suspected effective
Blurring	Common. Used for photo [24, 98, 124, 177] and video [301]	Less-preferred	Ineffective	Ineffective [161, 195]
Pixelating	Common. Used for photo [65, 158, 286] and video [33, 146].	Less-preferred	Ineffective	Ineffective [195]
As is	N/A	Preferred	Ineffective	Ineffective [220]

Table 4.4: Summary of photo obfuscation methods (body-obfuscations only because they are more effective; see “Effectiveness: Face-obscuring vs. Body-obscuring.”) Effectiveness is defined by the difference in the identification success percentage of *as is* and each body obfuscation (see Table 4.1). The misidentification of *as is* is 23% (100% minus 77%). An obfuscation that achieves at least twice of *as is* misidentification (46%) is defined as “Somewhat effective”, so the identification success should be no more than 54%. An obfuscation that achieves at least three times of *as is* misidentification (69%) is considered “Effective”, so the identification success should be at most 31%. Obfuscations are ordered from most to least effective.

Consistent with prior work on *blurring* and *pixelating* obfuscations against human recognition [27, 158], participants in our study were able to identify humans who were blurred and pixelated in photos. This may be because these obfuscations fail to hide body shape, skin and hair color. Color cues are important in face recognition. The effect of color becomes more evident when shape cues are degraded [299]. As we mentioned in section “Controlling Content”, these features may also allow the obscured photo to be re-identified by machines. For example, generative adversarial networks (GAN) [161], and artificial neural networks [195] worked well to identify blurred and pixelated faces [195].

On the other hand, *inpainting*, *masking*, *bar*, *point-light*, and *avatar* are much more effective in obfuscating the target in each photo (Table 4.1). Participants are also less confident about their ability to identify people who are de-identified using these obfuscations (Figure 4.2 and Table 4.2), regardless of whether their identification is correct or incorrect. In other words, these obfuscations are effective and viewers feel less confident in their ability to recognize targets when viewing them. This finding is consistent with prior work about the relationship between activity visibility and perception confidence that the less visible an activity is, the perceptions are more likely to derive from participants’ own experiences, thus lower accuracy and confidence they have [20].

Perhaps surprisingly, participants were only somewhat unconfident (with mean ratings around three), rather than very unconfident about their ability to recognize targets in effective obfuscations. Partly, this may be due to the effect of our experimental interface. Participants were forced to make a choice even when they did not know which target was present. Once they made their choice, cognitive dissonance may have led them to report they were “somewhat” rather than “very” unconfident in that choice. Alternatively, or additionally, Americans are more likely to choose a Likert option that indicates positive emotion [164]. Participants may have chosen the most positive choice (somewhat unconfident) among the options on the unconfident side of the scale. These speculations do not take away from the key finding: participants were less confident in their recognition of effective obfuscations.

*Inpainting*, which removes all visual clues about the person in a photo, is the most effective obfuscation, yielding a mere 19% identification success, which is notably, less than chance (25%). When target is present, this rate decreases to 7% which is much lower than chance, indicating, as expected, participants were unable to identify a target in this condition. When the target is absent most participants chose “None of above” because the target is completely removed, resulting in more

correct rejections (67%). In a sense, this is an ideal scenario: rather than trying to guess the target’s identity, it is better for viewers to simply assume that there is no one there at all.

### 4.4.3 User Experience

The upward slopes in Figures 4.3 to 4.6 and 4.7 demonstrate that overall, there is a trade-off between obfuscation effectiveness and user experience: obfuscations with a higher effectiveness have a lower user experience in terms of satisfaction, information sufficiency, enjoyment, social presence, and likability. The scatter plot in Figure 4.8 , which plots likability against identification success, also demonstrates this trend.

*Blurring* and *pixelating* are subtle; they preserve many visual features of an image such as the colors and shapes [299]. Because of the subtlety, people may not notice the affected region at first glance. While this subtlety may contribute to relatively high levels of satisfaction, it also likely results in their relative ease of recognition. Conversely, *masking*, *bar*, and *point-light* are more effective, but they are less satisfying, give insufficient information, are less enjoyable, and lack a sense of social presence. This is also reflected by the preference percentages (only around 2%) and qualitative feedback of these three obfuscations. For example, participants thought dots or lines damaged the photo aesthetics. However, there might be solutions. Another study that I contributed to uncovers that viewers’ satisfaction can be restored by adding beautification filters to enhance the aesthetics of an obscured photo [110]. As we will discuss in the following section, *inpainting* and *avatar* are exceptions that are both effective and provide a relatively good user experience.

### 4.4.4 Better Options: Inpainting and Avatar

Although effective obfuscations are generally less satisfying (Figure 4.8 ), *inpainting* and *avatar* are outliers to this trend. They are very effective, and, as compared to other effective obfuscations, have high levels of satisfaction, information sufficiency, enjoyment, social presence and likability.

We could argue that the user experience ratings of *inpainting* should be similar to the *as is* condition, because *inpainting* does not add unrelated content (e.g. a large gray box) to the image. Ostensibly, with the target completely removed and the area filled in by existing photo content, the only difference is the number of people in the group photo. However, after removing the target,

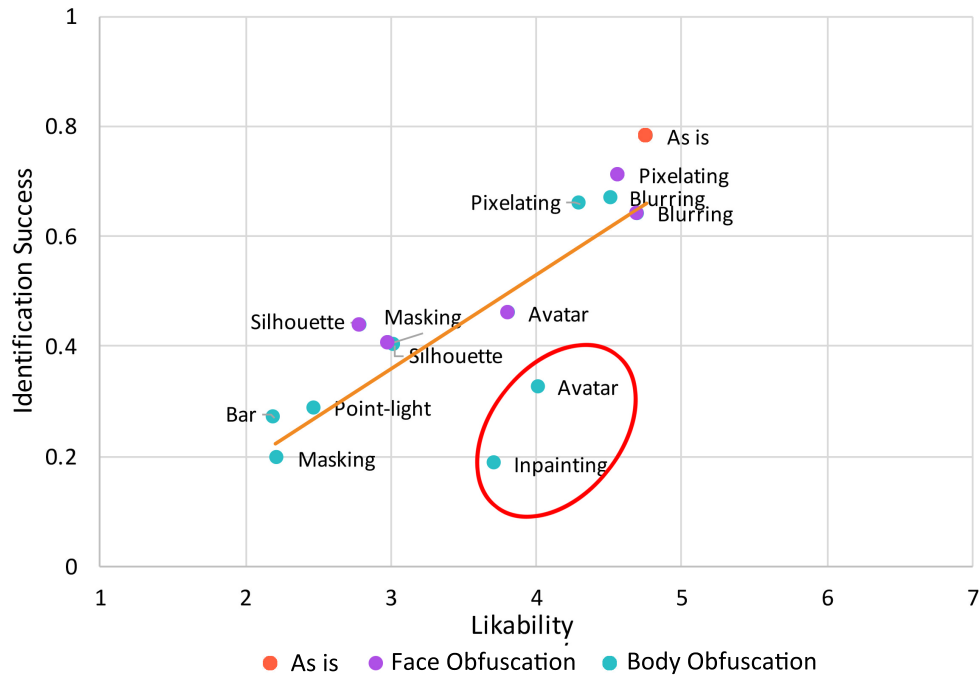


Figure 4.8: Scatterplot of Likability (X axis) against Identification Success (Y axis). This plot shows the general trade-off between effectiveness and user experience. However, body avatar and body inpainting are outliers. They are both effective and provide a good user experience.

an unnatural seeming space were left, damaging the composition [167]. Furthermore, participants probably knew, based on in-study experience, that there likely used to be a person in the gap they saw in the photo. The awkwardness of this gap may have reduced the user experience.

Future versions of *inpainting* could improve the user experience by using more sophisticated image re-construction techniques to recompose the picture after removing the target [58, 104]. Using these techniques, we can imagine a system that could first extract the people in the photo, identify and fill gaps in the background, and finally put the people back to achieve an optimal composition, but without the target. Indeed, we see promising commercial applications that have elements of this functionality already present such as Photoshop’s “content aware patch” and Snapseed’s “Expand [276] ” feature. While currently these advanced editing options are not suited to the task (e.g., sometimes these generate unnatural double images), we see promise for techniques such as these to create *inpainting* obfuscation options that provide a better user experience.

Unlike *inpainting*, *avatar* does add content to an image, so we might not expect the user experience ratings of *avatar* to be similar to *as is*. However, the user experience ratings of the *avatar*

obfuscation were higher than all other effective obfuscation methods, and therefore should be further explored. One noteworthy thing about *avatar* is that the photo enjoyment rating of *body avatar* is slightly higher than *face avatar*, indicating that people may prefer to see a complete cartoon character rather than just a cartoon face. However, except for Bitmoji [30], existing approaches to avatar creation mainly focus on generating avatars that look somewhat like the target, but with an exclusive focus on the face (e.g., facial component matching [235]). Future avatar-based approaches for photo obfuscation may benefit from the development of methods to create full-body avatars instead of face avatars as this may improve the user experience.

Our results show that *avatar* and *inpainting* also provide a greater sense of human contact than other effective obfuscations. For *avatar*, the cartoon human character preserves the target’s facial expression and gesture. Viewers may feel this feature allows them to perceive human contact both among the people within the photo, and between themselves and the people in the photo [22]. For *inpainting*, since the target is totally removed, people may be less likely to be jarred by the visual indicator of the lack of presence of a person in the photo. People tend to select the medium that they perceive to have the highest human contact, hence enough human contact in a photo or on SNSs would increase their participation and encourage them continue using the medium [94].

Participants qualitative input was consistent with the quantitative results about users’ experience. For example, *inpainting* and *avatar* are more likable and preferable than other highly effective obfuscations. In open response format, participants mentioned that they liked *inpainting* because “it is the most thorough privacy technique,” “it seems you were never there in the first place and there is no way to identify you,” “it is the best way to make the photo visually appealing,” and “provides true privacy.” For *avatar*, they stated it is “cute and fun,” “catches the eyes,” “still shows the person in a positive light,” and “inserts personality, you can customize it.” Participants’ comments revealed that they preferred *face blurring* for many of the reasons they liked *inpainting* and *avatar*, but were unaware of the ineffectiveness of *blurring*. Participants reported that they thought *blurring* “adequately hides identity while still giving information about the original photo and person’s attitude,” “the body without a clear face doesn’t tell much,” and “preserves the integrity of the picture while providing some form of privacy.” This implies that, perhaps because of its widespread use, people are unaware that *blurring* is ineffective against both human (Table 4.1 and Table 4.4) and machine identification [161, 195]. One clear implication is that if users are provided with obfuscation options, they should be clearly informed about the benefits and drawbacks of each



(e.g., *blurring* is ineffective as a privacy-enhancement).

Finally, *inpainting*, *avatar* and *face blurring* were the most commonly selected obfuscations people would want to use to obscure a photo they had previously declined to upload to a SNS for privacy reasons. Participants reported they would be willing to upload that photo, if they had an obfuscation available as a solution to their privacy dilemma. **These results indicate that obfuscation methods, especially *inpainting* and *avatar*, have the potential to dramatically increase the privacy options available to people who want to share photos on SNSs.** In SNS scenario, users can *inpaint* themselves in a photo which their friend uploads, that generates a re-constructed photo that is closest to the original one. Hence, the photo uploader will not feel much sharing loss, and viewers will not even be aware. On the other hand, though both the uploader and viewers will be aware of the *avatar* obfuscation, it brings positive emotions, as our participants stated: “cute” and “fun.” It is similar to other frequently used applications which add cartoon figure or emoji in a photo. *Avatar* makes photos more interesting and protects privacy. Both obfuscations reduce privacy conflicts in photo sharing on SNSs.

Though in this study, we only applied *avatar* and *inpainting* on human in photos, but they can also be useful to protect other sensitive content we identified in the first study, for example, beer can, pet, and vehicle license plate.

#### 4.4.5 Obfuscation Timing: at capture, upload or share?

Besides the obfuscation methods, referring back to the behavioral privacy model [50], in the Background section, we have introduced that privacy protection can happen at different phrases, hence obfuscation timing is important to consider when developing photo privacy control mechanisms. Obfuscations can be applied at various stages of photo processing: at the time of capture, on the device, at the time of upload, or at the time of sharing. Applying an obfuscation at each stage has different privacy benefits and addresses different concerns [151]. Applying at the time of sharing means that a SNS, for example, would gain access to a raw photo (including identifiable information) but “friends” may not see the raw photo because of the obfuscation, which addresses concerns about social threats [118, 151]. On the other hand, we could apply an obfuscation at the time of photo capture, such that only an obfuscated image is captured on a device (e.g., phone). In this example, an image would be obfuscated before upload, and a SNS would never gain access to the raw photo, which addresses organizational threats related to SNS providers and various third

parties [151].

We already see somewhat related commercially available examples of the first situation, where “filters” are applied at the sharing stage (e.g., Facebook and Snapchat photo filters [112, 258]). However, these obfuscations are not designed to obfuscate, nor are they likely effective as a privacy-enhancement, though this warrants future investigation. From our qualitative data, we know that at least some participants have privacy concerns about the privacy of their photos as shared with platforms (e.g., a SNS like Facebook). One participant raised his/her concern about this by saying: “*Facebook identifies you first, then blurs you, but Facebook already tracks you.*” People fear that their information will be collected, stored, sold, and reuse by SNS providers or other third parties [151].

One possible solution, therefore, is to apply privacy obfuscations before uploading images to such platforms. However, identifying and obfuscating sensitive parts in an image is computationally resource-intensive, for example, increasing CPU usage [150]. A remarkable degradation in efficiency in case of devices is observed with the increase in the number of people in the photos being processed. Thus, accomplishing privacy obfuscation on a device (vs. offloading to the cloud, for example) will require a re-thinking of desired device capabilities. Should new cameras be designed with computer vision capabilities on board, as suggested by [50]? How can automated redaction be done in a more efficient (from a memory, power, etc. perspective) on the device rather than on the cloud so that private information does not need to leave the device thus putting it at greater risk for leaking, hacking, etc.?

## 4.5 Limitations

First, other information, besides a visual representation of the face and body of a person can be revealing. The risk of contextual cues in identification is particularly acute in OSNs because of their social nature. For example, in an OSN un-obscured mutual friends, the background [263], any text or personal belongings, the comments under the photo [124], and the time and location [283] may lead to identification of an obscured person. The approaches we have identified here may be similarly effective and satisfying when applied to these contextual photo elements. However, this is certainly worth investigating in the future.

Second, participants were recruited via mTurk. Though Turkers are relatively more demo-

graphically diverse than the sample collected by other means [39], recruiting this way has drawbacks. For example, MTurk only allows studies to be conducted online and has a unique nature of labor [193]. Future work should replicate this study with non-mTurk participants.

Third, the user experience rating of *masking* is the worst compared to all other obfuscations. Besides the box covering content in the photo beyond just the target, the gray color used may also be an issue. We decided to use gray for the box to ensure a consistent experience and to prevent any gender indication, (e.g., a pink box could indicate the obscured target is a female [133]). However, gray may evoke negative emotions compared to warmer and brighter colors and result in less perceived human contact both within the photo and between the viewer and the photo [190].

Third, a research question around computer vision emerge from this work: What automated interpolation techniques are most effective at creating effective, acceptable and seamless inpainting? For example, reconstructing 3D models of world landmarks using collections on photo sharing sites may allow recovery of obfuscated parts of photos [58]. Future studies need to further investigate how platforms and devices might implement the user experience enabling these obfuscations.

Finally, we only applied obfuscations to people that viewers were not familiar with. However, on SNSs, viewers are likely to know the people in a photo, which limits the applicability of obfuscations. Hence, investigating obfuscations which are robust in de-identifying both familiar and unfamiliar people is important. I have addressed this limitation in the next chapter.

## 4.6 Chapter Conclusion

In Study One, we identified sensitive content in photos, hence in this study, we aimed to investigate effective and usable obfuscations that can be applied on sensitive content. The results show that the two most commonly studied and used obfuscations, *blurring* and *pixelating*, are not effective at preventing humans (or, drawing from related work, machines) from recognizing the content of a photo. Thus, the most commonly used privacy obfuscations do not provide privacy protection. We then introduce novel obfuscations that *are* effective at preventing humans from recognizing content in an image. Of the highly effective obfuscations we introduce, we then analyze these from the perspective of user experience finding that *inpainting*, which totally removes the content from the photo, and *body avatar*, which replaces the content with an avatar, outperform other obfuscations. We suggest that *body inpainting* and *body avatar* show promise as photo privacy-

enhancing technologies because they are effective from a human recognition perspective and provide a good user experience.

## Chapter 5

# Study 3: De-identifying Familiar and Unfamiliar People

Note: This work was rejected from CHI 2018 and USENIX 2018.

### 5.1 Introduction

From the last chapter – Study Two, we know that some photo obfuscations are effective at de-identification by humans and machines from the last chapter-Study Two. However, one serious limitation of this work is that we only applied obfuscations on people who are unfamiliar to viewers. This seriously limits the obfuscations’ application in SNSs context. Most privacy conflicts in SNSs are related to familiar people identification [25]. Viewers who are familiar with people depicted in a photo can identify them more easily when they are not obfuscated [36]. Some obfuscations may be robust across familiar and unfamiliar people, whereas others may not be, as shown in Figure 5.1 where it is trivial to identify former US president Barack Obama. To bridge this gap, in this study, we investigated the de-identification effectiveness and users’ experience of obfuscations which were applied on both familiar and unfamiliar people in photos. We selected *blurring*, *silhouette*, *avatar*, *masking*, *inpainting* obfuscations from our last study, and included *morphing* [127]. We predicted *morphing* to have high effectiveness and to make for a better user experience, because it could obscure a photo seamlessly (for detailed introduction of these obfuscations, please see Table 5.1).



Figure 5.1: An example of a blurred familiar person

In general, the goal of this study is to investigate the effectiveness against human recognition and the user experience (by measuring satisfaction, information sufficiency, enjoyment, social presence, likability and preference) among six obfuscation methods applied on both familiar and unfamiliar people. Overall, we would like to uncover which obfuscations are effective and can at the same time provide a good user experience across familiar and unfamiliar cases.

We found that most of these obfuscations (e.g. *inpainting*, *masking*, *avatar*, and *morphing*) were effective and that there was no major difference between identifying familiar and unfamiliar people across all obfuscations, except for *blurring*, which is much less effective in familiar cases. Furthermore, we found that familiarity did not influence photo satisfaction, information sufficiency, and social presence. *Inpainting*, *avatar*, and *morphing* provide a good user experience. In terms of likability and user preference, *morphing* is less preferable, though. In brief, *inpainting* and *avatar* have a solidly high effectiveness and a good user experience across familiar and unfamiliar cases.

To summarize, the primary contribution of this work is identifying obfuscations that are 1) robust in the increased likelihood of recognition associated with familiarity and 2) provide a good viewer experience.

The remainder of this chapter is structured as follows: First, we summarize previous work on withholding behavior, two approaches to protect privacy, and specify the importance of familiar people de-identification. Next, we introduce our method and use statistical analysis to study obfuscation's effectiveness, viewer experience, and our obfuscations' impact on photo withholding. We







Example	Name & Definition	Related Work	Example	Name & Definition	Related Work
	<b>Blurring.</b> Reduces image detail by generating a weighted average of each pixel and its surrounding pixels.	[24, 301, 98, 149, 158]		<b>Morphing.</b> Merging two people's bodies to create an average representation.	[76, 147, 127]
	<b>Silhouette.</b> Replaces content with a monochrome visual object that mirrors the extracted shape of the original content.	[49, 149, 216, 305]		<b>Avatar.</b> Replaces content with a graphical representation that preserves some elements of the underlying content. For example, a human avatar can preserve facial expression and gesture, but hide biometrically unique elements (e.g., face) of identity.	[216, 235, 262]
	<b>Masking.</b> Replaces content with a monochrome solid box that covers the content to be protected and surrounding image content.	[148, 149, 305]		<b>Inpainting.</b> Completely removes content fills in the missing part of the image in a visually consistent manner.	[149, 216, 272, 305]

Table 5.1: Six obfuscation methods. In the example figures, we applied the methods on familiar people. They were also applied on unfamiliar people in the study, yielding in 14 conditions (We added *as is* as the baseline condition).

subsequently discuss the results of our analysis. Finally, we also state our limitations in this study.

## 5.2 Method

To answer these questions, we conducted an experiment with 230 participants to investigate the effectiveness and user experience of privacy-enhancing obfuscations.

## 5.2.1 Participants

We recruited 285 participants located in United States through the Amazon Mechanical Turk crowd-sourcing service. Participants were paid \$2.00 to complete the study which took about 30 minutes based on a suggested payment on MTurk [239]. We set restrictions to ensure high data quality: MTurk workers must have a good reputation (above 95% approval rate) with more than 1000 HITs approved [223]. After collection, we excluded the data from participants who failed more than one attention check question. Additionally, to ensure the majority of participants knew the targets in familiar cases, we also excluded the data from participants who failed to identify three or more famous people in familiarity questions, resulting in a final sample size of 230.

One hundred and twenty-two reported being male, 107 being female, and one person preferred not to reveal the gender. Participants' ages ranged from 18 to 55+ years, with 13% age 18-24, 47% age 25-34, 23% age 35-44, 10% age 45-54, and 7% age 55+. Seventy-two percent was White. This sample is representative of the US population [281]; in general, recruiting via mTurk results in a more diverse sample than other recruitment means [39]. Ninety-nine percent of participants reported using Internet most of the day or several times a day; and 73% used SNSs most of the day or several times a day.

## 5.2.2 Experimental Design

Our experiment was a within-subject design, with seven privacy-enhancing obfuscation conditions by two familiarity levels (familiar vs. unfamiliar). The seven obfuscations were *as is* (obfuscation-free), *blurring*, *morphing*, *silhouette*, *avatar*, *masking*, and *inpainting*.

### 5.2.2.1 Obfuscation Methods

We adopted six obfuscations from prior work on online photo privacy and video surveillance [180, 124, 127, 216, 305] (listed in Table 5.1). We chose the most effective and user-friendly ones, for example, *inpainting* and *avatar*, and the commonly adopted *blurring*. We did not include *pixelating*, *bar*, and *point-light*, because they are neither effective nor user-friendly from Study Two. We are particularly interested in *morphing* [127], which merges another person with the target person in the photo to create a general representation. We expect *morphing* to have high effectiveness and good user experience, because it is the most seamless obfuscation among our six methods.



## 5.2.3 Stimuli

### 5.2.3.1 Targets

The *target* is the person in a photo to be identified. In total, we had 14 targets (seven familiar people and seven unfamiliar people). To create unfamiliar targets, we took photos of seven people who were unknown to our participants. We included models from racial categories broadly representing the racial makeup of the United States, including white, African American, Asian, and Hispanic and Latino. For familiar targets, we chose seven people who were shown to be familiar to participants from [96, 294], for example, Barack Obama (Table 2.2). Next, we located images of each of these people that allowed for noncommercial reuse and modification and downloaded these for use in the study.

Though famous people may not be a perfectly representative of friends, they are a common proxy in significant prior work that investigates recognition of familiar people, and these work indicates that the cognitive process of recognizing famous and familiar people are similar (e.g., [37, 81, 244, 245]). One reason for our choice and this choice in prior work is because participants have different levels of familiarity with people in their friend circle whereas the familiarity of famous persons is more consistent across participants. Using famous persons is thus likely to result in a more consistent recognition rate. For the purposes of experimental control, we traded off some amount of external validity. Furthermore, we explicitly tested participants' familiarity with the celebrities we used as stimuli. As shown in the results of familiarity questions (Table 5.2), with the exception of Taylor Swift, all of the famous people we used as familiar targets achieved a named identification rate of 90% or higher. This means that at least 90% of participants were able to correctly produce (recall rather than recognize) the name of the famous person. The subjective familiarity means are five or above, indicating a relatively high level of self-reported familiarity with all familiar targets.

### 5.2.3.2 Photo Creation

To create photo backgrounds and background people, we again chose online photos that allowed reuse and modification and photos taken by our researchers. We used Photoshop to reassemble them with the targets. To be consistent, each photo has one target (obfuscation condition applied), three background people, and a similar background scenery (park, campus etc.) (Figure. 5.2). We employed each of the seven obfuscations to one familiar target and one unfamiliar target, each with

Name	%Named	Mean Familiarity (SD)
Jennifer Aniston	95%	5.55 (1.50)
Angelina Jolie	93%	5.67 (1.52)
Taylor Swift	83%	4.97 (1.92)
Oprah Winfrey	99%	6.08 (1.10)
Barack Obama	100%	6.55 (0.71)
Brad Pitt	96%	5.90 (1.36)
Leonardo DiCaprio	98%	6.16 (0.97)

Table 5.2: Participants’ familiarity with the famous people in our stimuli. The three columns show the famous people’s names, percentage of being named, and means of familiarity with standard deviations.

two different backgrounds resulting in 196 unique photos (7 familiar targets \* 7 obfuscations \* 2 backgrounds + 7 unfamiliar targets \* 7 obfuscations \* 2 backgrounds).

### 5.2.3.3 ID Photo

We gathered an ID photo of each target, and three ID photos of people who look similar to the target, for example, with same gender, same race, similar hair color, and body figure (e.g. the second and fourth person in the choices in Figure. 5.2). The ID photos of famous targets are also all famous people’s photos, for example, Jennifer Aniston vs. Nicole Kidman.

## 5.2.4 Measurements

We measured obfuscation effectiveness via identification rate and confidence, and user experience through existing validated Likert scales.

### 5.2.4.1 Obfuscation Effectiveness

- **Identification Rate.** We asked “Please identify the person indicated by the orange arrow.” with four choices (three ID photos and “None of above”).
- **Identification Confidence.** Afterwards, participants were shown the question “How confident do you feel that you correctly identified the person?” with a scale from 1 ‘Completely unconfident’ to 7 ‘Completely confident’ [230].

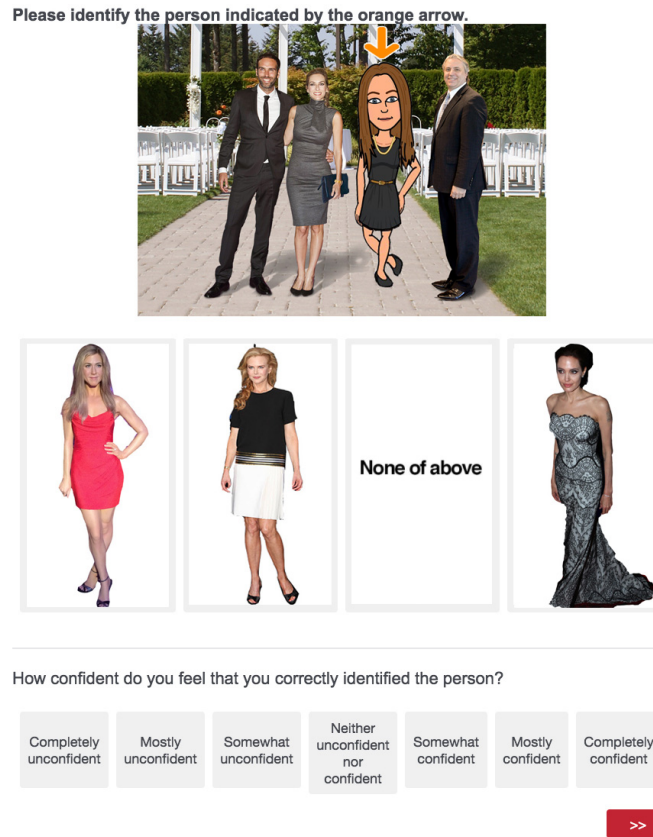


Figure 5.2: Experiment interface with one stimuli and ID photo examples

#### 5.2.4.2 Users' Experience

We measured the obfuscation's user experience from four perspectives. All the responses used 7-point Likert scale from 1 'Strongly disagree' to 7 'Strongly agree.'

- Photo Satisfaction.** We adapted the satisfaction item "The photo is satisfying" from the validated image appeal scale [59]. Image appeal, validated across cultures, is the extent to which images are perceived as "appropriate and aligned to user expectations, satisfying, or interesting" [59].
- Perceived Photo Information Sufficiency.** Various privacy filters may cause different levels of information sufficiency. We selected a single item "The photo provides sufficient information" from the photo information quality scale [72]. This scale measures "the satisfaction of users who directly interact with the computer for a specific application". Our selected item

loads onto the “content” factor and was correlated to the item “is the system successful?” [72]

- **Photo Enjoyment.** We measured perceived photo enjoyment using the single-item photo enjoyment scale [233].
- **Perceived Social Presence.** We adapted an item “There was a sense of human contact when I saw the photo” from perceived social presence scale which measures the feelings of intimacy and warm [154]. Human contact here means both contact between the viewers and people in the photo, and people’s interaction within the photo.
- **Obfuscation Likability.** We measured likability of each obfuscation using the item “I like the \_\_\_\_ obfuscation” which derived from the interface preference scale [203].
- **Obfuscation Preference.** Participants reported their preference for each obfuscation with the question “If you could use any of the obfuscations for photos you post on SNSs, which one, if any, would you like to use?”, followed by an open-ended question about the reason, and 7-point Likert scale item measuring their willingness to use this obfuscation. Next, they answered “Have you ever declined to upload a photo to an online social network for privacy reasons?” If “Yes,” they were asked given the access to one obfuscation, if they were willing to share again. If “Yes,” we asked which obfuscation would have moved them to upload the photo, and their reasons for doing so.

### 5.2.5 Procedure

Prior to the study, we conducted pilot tests with our lab members to check for bugs, gather data about the length of the study and ensure that the data collection worked well. In the actual test, first, participants accessed the experiment website Qualtrics through the link in our MTurk HIT. After providing consent, they answered six demographic questions and two social network usage questions. Next, we asked participants to test the browser size and resize their browser to make sure all participants viewed stimuli in a similar visual environment. Afterwards, they saw an overview of the seven obfuscation examples along with the description of each obfuscation.

Next, we trained participants about the experimental task. During training, participants learned about the tasks they would perform and completed two training trials. To help participants gain confidence about the identification task, next they completed two pre-trials. The photos used

	% identified Total	% identified Familiar	% identified Unfamiliar	OR (familiar vs. unfamiliar)	p-value	OR (vs. As is, regardless of familiarity)	p-value	% $\Delta$
Masking	21%	25%	17%	1.62	<.05*	0.05	<.001***	-75%
Inpainting	23%	22%	24%	0.89	0.59	0.06	<.001***	-72%
Avatar	33%	27%	39%	0.58	<.01**	0.10	<.001***	-60%
Silhouette	33%	30%	35%	0.80	0.25	0.10	<.001***	-61%
Morphing	36%	34%	37%	0.88	0.60	0.12	<.001***	-57%
Blurring	72%	78%	67%	1.75	<.01**	0.53	<.001***	-13%
As is	83%	90%	77%	2.69	<.001***	NA	NA	NA

Table 5.3: Identification rate in all cases (including both target present and absent), odds ratio and p-value between familiar and unfamiliar cases for each obfuscation, and odds ratio and p-value between each obfuscation and the baseline *as is* regardless of familiarity. The obfuscations are ordered by identification rate of total cases (familiarity + unfamiliar) from lowest (most effective) to highest (least effective).

in the pre-trials were obfuscation-free, making the target easily identifiable.

Participants then completed 14 trials where they saw photos with semi-randomly assigned obfuscation conditions and targets, and identified the target person (Figure. 5.2). Participants saw all 14 conditions and 14 targets during the experiment. No conditions or targets were repeated. For example, in the first trial, if the photo contains condition *blurring* (on familiar target) and Jolie, then the rest of the photos that include either *blurring* (on familiar target) or Jolie would be excluded from the subsequent trials. In most cases the target was among the four choices offered, but there was a 21% chance that the target was NOT present among the choices. Afterward identifying the target, participants rated their confidence and experience.

After finishing all trials, participants were shown each of the seven obfuscation conditions individually, and rated each likability. Then they answered several obfuscation preference questions. Next they saw the seven famous people’s photos used in the familiar trials, and were asked to write down each famous person’s name and rate their familiarity with each person [261]. Finally, participants responded to a set of privacy attitude questions. After completing all tasks, a random code was generated. Participants copied this code to MTurk to receive remuneration.

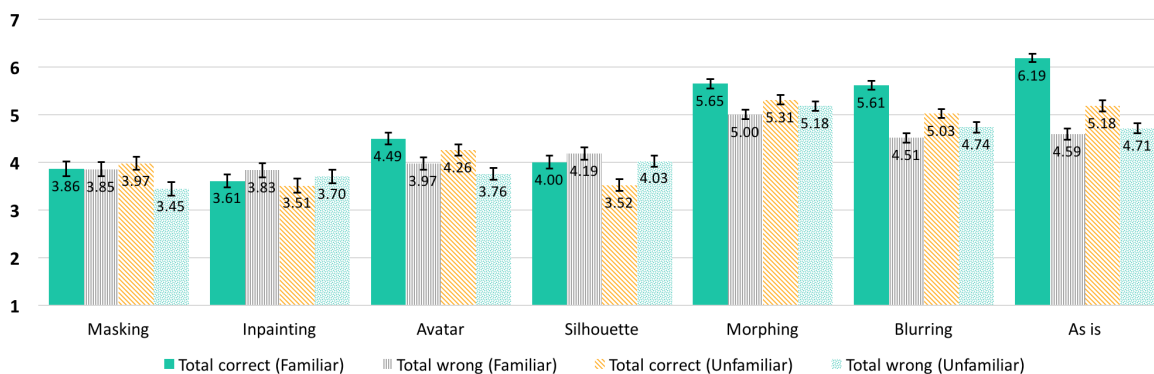


Figure 5.3: Means and standard errors of identification confidence of Total Correct and Total Wrong separated by familiarity.

## 5.3 Results

### 5.3.1 Obfuscation Effectiveness

Obfuscation effectiveness is measured through identification rate, which is the percentage of trials that participants correctly identified a target in question. Identification confidence is a self-reported measure of how confident a participant was in his or her identification. In this work, because we are interested in privacy, higher identification rate means lower obfuscation and privacy effectiveness. For example, if one obfuscation has a 40% identification rate, that means its effectiveness as a privacy-enhancing obfuscation is 60% (100% - 40%).

#### 5.3.1.1 Identification Rate

Using a Tukey post-hoc test on a logistic mixed-effects model of all cases, we found that regardless of familiarity, *as is* (90% and 77%) and *blurring* (79% and 67%) have higher identification rates compared to any other obfuscations (all  $p < .001$ ), which indicates *blurring* is ineffective against human recognition. All other obfuscations perform well, with identification rates ranging from 17% to 39% (Table 5.3).

*Effect of Familiarity.* If we look across all obfuscations, familiarity does not have an effect on identification rate,  $\chi^2(1) = 1.01$ ,  $p = .31$ . However, breaking it down into individual obfuscations, participants were able to identify familiar targets obfuscated by *as is*, *blurring* and *masking* more easily than unfamiliar targets (all  $p < .05$ ; differences of 8–13%). For *inpainting*, *silhouette*, and *morphing*, there was no major difference between familiar and unfamiliar cases (all  $p \geq .25$ ).

Oppositely, a familiar person was less likely to be identified than an unfamiliar person when the *avatar* obfuscation was applied, with a 12% difference ( $p < .01$ ). Note that *inpainting*, *silhouette* and *morphing* are effective across both familiar and unfamiliar cases.

### 5.3.1.2 Identification Confidence

Using a Tukey post-hoc test on a linear mixed-effects model of all cases we found that regardless of familiarity, people are more confident with their identifications for three visually less distorted obfuscation methods, *as is* ( $M = 5.56$ ,  $SD = 1.64$ ), *blurring* ( $M = 5.15$ ,  $SD = 1.52$ ), and *morphing* ( $M = 5.22$ ,  $SD = 1.51$ ), than the remaining four obfuscations (all  $d \geq 0.52$ , all  $p < .001$ ), with all scores above four. When identifying the targets in photos using *inpainting* ( $M = 3.71$ ,  $SD = 2.20$ ), *masking* ( $M = 3.69$ ,  $SD = 2.19$ ), *avatar* ( $M = 4.03$ ,  $SD = 1.92$ ) and *silhouette* ( $M = 3.99$ ,  $SD = 1.92$ ), their confidence decreases (Figure. 5.3).

*Effect of Familiarity.* Familiarity has an overall effect on the identification confidence ( $\chi^2(1) = 48.10$ ,  $p < .0001$ ): people are more confident about their identifications when they view familiar people’s photos. For *as is* and *blurring*, when viewing photos of familiar people, people feel more confident than viewing photos of unfamiliar people (both  $d \geq 0.23$ , both  $p < .01$ ), especially when they are correct.

Since people can hardly see anything related to the target’s identity in *inpainting* and *masking*, the confidence ratings are similar regardless of the familiarity or the identification correctness (all between three to four), which indicates people are generally not confident when viewing photos with these two obfuscations.

## 5.3.2 Users’ Experience

Knowing that *inpainting*, *masking*, *silhouette* and *morphing* are effective, how do our participants feel about these obfuscations? We created four linear mixed-effect models to analyze users’ experience of the obfuscations, then conducted Tukey post-hoc tests on the pairs we were interested in, for example, comparisons between *morphing* and other obfuscations.

### 5.3.2.1 General Effects of Obfuscation and Target Familiarity

As shown in Figure 5.4 to 5.7, the results of four linear mixed-effect models on four scales show significant variations among the seven obfuscation methods: with  $\chi^2(6) = 697.01$ ,  $p < .0001$

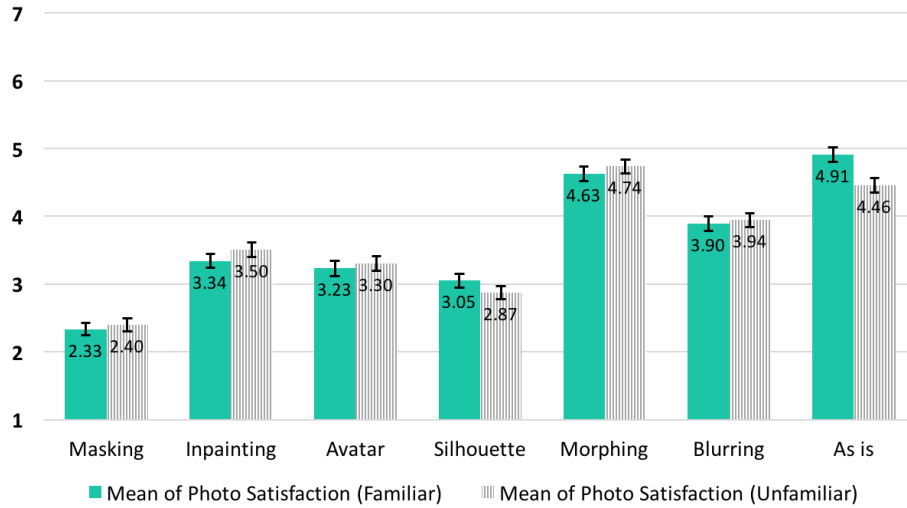


Figure 5.4: Photo satisfaction rating ( $M$  and  $SE$ ). Obfuscations ordered from most to least effective.

for satisfaction;  $\chi^2(6) = 924.43$ ,  $p < .0001$  for information sufficiency;  $\chi^2(6) = 588.59$ ,  $p < .0001$  for enjoyment; and  $\chi^2(6) = 549.25$ ,  $p < .0001$  for social presence.

Familiarity does not have an overall effect on satisfaction, enjoyment, and social presence. Information sufficiency is an exception with  $\chi^2(1) = 4.58$ ,  $p < .05$ , and obscured photos with unfamiliar people have lower information sufficiency compared to those with familiar people. Only in *as is*, the satisfaction, information sufficiency, and enjoyment of photos with familiar people are higher than those with unfamiliar people (all  $d \geq 0.25$ , all  $p < .001$ ) (Figure 5.4 to 5.6).

### 5.3.2.2 Effective Obfuscations That Provide A Good User Experience

One obfuscation, in particular, stood out as being both effective and providing a good user experience. *Morphing* has the highest ratings compared to other obfuscations across all four measurements (all  $d \geq 0.27$ , all  $p < .001$ ). There is no difference between *morphing* and *as is* on satisfaction, information sufficiency, enjoyment and social presence in unfamiliar cases (all  $p > .05$ ), and no difference on satisfaction and social presence in familiar cases (both  $p > .05$ ), indicating that overall *morphing* a person makes the photo as satisfying, enjoyable, and provides the similar amount of information and human contact as the original (*as is*) photo.

Although the user experience scores are not as high as *morphing*, *inpainting* (the most effective obfuscation) has a relatively good performance in satisfaction, information sufficiency, enjoyment and social presence, compared to next-most effective obfuscation *masking* (all  $d \geq 0.44$ , all



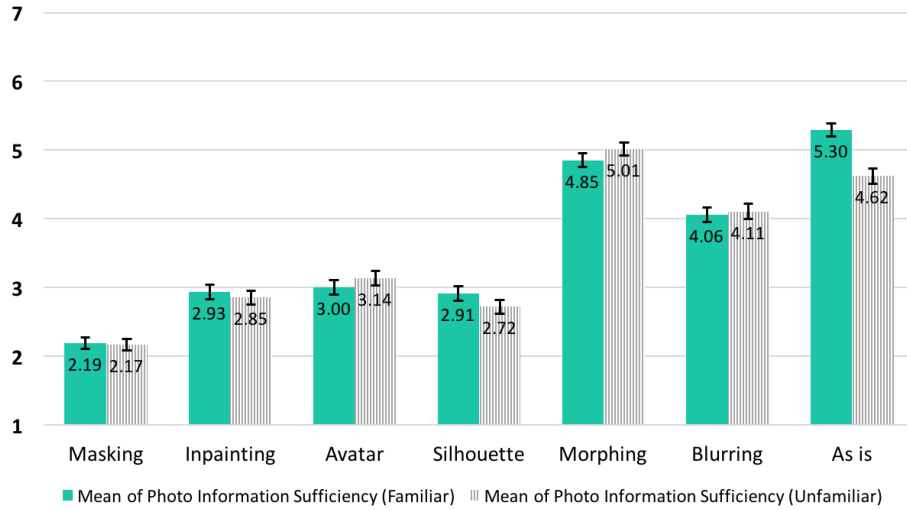


Figure 5.5: Information sufficiency rating ( $M$  and  $SE$ ). Obfuscations ordered from most to least effective.

$p < .001$ ).

Among the four most effective obfuscations in Figure 5.4 to 5.7, *avatar* performs better than *masking* from all perspectives (all  $d \geq 0.29$ , all  $p < .001$ ). The ratings are also higher than *silhouette*, though some differences are not statistically significant, for example, in information sufficiency and social presence measurements.

### 5.3.3 Obfuscation Likability

The results of four linear mixed-effect models on obfuscation likability show a variation among the seven conditions,  $\chi^2(6) = 371.99$ ,  $p < .0001$  (Figure 5.8). Consistent with the other four user experience measurements, within effective obfuscations, users like *inpainting* (1) ( $M = 4.11$ ,  $SD = 2.18$ ) and *avatar* (2) ( $M = 3.86$ ,  $SD = 2.00$ ) more than *masking* ( $M = 2.16$ ,  $SD = 1.56$ ,  $d_1 = 0.80$ ,  $p_1 < .0001$ ,  $d_2 = 0.72$ ,  $p_2 < .0001$ ) and *silhouette* ( $M = 3.17$ ,  $SD = 1.78$ ,  $d_1 = 0.37$ ,  $p_1 < .0001$ ,  $d_2 = 0.29$ ,  $p_2 < .0001$ ). Note that the likability of *inpainting* is about as high as *blurring* ( $M = 4.54$ ,  $SD = 1.73$ ,  $p = .15$ ). However, the likability of *morphing* ( $M = 3.00$ ,  $SD = 1.82$ ) is much lower than *as is* ( $M = 5.14$ ,  $SD = 1.88$ ,  $d = 0.84$ ,  $p < .001$ ), even though these two have similar satisfaction, information sufficiency, enjoyment and social presence. *Morphing* is also less likable when compared to *inpainting*(1) and *avatar*(2) ( $d_1 = 0.40$ ,  $p_1 < .001$ ,  $d_2 = 0.34$ ,  $p_2 < .001$ ).

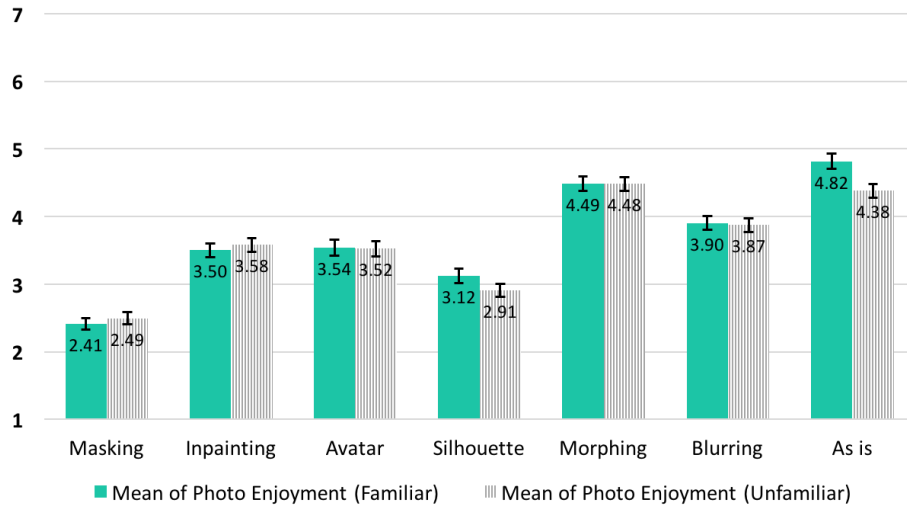


Figure 5.6: Enjoyment rating ( $M$  and  $SE$ ). Obfuscations ordered from most to least effective.

	General preference	Willingness to use	Preference given privacy concern
Masking	3%	5.00 (0.82)	0%
Inpainting	20%	5.73 (1.23)	33%
Avatar	18%	5.87 (1.24)	28%
Silhouette	3%	5.71 (0.95)	3%
Morphing	7%	5.38 (1.41)	6%
Blurring	26%	5.55 (1.23)	28%
As is	22%	6.04 (1.32)	1%

Table 5.4: Obfuscation preference, willingness to use, and preference given privacy concerns. Standard deviations appear in parentheses beside the means. Obfuscations are ordered from most to least effective.

### 5.3.4 Obfuscation Preference

After rating the likability of each obfuscation, participants responded which obfuscation they would like to use on their photos posted on a SNS. As shown in the second column of Table 5.4, the majority of participants chose *blurring* (26%), *inpainting* (20%), and *avatar* (18%). Twenty-two percent of them preferred not to use obfuscation. In accordance with the likability scores, only a few participants preferred *masking* (3%), *silhouette* (3%), and *morphing* (7%). Participants were generally willing to use the obfuscations they selected, with the means all above five. Besides *as is*, they reported the highest willingness to applying *avatar* on their photos ( $M = 5.87$ ,  $SD = 1.24$ ).

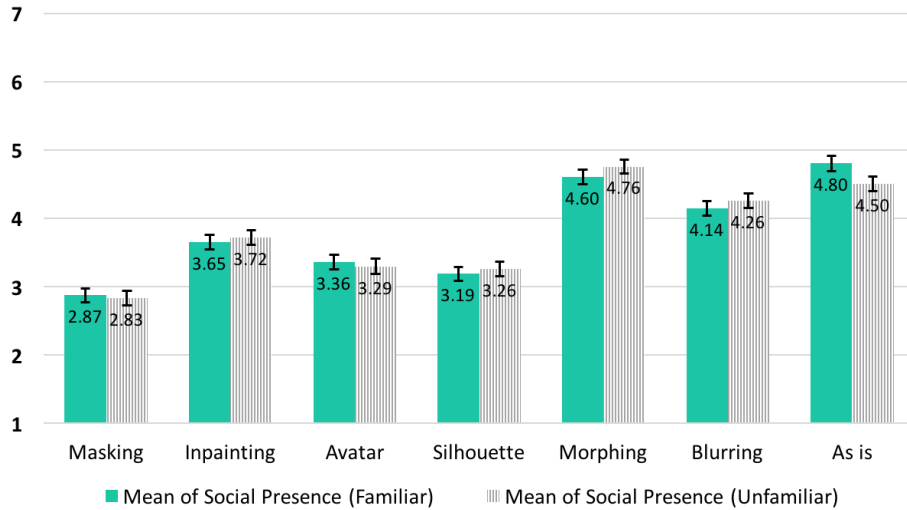


Figure 5.7: Social presence rating ( $M$  and  $SE$ ). Obfuscations ordered from most to least effective.

## 5.4 Discussion

### 5.4.1 It is Easier to Identify Familiar People

Neuropsychological studies show that humans process familiar and unfamiliar faces differently [131]. People rely more on so-called “internal features” (e.g. facial features) rather than “external features” (e.g. body or head contour) when identifying familiar people. On the other hand, people rely on both internal and external features equally when identifying unfamiliar people [81]. In our study, we found that for both obfuscation-free photos (i.e., *as is*) and photos with ineffective obfuscations (e.g. *blurring*, where both external and internal features remain visible to some extent), participants were more easily and more confidently able to identify people who were familiar to them. This replicates prior work on obfuscation-free photos (e.g., [36]), and extends this prior work to obfuscated photos. On the other hand, *silhouette* and *avatar* have lower identification rate in familiar cases than in unfamiliar cases. This is because the four choices (ID photos) gave participants a hint about whether the obscured target was familiar or unfamiliar. When identifying a familiar target, they were inclined to search for precise internal features, however *avatar* and *silhouette* provide few or no internal features.

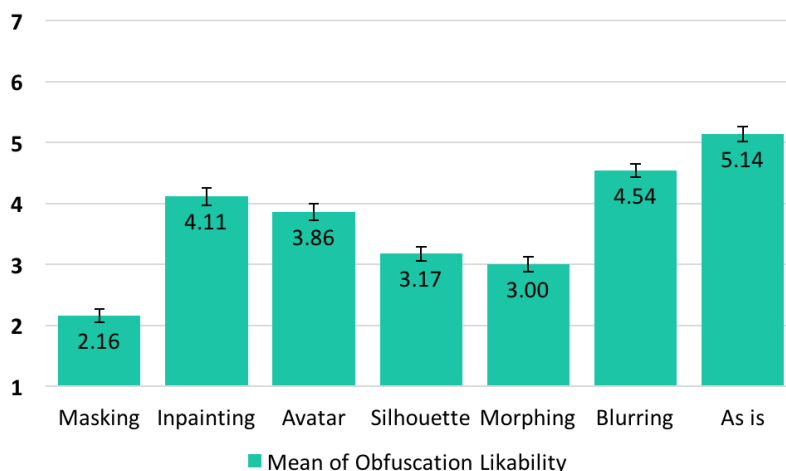


Figure 5.8: Obfuscation likability ( $M$  and  $SE$ ) from most to least effective.

### 5.4.2 Obfuscating Familiar People in Online Photos

To summarize, when there are familiar people in online photos, *inpainting*, *avatar*, *silhouette*, and *morphing* maintain their effectiveness against human recognition. For these effective obfuscations, high familiarity with the people in photos decreases neither the effectiveness nor the user experience, suggesting that these obfuscations are robust and can be applied to any online photos including photos in SNSs. Second, our results confirm that *blurring*, though used commonly both in prior research and in practice, is not effective against identification of unfamiliar people, and performs even worse when a viewer is familiar with the obfuscated person. In the following sections, when discussing *inpainting* and *morphing*, we do not consider familiarity, because familiarity does not affect their effectiveness and user experience.

### 5.4.3 Effective and Likable: Inpainting and Avatar

Aside from *morphing*, *inpainting* achieves the best balance between effectiveness and user experience compared to the other three effective obfuscations (*masking*, *avatar*, and *silhouette*, see Figure 5.9). *Inpainting*'s performance in likability is even better (see Figure 5.10). *Inpainting* has a low identification rate across familiar (22%) and unfamiliar cases (24%) (Table 5.3). This low rate includes an uncharacteristically high number of correct rejections in “target absent” cases (84% for familiar and 76% for unfamiliar cases) where we assume that, despite no evidence, participants were tempted to choose “None of above” due to the literal absence of a target. If instead we only



Figure 5.9: Scatterplot of Satisfaction (X axis) against Identification Rate (Y axis) which shows the general trade-off between effectiveness and satisfaction. *Morphing*, *inpainting* and *avatar* are below the regression line, which means they are relatively effective and satisfying.

focus on target present cases, the identification rate of *inpainting* goes down well below chance to 3% (familiar) and 7% (unfamiliar). Moreover, people feel unconfident about their identifications when viewing inpainted photos. *Inpainting* removes all possible clues of the target that may lead to identification as our participants pointed out in responses to open ended questions. They said, “it’s the perfect camouflage ‘invisible ink’ option,” and it “completely leaves out any traces.”

The user experience with *inpainting* is not as high as *as is*, *blurring* and *morphing*, but *inpainting* is more satisfying, enjoyable, and provides more information and human contact than the similarly effective obfuscation *masking*. It also has equivalent user experience with *avatar*. Excluding *as is*, *inpainting* has the second highest likability score, and 20% of participants prefer it to all other obfuscations. After we reminded participants about their privacy concerns when uploading photos, 33% reported they would rather use *inpainting* than any other obfuscation.

We might expect the user experience of *inpainting* to approach *as is*, if the obfuscation was indeed so good that viewers might not realize that a target was removed from the photo. However, because of the space left between the adjacent background people in our stimuli, viewers understand that a person was removed from the photo. Unlike many photos shared online, in our stimuli target and background people were not engaged in interactive gestures; when they are, the missing target may be even more obvious than in our stimuli. For example, imagine a photo with

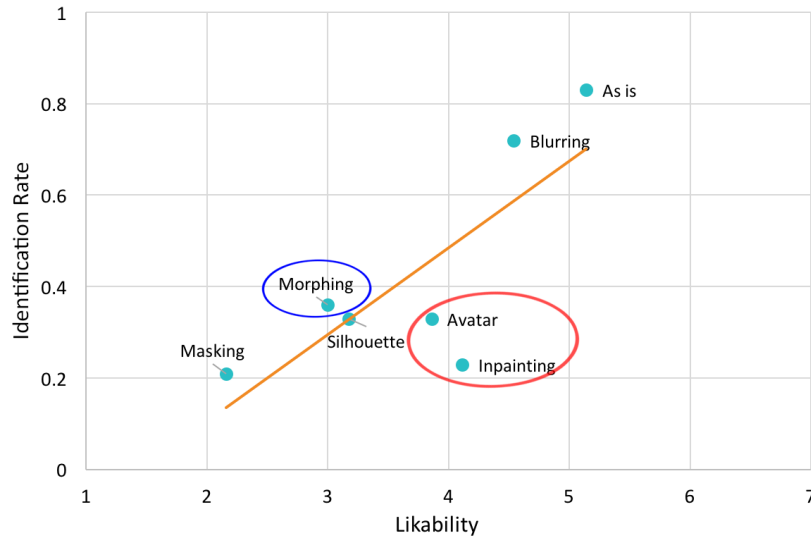


Figure 5.10: Scatterplot of Likability (X axis) against Identification Rate (Y axis) which shows the trade-off between effectiveness and obfuscation likability. *Inpainting* and *avatar* are below the regression line, and *morphing* is effective but not likable.

a person’s floating arm which is supposed to be on the target’s shoulder (Figure. 5.11). In this case, the missing target would be obvious and perhaps disturbing to viewers. In addition, the unnatural space damages the photo composition [167]. Fortunately, emerging image reconstruction techniques can solve this issue and further improve user experience [58, 104]. For example, using these techniques, a future photo obfuscation mechanism could first identify all people in a photo, distinguish each person from the background, remove the target, patch the missing background, modify the remaining people’s gestures if necessary, then reconstruct a seamless photo. There are some existing commercial applications that may facilitate this process to improve *inpainting*’s user experience, such as Photoshop’s “content Aware Patch [4]” and Snapseed’s “Expand [276].”

The second obfuscation that is a relatively good balance between effectiveness and user experience is *avatar*. From the user experience ratings, likability, and users’ preference, there is almost no difference between *avatar* and *inpainting*, which both score highly on all these measures. *Avatar* hides the identity, but preserves the facial expression and gesture of the target. It combines a common online social behavior—adding cartoon stickers on their photos on social media to enhance the aesthetic, emotions, and communications [117]—with their privacy protection goals. Participants mentioned that *avatar* “seems fun and different,” “looks cute,” is “the most visually appealing,” and is “hilarious and doesn’t ruin the photos.”



Figure 5.11: Example of awkward *inpainting*.

One thing that cannot be ignored is *avatar*'s surprisingly lower effectiveness in familiar cases compared to unfamiliar cases (27% vs. 39% identification). As we discussed in the second subsection of this Discussion, people identify familiar targets mostly using internal features, while in unfamiliar cases, they rely on both internal and external features [81]. *Avatar* provides some general clues (long/short hair, skin tone etc.), but no detailed internal features such as exact eye- and nose-shapes. Hence when the obscured person is familiar, people fail to acquire the detailed internal features that they need to make an accurate identification. In addition, when given three famous people as choices, participants arguably retrieved person-identity information from their long-term memory rather than comparing the avatar details with each choice carefully [38, 44], thus resulting in incorrect choices.

In most usage scenarios, this high identification rate in unfamiliar cases should not be a problem. First, when viewers try to identify an unfamiliar person's avatar in a photo outside of our study, they are not provided several choices to compare with this avatar. Moreover, the identification rate is likely related to the level of detail in the avatar. Reducing the details will likely increase the effectiveness, for example, changing the hair color of the avatar from brown to blond, or even switching the gender of the avatar. The confidence scores also show that people are not very confident in identification when the *avatar* obfuscation is used. In short, both *inpainting* and *avatar* are good options to protect photo privacy with a good user experience.

#### 5.4.4 A Promising Obfuscation: Morphing

For both familiar and unfamiliar cases, *morphing* is effective against human recognition. It also has the potentially desirable characteristic that it looks un-obscured; when participants see it in trials, they are likely unaware that the photo is *morphed* and may think it is *as is*. There are no obvious visual occlusions like in *masking*, or indicators of a person missing like the space left in *inpainting*. We speculate this characteristic is what we see reflected by its high satisfaction, information sufficiency, enjoyment, and social presence ratings, which are almost as high as *as is*. This indicates that *morphing* has the potential to be among the most preferable obfuscations (see Figure 5.9).

However, when participants are made explicitly aware that the image they are viewing has been *morphed*, and subsequently asked whether they would like to apply *morphing* on their own photos, we found that *morphing* was not at all preferable. It had a much lower likability score and preference percentage than *inpainting* and *avatar* (see Figure 5.10). These contradictory findings suggests that people may be skeptical about the concept of *morphing*, or may be unwilling to “blend” themselves or friends with other people, which is required for *morphing*. One participant mentioned *morphing* makes the photo look “ridiculous.” However, other participants saw real promise in the idea, saying, “Morphing is better for maintaining an anonymous ‘persona’ online. I think it’ll be easier to use morphing to show others that a person ‘exists’ behind my profile, but without revealing my real identity.” Another participant said that *morphing* was the obfuscation that “best keeps the integrity of the photograph while removing identity in a seamless way,” indicating the participant saw aesthetic value in this obfuscation.

Another possible reason for the low likability of *morphing* may be that we used a single person from the three “known person” choice options, rather than either an unknown person or an aggregate face, comprised of many people, to merge with the target. This may have enabled participants to identify both potential images that were morphed together. Consequently, people might have thought they would be easier to identify using *morphing* themselves. The high confidence scores indicate that participants were confident with their identifications of *morphed* photos. If their goal was to be obfuscated, they might assume others could also identify them if they used *morphing*.

Another intriguing possibility for the low likability could be something akin to the uncanny valley effect [202]. The images generated by *morphing* look somewhat like the target and somewhat



like a non-target, which could result in an eerie sense of the image being almost, but not exactly a known person. This possibility deserves additional attention. If, for example, the uncanny valley effect [202] is the reason for the low likability scores, we may be able to improve *morphing* by creating an “average” person by merging hundreds of people to create a more general target. This could push the new image further out along the similarity axis so that it is less similar to any one person, and thus out of the uncanny valley. Moreover, people tend to judge average faces as attractive [160], thus a *morph* that uses averages may make the person in the photo be viewed as more beautiful.

## 5.5 Limitations and Future Work

First, as we admitted in the method section, celebrities may not be a perfect representative of familiar people let alone friends. Moreover, in a SNS environment people’s recognition ability may be different under different levels of familiarity, which may affect the performance of the obfuscations, for example, there may be a difference in obfuscating close friends versus persons one has met only once. Future studies should test the effectiveness of obfuscations under different familiarity conditions.

Second, morphed photos were rated by participants to be satisfying, enjoyable, and able to provide sufficient information and human contact, but when applied on their own photos, they reportedly did not like *morphing*. We provided some possible explanations for this difference in the Discussion section, but we did not ask participants directly to describe their dislikes. Now that we know that *morphing* is promising, we will probe more deeply into this difference to gather feedback about how we might enhance the likability of *morphing*. Next, the intensity of *blurring* affects obfuscation effectiveness against both human and machine recognition [106, 158]. We used the blurring intensity level from [158]. Changing the intensity of blurring could increase identification rate and/or lead to a lower user experience. Last, there are two aspects of our social presence measurement: the human contact between viewers and people in a image and the human contact within the people in a image. In future research, we would like to distinguish them.

## 5.6 Chapter Conclusion

Controlling content using obfuscation may be an effective strategy for enhancing privacy while maintaining or increasing the potential audience. Of these obfuscations, *inpainting* and *avatar* are robust to the increased likelihood of recognition associated with familiarity and provide a good viewer experience. Another effective obfuscation, *morphing* produced complicated and contradictory results; it has the highest user experience scores when viewers are unaware of its use, but it is not likable or preferable when viewers know the resulting image is a combination of two images. We suggest that *inpainting* and *avatar* are useful tools for photo privacy enhancement. With a solid understanding of obfuscations, we need to explore whether they could combat photo self-censorship, help reduce people's privacy concerns, and encourage photo sharing. The following two chapters answer this question.

## Chapter 6

# Study 4: Obfuscation May Combat Self-Censorship

Note: This work was rejected from CHI 2020.

### 6.1 Introduction

We know that when people feel that certain content in their photos is risky, they are likely to turn to the self-censorship strategy, which refers to the act of withholding information from others without formal obstacles, is a sociopsychological phenomenon that may cause negative effects on both the personal level (personal distress) and the collective level (decreasing the free information flow) [18]. As a boundary regulation strategy, self-censorship is prevalent on SNSs [60]. SNSs contain few visual cues about the audience [48], so SNS users imagine an audience to share corresponding content. However, they may censor content if they feel the imagined audience is not appropriate [191]. For example, people censor content that may harm their self-representation towards certain social groups, content may offend others, or bore their audience [255]. Existing work on online self-censorship investigates the common topics that people censor (e.g., entertainment, politics, personal opinion [255]), and much work specifically focuses on the political self-censorship phenomenon (e.g., [18, 99]), but there have been few studies that investigate specific types of content under censoring.

In this work, we particularly explore photo self-censorship. Visual data in SNSs carries more information than textual data, hence photo sharing plays an important role in online social communication. While enjoying the benefits of photo sharing, users also suffer from privacy leakage. Though most SNSs allow users to control their audience (e.g., share with only friends) [84], users are likely to censor photos as a safer option [191], in turn reducing the communicative ability of SNSs. Very few prior studies quantify the prevalence of online self-censorship. One work did a large-scale exploratory analysis and found out that post and comment censorship is common (71% of the participants have self-censored at least one post or comment) [60], but photo self-censorship, as a likely target for self-censorship, is still under-examined. Additionally, from the first three studies, we understand obfuscations can be usable and help prevent humans from recognizing sensitive content in photos. Hence, in this study, we would like to investigate whether obfuscation can combat photo self-censorship.

In this study, we aim to achieve three goals introduced below. We (1) quantified the prevalence of online photo self-censorship with 230 SNS users. Furthermore, because from previous chapters, we know that photo obfuscation can be an effective tool to protect online photo privacy, we also (2) interrogated whether privacy-preserving obfuscations such as blurring, might be useful for combating photo self-censorship. We know that there are gender and age differences in post/comment censorship, so we would like to know (3) if photo self-censorship had similar age or gender difference patterns compared to prior work [60], and if other factors such as privacy preference affected photo self-censorship.

First, we found that over half of the participants have self-censored photos due to privacy concerns on SNSs, which indicates that photo self-censorship is indeed prevalent. Among the participants who had censored photos, half of them would like to share that photo they had previously self-censored if they could apply obfuscations on it, which suggests that photo obfuscations may be useful to combat photo self-censorship. Next, we observed that people with higher privacy consciousness were more likely to censor photos. Last, we also found that women were more willing to share obfuscated photos about which they previously had privacy concerns. Additionally, people with higher privacy consciousness about their personal information were also inclined to share a previously censored photo after obscuring it, which implies that obfuscations could potentially address their concerns on photos containing sensitive information. Through this work, we gain a better understanding of photo self-censorship, and suggest that in addition to SNSs' current recipient control

approach, researchers can adopt obfuscations that may combat photo self-censorship.

## 6.2 Method

We conducted an online survey and elicited data from 230 participants to investigate 1) the prevalence of photo self-censorship on SNSs, 2) whether photo obfuscations can be a potential solution to reduce photo self-censorship, and 3) what factors may be associated with photo self-censorship and the intent to share photos with obfuscations.

### 6.2.1 Participants

We recruited 285 participants located in the United States through the Amazon Mechanical Turk crowd-sourcing service. Participants were paid \$2.00 to complete the study based on a suggested payment on MTurk [239]. We set recruiting restrictions to ensure high data quality: MTurk workers must have a good reputation (above 95% approval rate) with more than 1000 HITs approved [223]. We inserted three attention check questions in the Qualtrics survey. After collection, we excluded the data from participants who failed more than one attention check question and the final sample size was 230.

One hundred and twenty-two reported being male, 107 being female, and one person preferred not to reveal their gender. Participants' ages ranged from 18 to 55+ years, with 13% age 18-24, 47% age 25-34, 23% age 35-44, 10% age 45-54, and 7% age 55+. Seventy-two percent were White. This sample is demographically representative of the US population [281]. Ninety-nine percent of participants reported using the Internet several times or most most of the day; and 73% used SNSs several times or most of the day.

We also collected three aspects of participants' privacy preferences or consciousness: the information receiver, the information content, and the privacy preference in social contexts. Similar to the findings from the Pew Research Center [186], 97% of the participants considered "being in control of who can get information about you" as important, from which 86% consider it very important. Eight-one percent of the participants rated "controlling what information is collected about you" as important or very important. Ninety percent considered not revealing highly personal information on SNSs as important.

## 6.2.2 Obfuscation Methods

We selected six obfuscations from prior work on online photo or video surveillance privacy (Table 6.1).







Example	Name & Definition	Related Work	Example	Name & Definition	Related Work
	<b>Blurring.</b> Reduces image detail by generating a weighted average of each pixel and its surrounding pixels.	[24, 180, 301, 149, 158, 178]		<b>Morphing.</b> Merging two people's bodies to create an average representation.	[76, 147, 127]
	<b>Silhouette.</b> Replaces content with a monochrome visual object that mirrors the extracted shape of the original content.	[49, 180, 149, 216, 305]		<b>Avatar.</b> Replaces content with a graphical representation that preserves some elements of the underlying content. For example, a human avatar can preserve facial expression and gesture, but hide biometrically unique elements (e.g., face) of identity.	[180, 216, 235, 262]
	<b>Masking.</b> Replaces content with a monochrome solid box that covers the content to be protected and surrounding image content.	[180, 148, 149, 179, 305]		<b>Inpainting.</b> Completely removes content and fills in the missing part of the image in a visually consistent manner.	[180, 149, 216, 272, 305]

Table 6.1: Six obfuscation methods. Study participants were shown images, but not provided name, definition, or related work citations.

## 6.2.3 Variables

The demographic variables in our models include: 1) gender, 2) age, 3) Internet usage frequency (from ‘most of the day’ to ‘never’), 4) SNS usage frequency (from ‘most of the day’ to ‘never’), and 5) privacy preference/consciousness about personal information (four-point Likert scale). The two binary outcome variables are 1) if the user has declined to upload a photo to an online social network for privacy reasons, and 2) if the user is willing to upload the photo which he/she has previously refused to share if they are able to obscure the sensitive portion.

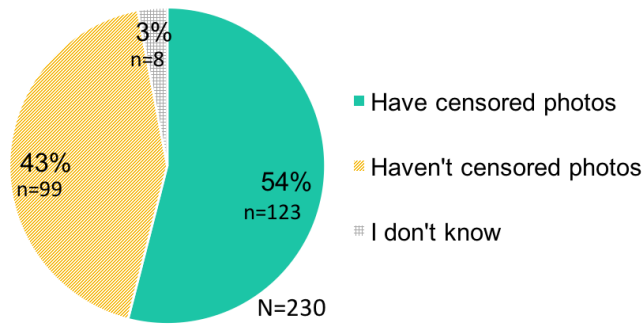


Figure 6.1: The percentage of people who have censored photos and who have not.

## 6.2.4 Procedure

First, after providing consent, participants answered six demographic questions and two social network usage frequency questions. Next, they were shown six example photos with six types of obfuscations in a randomized order. The example photos were composed of one target person to be obscured, three background people, and a campus scene. We included each obfuscation’s introduction on each example photo. To study the prevalence of photo self-censorship, we then asked participants “Have you ever declined to upload a photo to an online social network for privacy reasons?” There were three options: “Yes,” “No,” and “I don’t know.” Next, to uncover if obfuscations may be useful for reducing photo self-censorship, we asked participants who answered “Yes” in the last question: “In the last question, you said you had declined to upload a photo to an online social network for privacy reasons. If you had access to one of the privacy filters here, would you be willing to upload this photo using one of the filters?” Again, they responded “Yes,” “No,” and “I don’t know.” Afterwards, they answered three questions to measure their privacy preference or privacy consciousness. After completing all tasks, a random code was generated. Participants copied this code to MTurk to receive remuneration.

## 6.3 Results

### 6.3.1 Prevalence of Photo Self-Censorship

Just over half of participants (54%) reported they had self-censored photos on SNSs. Forty-three percent reported they had never declined to upload a photo due to privacy concerns, and 3%

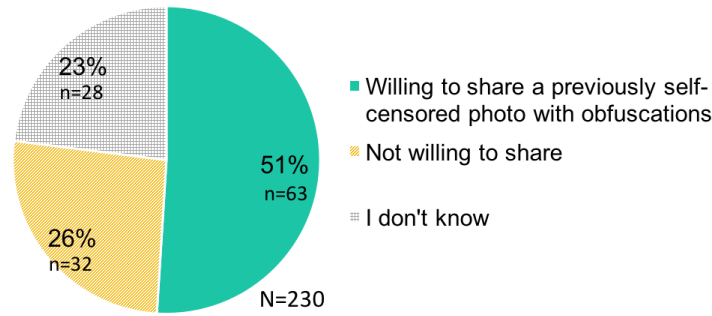


Figure 6.2: The percentage of people who are willing to share photos which they previously had privacy concern using obfuscations and who are still withholding even with obfuscations.

	Coefficient (SE)	Odds Ratio	OR 95% CI
Intercept	-4.16 (1.81)		
Gender	-0.25 (0.28)	0.78	[0.45, 1.35]
Age	0.22 (0.14)	1.24	[0.96, 1.64]
Internet usage	0.35 (0.30)	1.43	[0.80, 2.58]
SNS usage	0.03 (0.16)	1.04	[0.76, 1.41]
Privacy preference	<b>0.68 (0.31)*</b>	1.97	[1.09, 3.70]

Table 6.2: Coefficient, standard error, and odds ratio of photo censorship model (if participants have declined to share a photo due to privacy concern)

answered “I don’t know” (Figure. 6.1).

We constructed a logistic regression model to investigate whether gender, age, Internet usage frequency, SNS usage frequency, and participants’ privacy preferences influence photo censorship.

As shown in the first model (Table 6.2), only privacy preference has an effect on photo self-censorship. This indicates that people who have higher privacy consciousness are more likely to censor their photos ( $b = 0.68$ ,  $p = .03$ ). Controlling for other variables, each one point increasing in privacy consciousness leads to a 97% increase in the odds of photo self-censorship (*Odds Ratio* ( $OR$ ) = 1.97).

### 6.3.2 Could Obfuscation Encourage Sharing?

We asked the 124 participants who reported they had self-censored photos whether they would share the photo they self-censored if they had obfuscation options available. Half (51%) of them reported they would be willing to share the self-censored photo on a SNS if they were able to apply an obfuscation to the sensitive content. However, 26% maintained that they would still



	Coefficient (SE)	Odds Ratio	OR 95% CI
Intercept	-3.62 (3.08)		
Gender	<b>1.15 (0.49)*</b>	3.16	[1.27, 8.70]
Age	-0.20 (0.27)	0.82	[0.48, 1.40]
Internet usage	-0.15 (0.47)	0.86	[0.34, 2.17]
SNS usage	-0.31 (0.30)	0.73	[0.41, 1.34]
Privacy preference	<b>1.17 (0.54)*</b>	3.23	[1.19, 10.47]

Table 6.3: Coefficient, standard error, and odds ratio of the model of willingness to share the photo again with obfuscations applied

refuse to share the self-censored photo even if obfuscation options were available. Nearly one fifth of participants (23%) were not sure if they would share the self-censored photo if obfuscation options were available (Figure. 6.2).

We constructed another logistic regression model to investigate whether gender, age, Internet usage frequency, SNS usage frequency, and privacy preferences have effects on users' willingness to share a self-censored photo with obfuscations applied.

This model (Table 6.3) suggests that people with higher privacy consciousness about their personal information were more inclined to report they would share a previously censored photo if they could obscure sensitive content ( $b = 1.17, p = .03$ ). Controlling for other variables, for each one point increase in privacy consciousness there is a 223% increase in the odds of photo self-censorship ( $OR = 3.23$ ). We also find that gender influences willingness to share a previously self-censored photo when an obfuscation is applied ( $b = 1.15, p = .02$ ). Specifically, women are more willing than men to share a photo they had previously self-censored if they were able to obscure the photo to improve privacy. Controlling for other variables, the odds of willingness for women are 3.16 times higher than men ( $OR = 3.16$ ).

## 6.4 Discussion

### 6.4.1 Self-Censorship of Photos Is Prevalent Among SNS Users

Photo sharing is an important activity on SNSs. It helps users maintain a real-world relationship with their family and friends [211]. Beyond their existing off-line social networks, users also expand their social graph through sharing photos, for example, receiving attention from a wider audience, even from the public [211]. Users also manage their impression via photo sharing, such as

selecting photos that emphasize socially desirable characteristics to show an ideal version of themselves [74]. However, privacy issues often hamper photo sharing [6]. The results of our study show that over half of the participants did censor their photos, which is very prevalent. Some common sensitive content that they have concerns about include the identity of people, nudity, appearance, facial expression, inappropriate behavior, and personal identity information [25, 175]. Beyond the sensitive content, there are three main reasons behind photo self-censorship: first, maintaining a good impression; second, personal, family, and property safety; third, sharing certain content may get photo posters into trouble, for example, sharing a confidential work photo [175].

#### **6.4.2 Photo Obfuscations May Be Useful For Encouraging Photo Sharing**

From previous chapters, we know that photo obfuscations can be an effective and satisfying tool to protect users' privacy. Two promising obfuscations—avatar and inpainting—can achieve a very low identification rate and a good user experience. In this study, half of the participants who had declined to upload a photo with sensitive content changed their mind and were willing to share the photo if they could apply one of the obfuscations to their photo. This result suggests that photo obfuscations may have the potential to encourage photo sharing. However, we can see that 23% of participants had no idea if they would like to share the obscured photo. A possible reason may be that in the experiment, the obfuscation photo examples they saw were not their own photos, hence they might be skeptical about the actual application on their own photos. Overall, the result shows users' positive attitudes towards photo obfuscations in photo sharing on SNSs.

#### **6.4.3 People With Higher Level of Privacy Consciousness And Women Are More Willing To Share Privacy-Enhanced Obfuscated Photos Which They Previously Had Privacy Concerns**

First, for photo self-censorship, as we expected, people with higher levels of privacy consciousness of their personal content are more likely to censor photos. Regarding the gender, previous studies do not agree as to the impact of gender difference in self-disclosure that may lead to distinct photo self-censorship patterns. One work suggests that men are likely to disclose more about themselves and possibly have fewer privacy concerns [189], while other work indicates that men are less comfortable to disclose themselves, and men censor more posts than women [60, 246]. However, in

the context of photo sharing, gender does not have an effect on photo self-censorship. Additionally, we know that older users censor fewer posts than younger users [60], so we expected to find similar patterns on photo self-censorship. However, the results indicate that age does not have an effect.

Second, when we asked if they were willing to share the photo with the access to one of the obfuscations, people who have higher privacy-consciousness are more likely to be willing to share photos which they previously refused to share due to privacy concerns, if they could apply obfuscations. This result again suggests that from the users' perspective, photo obfuscations may sufficiently reduce their privacy concerns on sensitive portions of their photos, and thereby encouraging or enabling sharing. Objectively, from a technical perspective, some promising obfuscations (e.g., inpainting, avatar) are indeed effective in hiding the sensitive content, so it seems reasonable that participants would be willing to share more photos if they had privacy-enhancing obfuscations available for use.

Besides the privacy consciousness, there was a difference between men and women. Women are more willing to share obfuscated photos about which they previously had privacy concerns. This finding may be in line with the statement that men are still less comfortable to disclose their personal information [246], even if they are given the choice to obscure the sensitive content in a photo. Women use SNSs more frequently [139] and they are more likely to post photos than men on SNSs [184, 204]. On the other hand, they are also more concerned about their self-presentation [67, 159] and safety [171], and usually actively seek strategies to protect privacy [184]. Hence, women may have more demand for effective privacy protection mechanisms, such as obfuscation, that allow them to share more photos while at the same time preserving privacy. Other possible reasons may be that men consider that altering their photos has a negative effect on their self-representation (e.g., not a true self). While compared to men, women reported editing and beautifying photos more frequently before posting [97], such as using photographic filters or Photoshop, so there is less resistance when adopting obfuscations. These potential explanations offer a rich space for additional exploration about privacy, photo sharing, and the potential of obfuscations to differentially benefit different user groups.

## 6.5 Limitations and Future Work

First, our study is based on a participants' self-report, which is very prevalent in most areas of the social sciences [194]. However, self-report has weaknesses, for example, people may respond in a way that presents them in a positive light [221], or respond without carefully considering the question [222]. This concern can be addressed by directly observing users' self-censorship behavior on SNSs. Next, though we uncovered half of the participants were willing to upload a previously self-censored photo if they could use obfuscation, the obfuscation example photos they saw were not their own sensitive photos. Our results can be strengthened if we ask them to upload their own photos about which they have privacy concerns, then apply different obfuscations to their photos. Moreover, the best solution would be to create an obfuscation mechanism on Facebook combining with their existing recipient control approach and let participants actually upload photos as they usually do, in which we can confidently conclude if photos obfuscations reduce photo self-censorship and encourage sharing. This was implemented in my next study. Moreover, though the 230 participants in our study are representative in which most of the participants were frequent SNS users, compared to the previous Facebook self-censorship study which has five million Facebook users [60], we believe increasing the sample size may further strengthen our results. Lastly, because we used standardized demographic questions from Pew Research Center [227], we only offered participants gender response options that do not reflect the spectrum of gender identities. Because of this limitation we were only able to analyze gender difference between men and women. In future studies we will include an option "Other: specify [text box]" as suggested in [128].

## 6.6 Chapter Conclusion

Photo self-censorship is common. In our study of 230 participants from the United States, over half of the participants have self-censored photos due to privacy concerns. Further, we found that privacy-conscious people were more likely to self-censor photos. We also uncovered that photo obfuscations might be useful for combating photo self-censorship. Among the participants who reported they had self-censored photos, half of them were willing to share the previously self-censored photo if they would be able to obfuscate portions of the photo to enhance privacy. Additionally, we learned that people with higher levels of privacy consciousness and women were more willing to share photos about which they previously had privacy concerns if they were able to apply obfuscations on

them. This work indicates that obfuscations are a promising photo privacy protection mechanism. However, as I stated in the Limitation section, self-report has its own limitations. Hence, in the next chapter, I describe a study in which participants were allowed to see the effect of obfuscations on their own photos which they had privacy concerns with.

## Chapter 7

# Study 5: An Experiment to Determine Whether Obfuscation Reduces Privacy Concerns and Increases Willingness to Share

Note: We plan to submit this chapter to CHI2021.

### 7.1 Introduction

In the previous chapters, I described the phases of identifying sensitive content in photos, understanding people’s sharing preferences, investigating effective and usable obfuscation methods, understanding sharing loss due to self-censorship, and understanding obfuscation’s potential to prevent sharing loss. The next step is to, based on our prior studies, design and test an interface that enables privacy-enhanced photo sharing. We propose a privacy-enhanced photo sharing interface which helps users identify potentially sensitive content and provides them with easy to use obfuscation options.

Specifically, in this chapter, I present a study in which we compare three possible interfaces

for a photo-sharing system: control, warning, and obfuscation. The control condition is a replica of Facebook’s current photo-sharing interface. The warning condition identifies sensitive content and provides a visual privacy warning about it to users. The warning condition is similar to work by Want et al. [290]. The obfuscation condition identifies sensitive content in photos and offers users the opportunity to choose an obfuscation to obscure the sensitive content in the photo. This is a novel interface option we have invented on the basis of the work described in this dissertation.

We evaluated the three conditions using the following set of metrics: perceived privacy risks, willingness to share, ease of use, perceived system effectiveness, and system satisfaction. We found that the obfuscation condition performs the best among all three versions in terms of reducing perceived privacy risks and increasing willingness to share. People also perceive it to be effective and satisfying. On the other hand, perhaps because the warning version does not provide any actionable solutions to protect privacy, it increases the perceived privacy risks, and in turn, decreases willingness to share.

## 7.2 Photo Privacy Protection Interface Design

The obfuscation interface includes two core design features—detecting and highlighting any potentially sensitive content in a photo and obfuscating the sensitive content. The first step is for a user to identify a photo they would like to share. The photo could be stored on their phone for example. Next, the user receives a prompt via the interface that potentially sensitive content has been detected in the photo (e.g., the sensitive content in the example Figure 7.1 is the face of a user’s brother). The identification of sensitive content can happen on the users’ phone (for maximum privacy) or on an SNS’s server upon upload with cryptographic guarantees in place that processing will not reveal the content to the SNS [71, 275], for example.

Next, the user can choose which obfuscation method they would apply to the identified sensitive content (Figure 7.2). In this study, we provided users five obfuscation options—blurring, masking, avatar, inpainting, and no filter. In previous studies, we identified that avatar, inpainting, and masking are effective against human recognition. Though blurring is not as effective as the other three obfuscation methods, it is used extensively and people perceive it likable [180]. Finally, the user can make a decision about sharing the obfuscated photo.

The primary option for users to manage the privacy of photos on SNS has been untag-

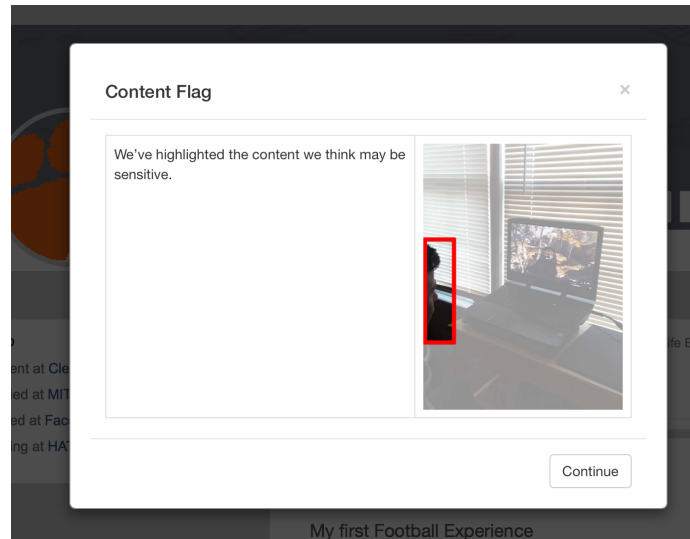


Figure 7.1: Photo protection interface– content detection and obfuscation options.

ging [25]. Untagging is when a user removes the link between a photo and their identity. Researchers have also explored soft paternalism or nudging. For example, a privacy wizard nudges users to adjust their privacy settings [88]. It asks users to assign privacy labels to selected friends, then the trained classifier automatically categorizes the remaining friends. Wang et al. proposed three nudging systems that encourage users to make adjustments to their Facebook posts in real-time [290]. However, our work, to the best of our knowledge, is the first to propose using obfuscation to enhance the privacy of shared photos using obfuscation.

### 7.3 Method

This is a two-step experiment. In the first step, we collected photos that people wanted to share via an SNS but have not due to privacy concerns. During this step, we also asked participants to identify the sensitive portion in their photos. These photos were subsequently used as stimuli in the second step to providing participants with a personal, realistic experience that we expect will have more ecological validity than if we had used generic photos. The photo that each participant was asked to make decisions about and provide input about their perceptions was were their own photo. We did not have to ask participants to assume it was their own photo as is a common method in many photo privacy studies. Between the two steps, I manually processed and obfuscated



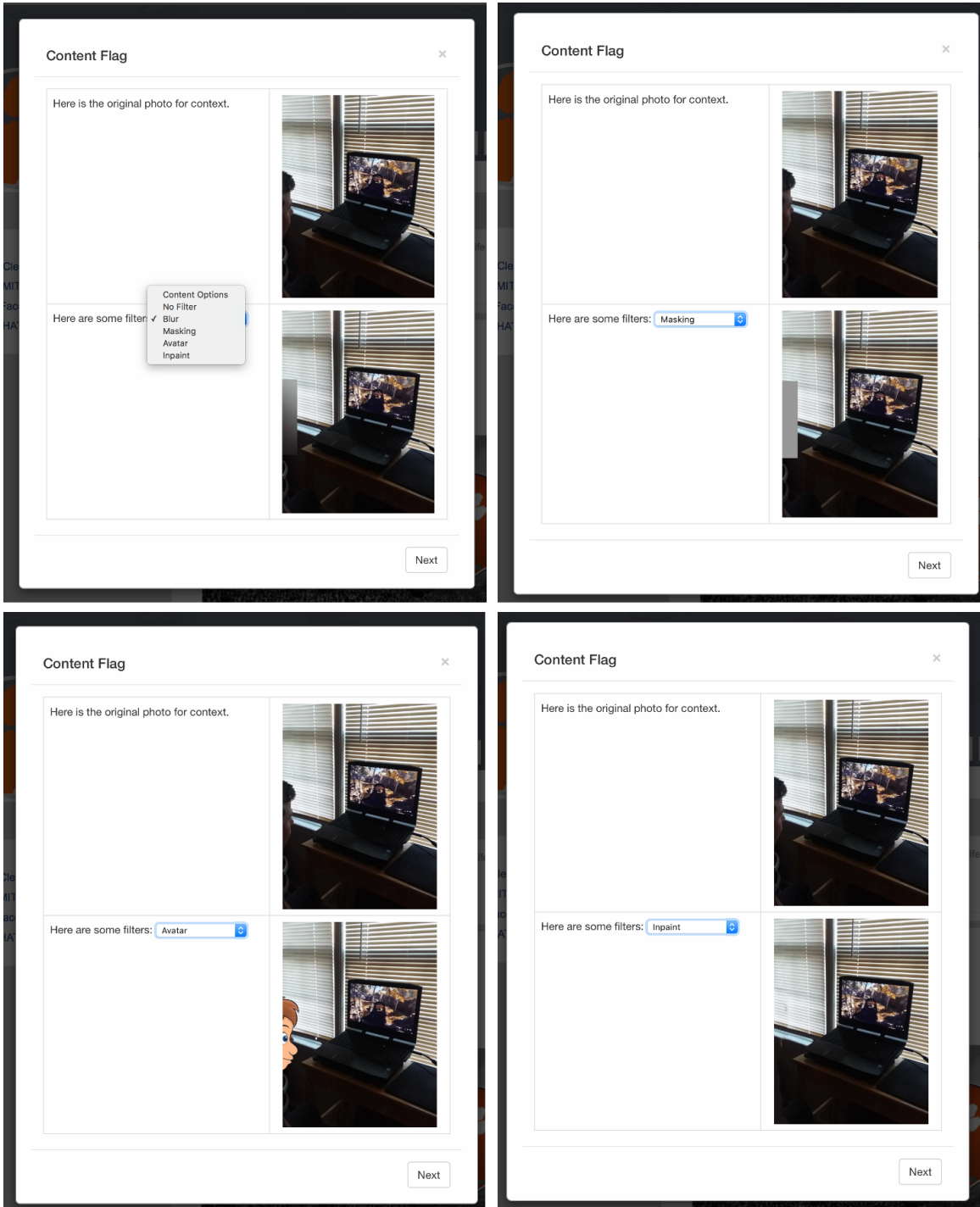


Figure 7.2: Photo protection system interface—obfuscation options.

participants' photos so that these would be available as stimuli for the second step.

In the second step, participants were randomly assigned to one of three conditions: the control interface, the warning interface, or the obfuscation interface. After using the interface they were randomly assigned to they evaluated the interface. We used these data to determine whether there the obfuscation interface can 1) reduce users' perceived privacy risks, 2) prevent sharing loss, and 3) be usable. In the following subsections, I will describe the two steps in detail.

All studies were IRB approved.

### **7.3.1 Step One: Photo Collection**

In the first step, we collected photos from participants via Qualtrics. We first asked participants demographic questions, Internet usage frequency, SNS usage frequency, Facebook usage frequency, photo uploading frequency, and whether they have ever declined to upload a photo to an SNS such as Facebook due to their privacy concerns. They also needed to provide their email addresses to be contacted for the second step of the study.

For participants who passed the screening questions (see inclusion criteria in the "Participants" subsection), they then provided one to five photos that they would like to upload to an online social network such as Facebook, but have NOT because of privacy concerns. We used these photos to create photo stimuli for the second step. If they were not willing to share a personal photo with us, they could upload an online photo that had similar content. For each photo, we asked two follow up questions:

- What content prevents you from posting this photo?
- What about this photo makes you want to upload if you did NOT have privacy concerns?

Through the first follow up question, we were able to identify the sensitive content they had concerns about. In the second step, we applied privacy-enhancing obfuscations to participants' self-identified sensitive content. By looking at the answers to the second question, we were able to assess the reasons participants had for wanting to share the photo if only the privacy concerns about the photo could be reduced.

### 7.3.1.1 Participants

We recruited participants in two rounds. In total, we recruited 439 participants (126 in the first round and 313 in the second round) located in the United States through the Qualtrics sourcing platform. We included screening questions to identify suitable participants. Please see the inclusion criteria in the parenthesis after each question.

- Please browse your photo album for at least three minutes. How many photos are there in your album that you would like to upload to an online social network such as Facebook, but have NOT because of privacy concerns? (Must have at least one such photo)
- Are you willing to provide us these photos so that we can use them as your study material in the second step–Facebook study? (Must be willing to provide us these photos)
- Are you available to participate in the Facebook study a week later? (Must be willing to participate in the second step study)
- About how often do you use or visit Facebook? (Must have a Facebook account and use it at least a few times a year)
- About how often do you upload photos to an online social network such as Facebook? (Must have the experience of uploading photos to Facebook)

We paid Qualtrics \$17 per qualified participant. To ensure high data quality, we included three attention check questions in the Qualtrics survey. After excluding the participants who did not pass the screening and attention check questions, the final sample size was 310 (86 in the first round and 224 in the second round). The demographics can be found in Table 7.1.

### 7.3.2 Step Two: Experiment

For the second step in this study, we conducted a mixed between and within-subjects experiment with a pre-test post-test design. The between-subjects factor is the three experimental conditions (control, privacy warning, and obfuscation) and the within-subjects factors are grouping, rather than independent variables: two personal characteristic variables – interpersonal privacy concerns and trust in SNSs. Both were measured on 7-point Likert scales and then converted to categorical variables with two levels – high and low for analysis.

		No.	%
<b>Gender</b>	Male	108	35%
	Female	202	65%
<b>Age</b>	18-24	44	14%
	25-34	124	40%
	35-44	93	30%
	45-54	35	11%
	55+	15	5%
<b>Ethnicity</b>	White	217	70%
	Hispanic or Latino	24	8%
	Black or African American	45	15%
	Native American or American Indian	1	0%
	Asian/Pacific Islander	22	7%
	Other	2	0%
<b>Internet Usage Frequency</b>	Most of the day	211	68%
	Several times a day	98	32%
	About once a day	2	0%
	A few times a week	0	0%
	A few times a month	0	0%
	A few times a year	0	0%
	Never	0	0%
<b>SNS Usage Frequency</b>	Most of the day	105	34%
	Several times a day	168	54%
	About once a day	28	9%
	A few times a week	8	3%
	A few times a month	2	0%
	A few times a year	0	0%
	Never	0	0%
<b>Photo Uploading Frequency</b>	Many times a day	42	14%
	Several times a day	29	9%
	About once a day	31	10%
	A few times a week	102	33%
	A few times a month	57	18%
	A few times a year	50	16%
	Never	0	0%

Table 7.1: Participants' demographics in step one.

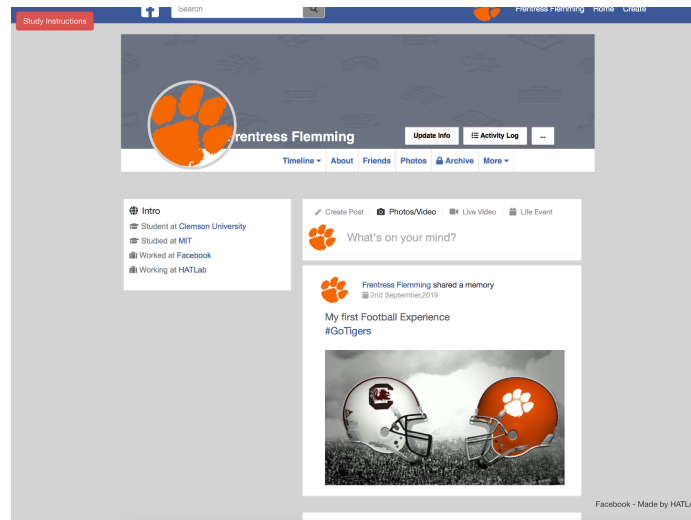


Figure 7.3: A screenshot of the control condition. The control condition mimics Facebook current photo sharing interface.

The five dependent variables are perceived privacy risks, willingness to share, ease of use, perceived system effectiveness, and system satisfaction. More details about the scales can be found in the “Measurements” subsection below.

Participants were randomly assigned to one of three groups. In the control group, participants used a prototype which mimics Facebook’s current photo uploading feature (Figure 7.3 and Figure 7.4). In the privacy warning group, participants used a prototype that identified sensitive content and provided a privacy warning (Figure 7.5). In the obfuscation group, participants used a prototype that identifies sensitive content and provides an opportunity to obfuscate that content (Figure 7.1 and Figure 7.2). Please note that in all three versions, we kept the recipient control feature that is available on Facebook. Participants could choose “public,” “friends,” or “only me” when they decided to share a photo.

We used the wizard of oz approach to simulate how the interface for each condition would perform. Simulating the functionality of a prototype allows researchers to explore and evaluating designs to test aspects of them and improve them before investing the considerable time, effort, and money in implementing the system [75]. We simulated the photo-sharing interface for all three conditions. We pre-selected one photo from photos that participants had shared with us during the first step. Then, during the second step, when participants clicked the “Photo/Video” button, the folder contained that photo. We also simulated the obfuscation step by manually obfuscating all

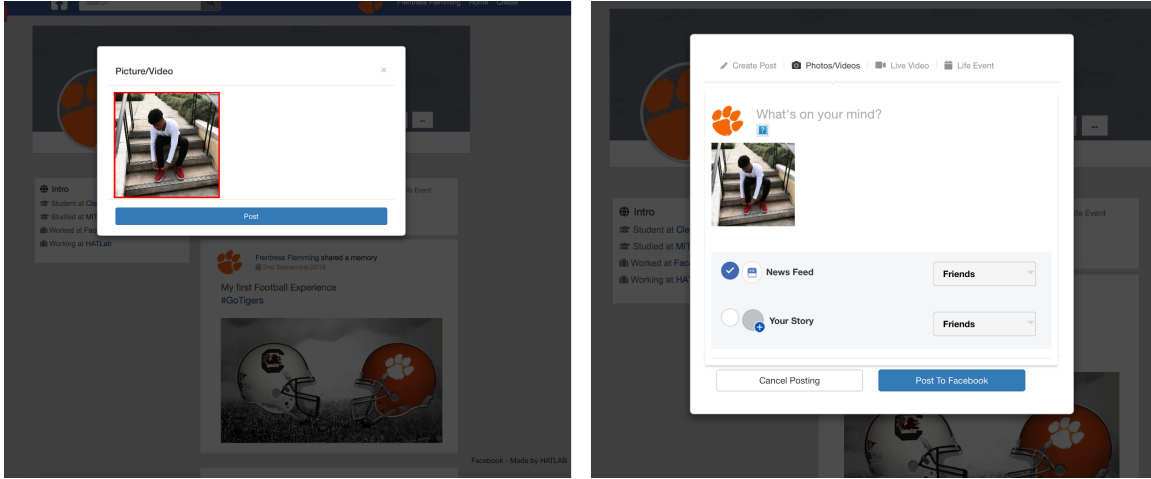


Figure 7.4: A screenshot of the control condition. The control condition mimics Facebook current photo sharing interface.

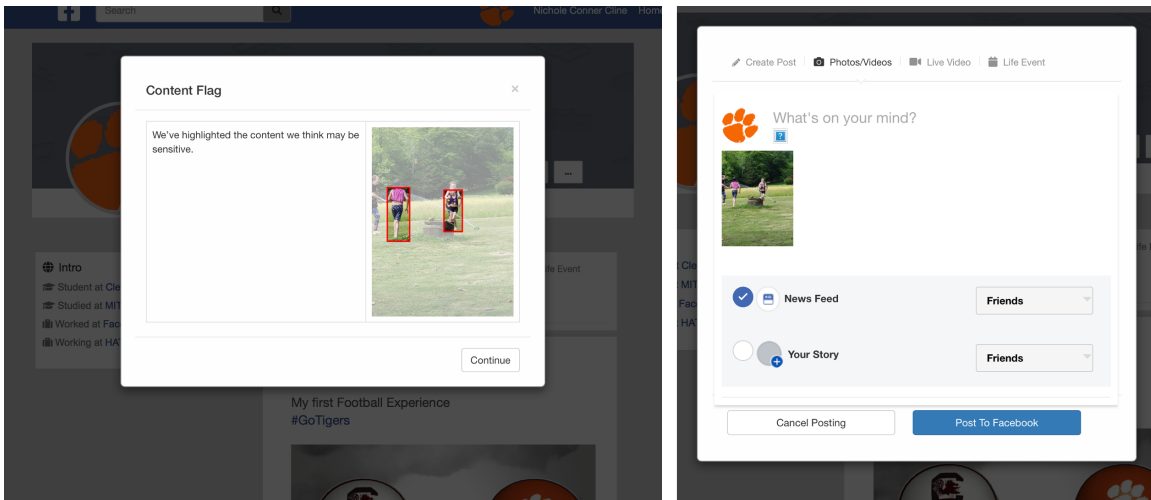


Figure 7.5: Privacy warning condition.

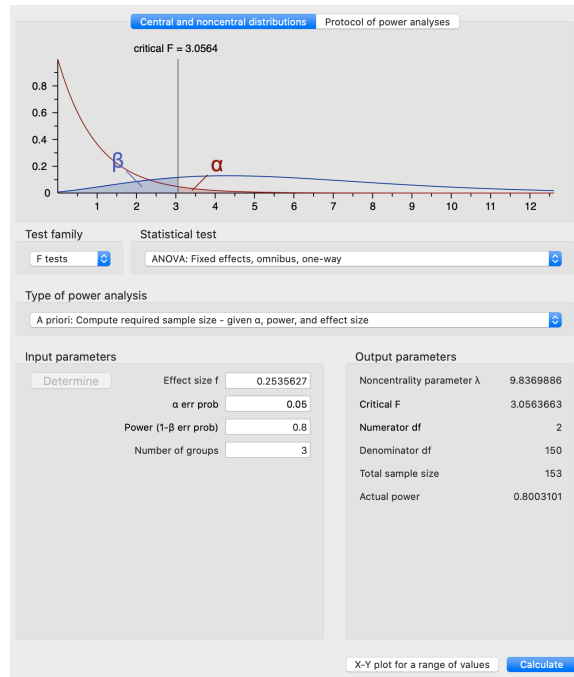


Figure 7.6: Results of the power analysis.

photos for participants in the obfuscation condition and stored them in a SQL database. Then when participants in the study chose an obfuscation option, it was available for them to view during the study period. To the participant, it seemed like the photo was being obfuscated in real-time.

### 7.3.3 Participants in Step 2: Experiment

We conducted a power analysis using the data from a pilot study with 15 participants. The power analysis indicated that to find a difference for variables we are interested in such as willingness to share with sufficient power, a sample size of 153 would be required (Figure 7.6).

We contacted participants who were in the first step via email. From the 310 qualified participants, 158 returned for the second step of the study resulting in a 51% return rate.

Participants were randomly assigned into three condition groups—54 in the control group, 51 in the privacy warning group, and 53 in the obfuscation group. Participants were given a \$10.00 Amazon eGiftcard upon completion of the 30-minute study. The demographic data can be found in Table 7.2.

		No.	%
<b>Gender</b>	Male	55	35%
	Female	103	65%
<b>Age</b>	18-24	19	12%
	25-34	63	40%
	35-44	49	31%
	45-54	18	11%
	55+	9	6%
<b>Ethnicity</b>	White	120	76%
	Hispanic or Latino	5	3%
	Black or African American	19	12%
	Native American or American Indian	0	0%
	Asian/Pacific Islander	13	9%
	Other	1	0%
<b>Internet Usage Frequency</b>	Most of the day	104	66%
	Several times a day	52	33%
	About once a day	2	1%
	A few times a week	0	0%
	A few times a month	0	0%
	A few times a year	0	0%
	Never	0	0%
<b>SNS Usage Frequency</b>	Most of the day	50	32%
	Several times a day	84	53%
	About once a day	16	10%
	A few times a week	7	4%
	A few times a month	1	0%
	A few times a year	0	0%
	Never	0	0%
<b>Photo Uploading Frequency</b>	Many times a day	17	11%
	Several times a day	17	11%
	About once a day	13	8%
	A few times a week	56	35%
	A few times a month	29	18%
	A few times a year	26	16%
	Never	0	0%

Table 7.2: Participants' demographics in step two.



### 7.3.4 Measurements

A Qualtrics survey was constructed which included two portions: 1) a pretest consisting of four parts (trust in Facebook, interpersonal privacy concerns, perceived privacy risks of the photo, and willingness to share), and 2) a posttest consisting of six parts (perceived sensitivity of the photo, willingness to share, ease of use, perceived system effectiveness, system satisfaction, and qualitative feedback). All the responses used 7-point Likert scale from 1 ‘Strongly disagree’ to 7 ‘Strongly agree.’ Please note that trust in Facebook and interpersonal privacy concerns are between-subjects covariates that were only measured at pre-test. All the measurements can be found in Table 7.3.

At the end of the post-test survey, participants provided their qualitative feedback by answering the following two questions:

- What did you like the most about using this system? Please tell us the reasons.
- What did you like the least about using this system? Please tell us the reasons.

### 7.3.5 Procedure

In the emails I sent to participants, besides introducing the goal of this study, I included a photo that would be used as photo stimuli, so that they could answer the two pre-test questions about perceived privacy risks and willingness to share this photo. I also provided participants with login credentials so that they could access the prototype designed for them after finishing the pre-test questions. Participants then went to the Qualtrics survey link and completed the pre-test questions. Afterward, they were guided to the prototype they had been randomly assigned to, performed the photo uploading task. Participants in the obfuscation condition could play around with the four obfuscation options provided – blurring, masking, avatar, and inpainting. They could see the effects of applying each obfuscation on their own photo. They got a validation code when they successfully finished the task, and went back to the survey. Participants pasted random codes generated by the prototype in the survey to show that they had finished the task. Next, they completed the post-test questions and received another validation code at the end of the survey. Upon completion, participants contacted me for incentives. I checked their survey data and prototype back-end data to ensure data quality and distributed Amazon eGiftcards. Data from nine participants were excluded either because they did not interact with the interface or failed more than one attention check question.

	Scale	Item
Pre-test	Trust in SNS [78]	<ul style="list-style-type: none"> <li>- I trust that Facebook will not use my personal information for any other purpose.</li> <li>- I feel that the privacy of my personal information is protected by Facebook.</li> <li>- I believe most of the profiles I view on Facebook are exaggerated to make the person look more appealing.</li> <li>- I believe most of the profiles I view on Facebook are exaggerated to make the person look more appealing.</li> </ul>
	Interpersonal privacy concerns [144]	<ul style="list-style-type: none"> <li>- It usually bothers me when people ask me something personal.</li> <li>- I will tell people anything they want to know about me.</li> <li>- I have nothing to hide from other people.</li> <li>- I am concerned that people know too many personal things about me.</li> <li>- To me, it is the most important thing to keep things private from others.</li> <li>- When people ask me something personal, I sometimes think twice before telling them.</li> <li>- I think it is risky to tell people personal things about myself.</li> <li>- I feel safe telling people personal things about me.</li> <li>- I feel comfortable sharing my private thoughts and feelings with others.</li> </ul>
	Perceived privacy risks [90]	<ul style="list-style-type: none"> <li>- How risky would you say it would be to post this photo on Facebook? (Not risky at all - very risky)</li> <li>- Posting this photo on Facebook would be risky.</li> <li>- Posting this photo on Facebook is dangerous.</li> <li>- Sharing this photo on my Facebook would add great uncertainty to my privacy.</li> <li>- Sharing this photo on my Facebook exposes me to an overall risk.</li> </ul>
	Willingness to share [188]	<ul style="list-style-type: none"> <li>- I am willing to share this photo on my Facebook.</li> </ul>
Post-test	Perceived privacy risks [90]	<ul style="list-style-type: none"> <li>- How risky would you say it would be to post this photo on Facebook? (Not risky at all - very risky)</li> <li>- Posting this photo on Facebook would be risky.</li> <li>- Posting this photo on Facebook is dangerous.</li> <li>- Sharing this photo on my Facebook would add great uncertainty to my privacy.</li> <li>- Sharing this photo on my Facebook exposes me to an overall risk.</li> </ul>
	Willingness to share [188]	<ul style="list-style-type: none"> <li>- I am willing to share this photo on my Facebook.</li> </ul>
	Ease of use (usage effort reversed) [142]	<ul style="list-style-type: none"> <li>- The system is convenient.</li> <li>- I do not have to invest a lot of effort in the system.</li> <li>- It takes many mouse-clicks to use the system.</li> </ul>
	Perceived system effectiveness [142]	<ul style="list-style-type: none"> <li>- This system has no real benefit for me.</li> <li>- This system is useful.</li> <li>- I can protect my privacy better using this system.</li> <li>- I can protect my privacy better using other approaches without the help of this system.</li> </ul>
	System satisfaction [205]	<ul style="list-style-type: none"> <li>- I am very satisfied when using this system.</li> <li>- I am very pleased when using this system.</li> <li>- Using this system made me contented.</li> <li>- I feel delighted when using this system.</li> <li>- I will strongly recommend it to my friends.</li> <li>- I will most likely use this system again.</li> </ul>

Table 7.3: Pre and post-test measurements.

## 7.4 Results

For the analysis, we used path analysis which is a special case of Structural Equation Modeling (SEM) and also can be viewed as an extension of a regression model [267]. A path model allows us to test the structural relations between all variables in a single model. For example, via regression models we can see the direct effects of experimental conditions on perceived privacy risks and willingness to share, while a path model also allows us to examine the relationship between the experimental conditions, perceived privacy risks, and willingness to share. In this way, we are able to see how variables are related to one another.

We first conducted a confirmatory factor analysis (CFA) which allows us to test whether measures of a construct are consistent with our understanding of the nature of that construct. From the results of CFA, we found that system effectiveness is highly correlated with system satisfaction ( $r = 0.80$ ), which means they are essentially measuring the same thing, hence we decided to remove system effectiveness from analysis. After removing effectiveness, the model fit was good:  $\chi^2/df = 2.207$ ;  $RMSEA = 0.050$ , 90% CI : [0.024, 0.074],  $CFI = 0.996$ ,  $TLI = 0.995$ .

Please note that in the obfuscation condition, when asked to select an obfuscation, 13 participants chose “no filter,” hence I removed their pre- and post-test data of perceived privacy risks and willingness to share from all the analysis in this section.

### 7.4.1 Effects of Experimental Conditions on Dependent Variables

Since we are interested in the differences between experimental conditions, before looking at the path model, I will show the effects of experimental conditions on the four dependent variables and the interaction effects between conditions and the two characteristic variables – interpersonal privacy concerns and trust in SNS (if there is one).

To test the interaction effects, I used median split to convert interpersonal privacy concerns and trust in SNSs into categorical variables (7-point Likert scale to binary high vs. low). Median splits are useful when examining interaction effects because it is easier to create dummy variables for the interactions between two categorical variables. Perceived privacy risks, usage effort, and system satisfaction are latent variables that have more than one item. To be able to analyze the data, we need composition scores for each scale. Instead of calculating the means of the items, we chose to calculate their factor scores which are more accurate [70]. In Figure 7.7, all the numbers

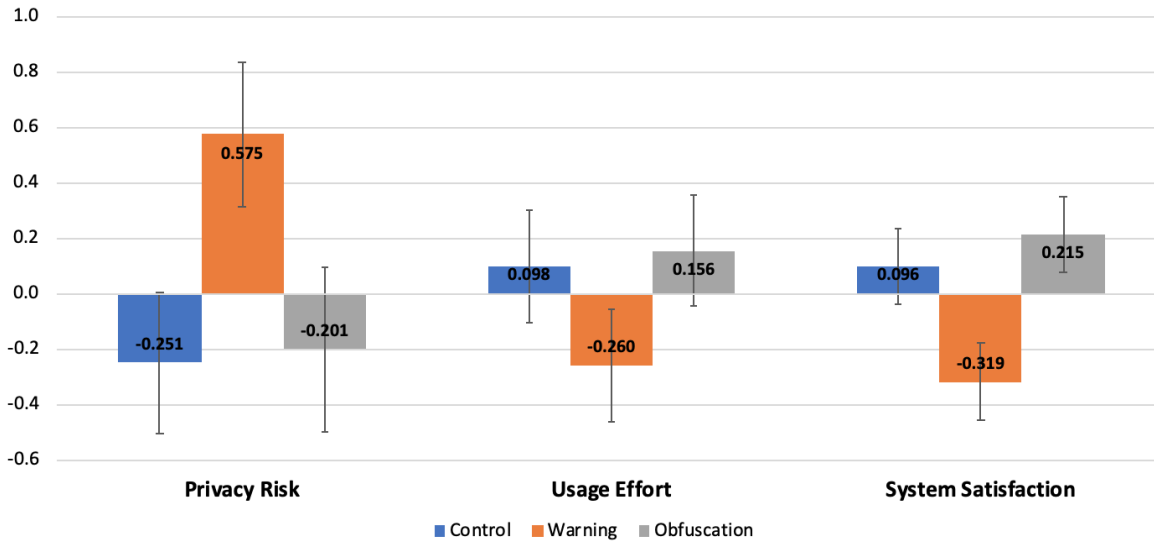


Figure 7.7: Marginal effects of between-subjects independent variable interface condition on perceived privacy risk, usage effort, and system satisfaction. All three metrics were measured in post-test.

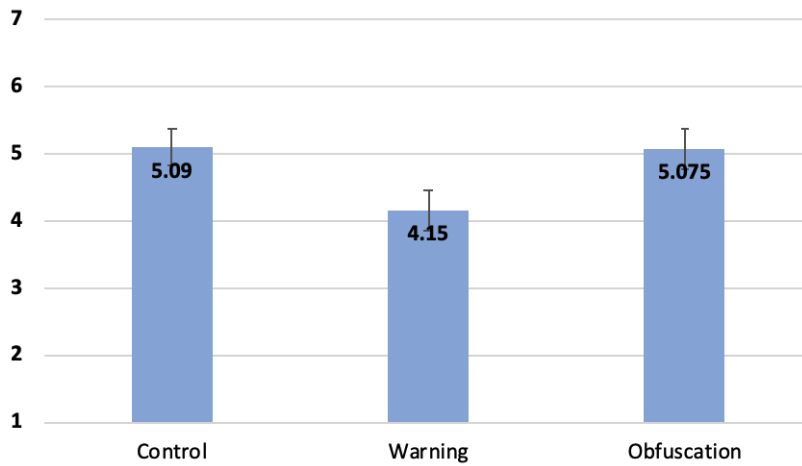


Figure 7.8: Marginal effects of privacy enhancing conditions on willingness to share (measured in post-test)

are factor scores and we are focusing on the comparisons among the conditions but not the numbers themselves. The scale of willingness to share only has one item, so I used means to plot the bar chart (Figure 7.8).

I created two linear regression models for each dependent variable. In the first model, I tested the main effects of conditions and interpersonal privacy concerns and the interaction effects between them. In the second model, I tested the main effects of conditions and the trust in SNS and the interaction effects.

For perceived privacy risks, the effect of experimental conditions is neither dependent on interpersonal privacy concerns nor trust in SNS. It shows that participants in the warning condition felt more privacy risks regarding their photos than participants in the control condition ( $p = .03$ ), while there is no difference between the obfuscation condition and the control condition ( $p = .89$ ) (Figure 7.7).

Similarly, for willingness to share, there is no interaction effect between the conditions and the two characteristic variables. Participants in the privacy warning condition were less willing to share their photos compared to those in the control condition ( $p = .02$ ), but there is no difference between the obfuscation condition and the control condition ( $p = .97$ ) (Figure 7.8).

Regarding ease of use, no interaction effect was found between the conditions and interpersonal privacy concerns. On the other hand, we found that the effect of conditions on ease of use is dependent on the trust in SNS (Figure 7.9). Specifically, people who have high trust in SNS feel that it is easier to use the warning version compared to the control version ( $p = .04$ ); and people who have low trust in SNS felt more effort was required in the warning version than the control condition ( $p = .04$ ); while for the obfuscation version, there is no difference in perceived ease of use between people who have low and high trust in SNS ( $p = .25$ ).

For system satisfaction, similar to ease of use, there is no interaction effect between the conditions and interpersonal privacy concerns. However, we found that the effect of experimental conditions on satisfaction is dependent on the trust in SNS (Figure 7.10). People who have high trust in SNS perceive the warning version to be more satisfying compared to the control version (both  $p = .04$ ). Additionally, people who have low trust in SNS perceive the warning version to be less satisfying than the control version ( $p = .03$ ).

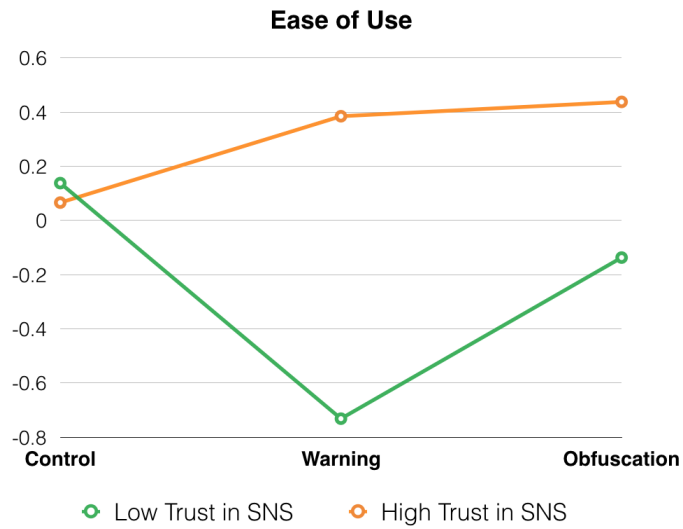


Figure 7.9: Interaction effects between experimental conditions and trust in SNS on ease of use (measured in post-test).

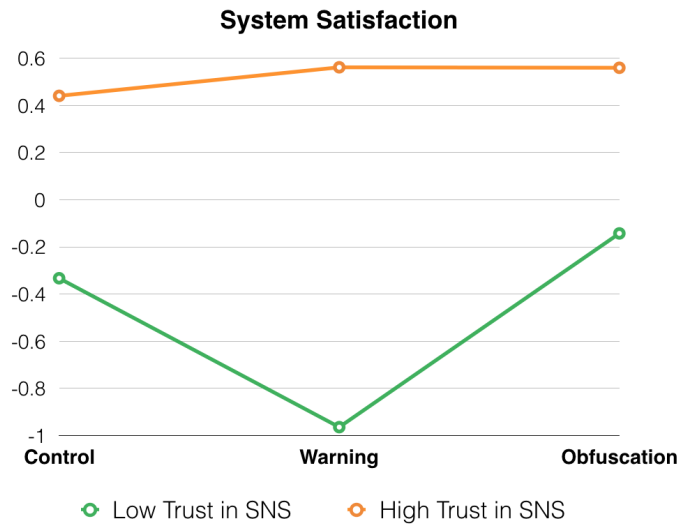


Figure 7.10: Interaction effects between experimental conditions and trust in SNS on system satisfaction (measured in post-test).

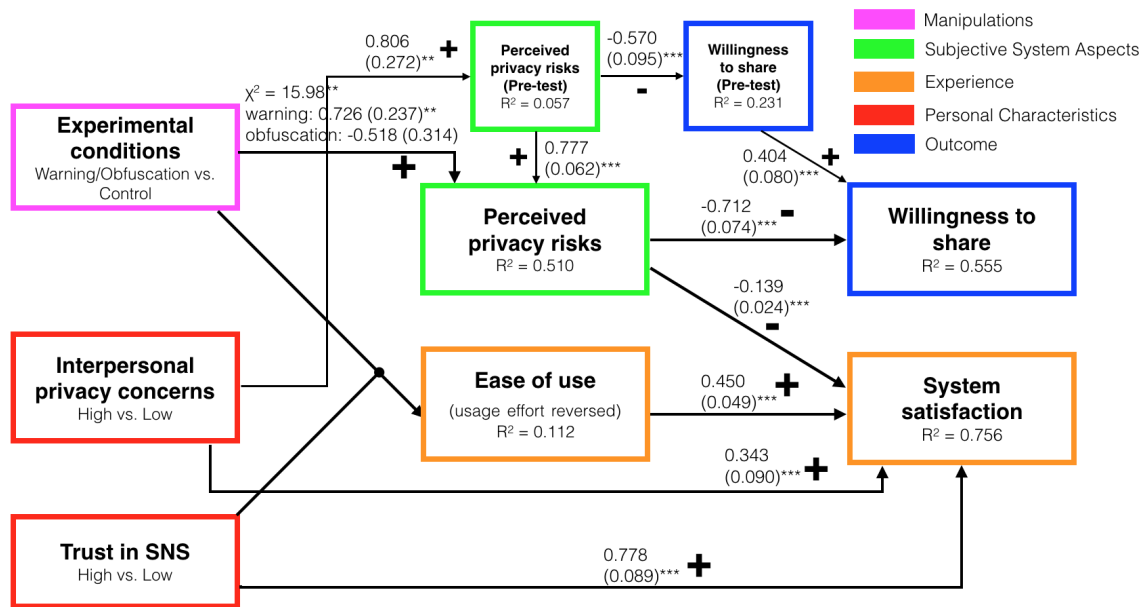


Figure 7.11: The path model shows that experimental conditions has effect on perceived privacy risks, and the increased privacy risks decreases willingness to share.

## 7.4.2 Path Model Results

To investigate the relationships among all variables, we constructed a path model. We started with a saturated model with all possible paths including the interaction effects between experimental conditions and the two characteristics variables (pre and post), then trimmed the insignificant paths. Figure 7.11 shows the final model. The model fit is acceptable:  $\chi^2(34) = 56.921$ ,  $p = .008$ ;  $RMSEA = 0.069$ , 90% CI : [0.035, 0.100],  $CFI = 0.952$ ,  $TLI = 0.929$ . Please note that I created two dummy variables for privacy enhancing conditions – privacy warning and obfuscation (vs. control) – to look at the differences between the two experimental conditions and the control condition, one for interpersonal privacy concerns (high vs. low), one for trust in SNS (high vs. low), and four for interaction effects between the conditions and the two personal characteristics variables.

I will first start by looking at perceived privacy risks and willingness to share. Figure 7.11 shows that the overall experimental condition has a significant effect on perceived privacy risks. After using the system, participants in the privacy warning condition felt more risks about their photo compared to those in the control condition, while there is no difference between the obfuscation condition and the control condition. Additionally, as we expected, pre perceived privacy risks increases

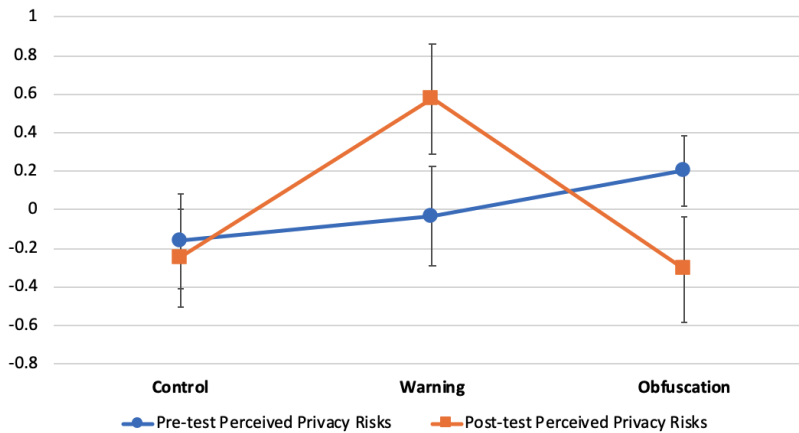


Figure 7.12: Pre and post tests results of perceived privacy risks.

post perceived privacy risks. Perceived privacy risks is in turn negatively related to willingness to share and system satisfaction. Between the two pre-test variables, we also found a relationship that pre perceived privacy risks decrease pre willingness to share. Similarly, pre willingness to share has a positive effect on post willingness to share.

In terms of the characteristics variable – interpersonal privacy concern, it has indirect effects on post perceived privacy risks and pre willingness to share, but I also find its direct positive effect on system satisfaction. Regarding the other characteristics variable – trust in SNS, for people have low trust in SNSs, they felt it is harder to use the warning condition compared to the control ( $p = .04$ ), while for people have high trust in SNSs, they considered it easier to use the warning ( $p = .04$ ) and obfuscation ( $p = .02$ ) versions compared to the control version (see Figure 7.9). Ease of use in turn increases the system satisfaction. There is also a positive direct effect of trust in SNS on system satisfaction.

### 7.4.3 Comparison Between Pre and Post Tests

Participants rated their perceived privacy risks and willingness to share before and after using the system. First, by looking at the pre-test results in each experimental condition, we can see that the random assignment we conducted spread participants evenly across conditions. We expected to see no difference in pre-test scores across the three groups. Regarding perceived privacy risks, we did not find differences between pre-test scores of experimental conditions (warning vs. control:  $p = .70$ , obfuscation vs. control:  $p = .27$ ). Similarly, there is no difference between pre-test



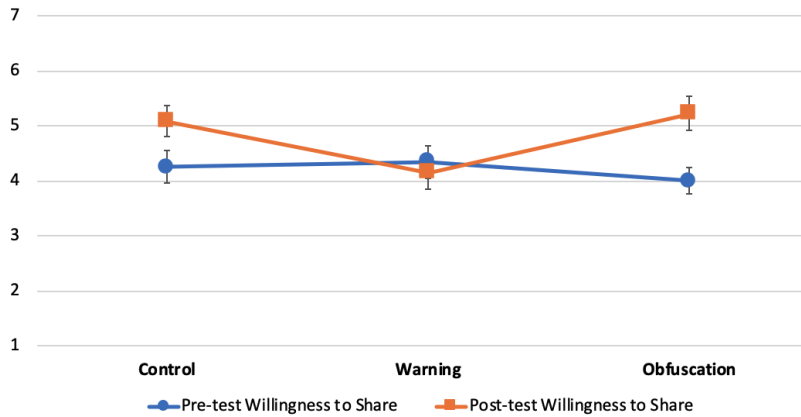


Figure 7.13: Pre and post tests results of willingness to share.

scores for willingness to share (warning vs. control:  $p = .82$ , obfuscation vs. control:  $p = .51$ ). The results confirm that participant assignments were successfully randomized.

Second, we conducted factorial ANOVA to test the differences between pre- and post-tests for the three conditions. Figure 7.12 shows the line chart of pre- and post-test scores of perceived privacy risks on which the y-axis is the rating of willingness to share on a 7-point Likert scale (since this scale only has one item instead of several, it is not a factor score as perceived privacy risks chart shows). There is an interaction effect between experimental conditions and pre- and post-test. For the control condition, there is no difference between the pre and post-tests in terms of the perceived privacy risks ( $p = .50$ ), For the privacy warning condition, participants perceived more privacy risks after using the system than before using the system ( $p = .01$ ), while in the obfuscation condition, the rating for perceived privacy risks in the post-test is lower than in the pre-test ( $p = .02$ ). We can see a larger difference between pre- and post-test scores in the obfuscation condition (pre mean = 0.20 vs. post mean = -0.31) than in the control condition (pre mean = -0.16 vs. post mean = -0.25), and this difference is significant ( $p = .02$ ).

Regarding willingness to share, again there is an interaction effect between experimental conditions and pre- and post-test. We noticed that participants reported they would be more willing to share the photo after using the control system ( $p = .0004$ ) (Figure 7.13). It is probably because when participants thought of a photo that they really want to share but have not due to privacy concerns, they did not consider the privacy-enhancing tool – recipient selection – that Facebook currently provides. After participants used the control system and were reminded they could select

	Recipient			Whether post to Facebook	
	Only me	Friends	Public	Yes	No
Control	6%	92%	2%	88%	12%
Warning	17%	83%	0%	81%	19%
Obfuscation	10%	90%	0%	88%	12%

Table 7.4: Percentages of each recipient group and whether or not participants posted photos.

recipients, they perceived the privacy risks slightly lower than when they were considering sharing this photo without considering recipient selection. Therefore, we find some evidence that recipient selection does provide some sense of privacy protection. For example, some participants reported they would not want to share their children’s photos because strangers might identify them, in such cases they could select sharing with “friends” instead of “public.”. For the privacy warning condition, there is no difference in the willingness to share ( $p = .26$ ). For the obfuscation condition, participants are more willing to share after using the obfuscation system than before using it ( $p = .00$ ). However, as we speculate for the control condition, recipient selection might have also contributed to the differences in the obfuscation condition. Additionally, unlike perceived privacy risks, the difference between pre- and post-test scores in the obfuscation condition and the difference in the control condition are similar ( $p = .16$ ).

#### 7.4.4 Behavioral Data

Besides the subjective measurements I discussed above, we also collected behavioral data via the prototype. Specially, we recorded 1) which recipient group participants chose (Only me, Friends, Public) and 2) whether the participant clicked on the “Post to Facebook” button or “Cancel posting” button. To understand the effects of interface versions on these two behavioral measurements, I created two (generalized) linear mixed-effects models in which I used the two behavioral measurements as dependent variables and included experimental conditions as an independent variable.

Regarding the recipient variable, from Table 7.4, we can see that across all three conditions, the majority of the participants chose sharing with friends, while in the warning condition, more participants (17%) chose “only me” compared to the control (6%) and obfuscation conditions (10%). The significance test also confirmed the above statement. I re-coded “Only me,” “Friends,” “Public” to numeric values based on the level of sharing. The result indeed shows that when using the warning version, participants were less likely to share with more people compared to the control version ( $p =$

	% selected	N
No filter	25%	13
Blurring	36%	19
Masking	6%	3
Avatar	26%	14
Inpainting	8%	4

Table 7.5: Percentages of obfuscation options selected.

.047), while there is no difference between the obfuscation and the control versions ( $p = .33$ ).

In terms of whether participants chose to post to Facebook, Table 7.4 demonstrates that fewer people (81%) in the warning condition chose to post photos to Facebook compared to the other two conditions (both 88%), however from the test results, we did not find any differences between conditions.

Additionally, to understand whether recipient selection has an effect on the outcome of sharing, we constructed another generalized linear mixed-effects model where the dependent variable is whether participants posted photos and the independent variables are recipient and experimental conditions. There is no interaction effect found between recipient and conditions, and the effect of recipient is not significant either ( $p = .99$ ).

All participants' selection of obfuscation options was recorded as well. Table 7.5 shows the percentages of obfuscation options selected. For people who chose an obfuscation, the majority of chose blurring and avatar, which is in line with study two of this dissertation that showed that people consider blurring and avatar likable.

#### 7.4.5 Qualitative Data About What Made People Like Most and Least about the Systems

We collected participants' qualitative feedback about what they like most and least about the system. In this subsection, I discuss the identified trends for each interface version.

For the control version, since it mimics the Facebook without privacy features added, the most common category of responses about what participants like most about the system were that the system was **easy to use and intuitive**. Participants commented that "it is a light and fast system, very easy to understand" "I like that it was easy and straightforward. I didn't have any problems, nothing froze up, and I felt like it was easy to use." Additionally, there were a few participants said they liked the **photo sharing functionality** and **interface design**. While when

asked about what they like least, the majority of participants stated that they disliked nothing about the system. However, a few participants noted that the system **needs the ability to customize privacy setting** and the **aesthetics of interface design should be improved**.

In terms of the privacy warning version, participants reported that they liked the **sensitive content highlight feature**. For example, one participant stated “It pointed out risks for me. It made me aware of the little things I might not have noticed before.” The other two reasons are **ease of use** and **good interface design**. Even older adults could use it without efforts “it was simple enough to use and even this 73-year-old senior adult figured it out and was able to post the picture.” However, when thinking about what they like least, participants mentioned that they **need solutions to protect content in addition to warning**. For example, one participant reported “there weren’t any privacy settings. I felt like I had no control over anything. I knew that whatever I posted could and would be seen by others.” They also complained that it **takes extra time** to upload a photo, “I can see how some users could get annoyed with the little extra time it takes to use.” Two participants stated that they **do not feel the need for privacy warning** because “I would never attempt to post anything with questionable or risky content.” Notably, one participant raised concerns about **the service provider who could store their sensitive photos when analyzing them**.

Regarding the obfuscation version, participants appreciated **the effectiveness of obfuscation feature and available obfuscation options**. For example, one participant said “it automatically flags what might be sensitive content so it can apply the filter directly to the correct area of the photo. The available filter options are very effective.” They also considered the system **easy and intuitive** and **efficient**. Participants commented “I loved that it was done automatically. you didn’t have to go out of your way to edit the picture and spend all kinds of extra time to change it just to feel comfortable with uploading it.” However, they were **skeptical about sensitive content detection**. Participants stated “maybe it’s getting too intelligent detecting private things in your photos” “I’m not completely sure I trust a computer program to tell me where the sensitive information is.” They also felt that **the system was inflexible** and they expected “more avatars available, and that I can move the avatar by myself, or that I can change the size of the avatar.” Additionally, similar to the warning condition, participants complained that it is “**took a lot longer to post a picture** then I normally would.”

## 7.5 Discussion

### 7.5.1 The Obfuscation Version Performs the Best among All Three Conditions In Terms of Reducing Perceived Privacy Risks and Increasing Willingness to Share

The path model in Figure 7.11 indicates that among the three versions, people using the privacy warning version perceived more privacy risks than people using the control version, which in turn, reduced their willingness to share. Moreover, from Figure 7.12, we can clearly see that before using the privacy warning version, when the sensitive content in their photos was not highlighted, their perceived privacy risks were lower than after using the system; while once they were reminded of the sensitive content in their photos, their privacy risks increased. This result is in line with prior work on the relationship between content sensitivity, privacy concerns, and willingness to disclose information [17, 188, 229]. For example, Bansal et al. found that perceived health information sensitivity is positively related to privacy concern about disclosing health information and higher privacy concern is negatively related to the intention to disclose health information [17]. Our result further emphasizes that in the field of online photo privacy, this relationship holds true.

Both the obfuscation and warning conditions identified sensitive content. The only difference between the obfuscation and warning conditions is that the obfuscation condition provided obfuscation options. Therefore, the lower perceived privacy risks in the obfuscation condition can be attributed to the application of the obfuscation. [295] shows similar results that privacy-enhancing technology increases people's perceived control and in turn decreases privacy concerns. The comparison between the pre- and post-test for perceived privacy risks also demonstrates that obfuscation leads to a larger decrease in perceived privacy risks compared to the control condition. In terms of willingness to share, while another work find that people are more likely to share a photo when they received a warning that asked them to think about the privacy of the person in a photo [12], our results show that people are less willing to share photos after receiving the privacy warning in general.

## 7.5.2 Obfuscation Increases System Satisfaction

The results of marginal effects show that the obfuscation version outperforms the warning version in terms of system satisfaction. People consider it satisfying and they feel their privacy is protected when obfuscations are available as a privacy-preserving option. For example, from the qualitative feedback about why they liked the obfuscation system, participants reported “it was amazing,” and made, “me feel comfortable with the system and very secure” “the available filter options are very effective” “its different and unique compared to the older system.” Participants also qualitatively expressed a willingness to use it in the future “This is a very good quality system that I could see myself using in the future to better my Facebook experience and to protect the privacy of myself and family.”

In early studies on self-disclosure behavior in SNSs, researchers found that people disclose a greater amount of personal information than they intend to [19, 209]. This phenomenon is termed the privacy paradox and refers to the seeming inconsistency between privacy attitudes and privacy behavior. Lee et al. stated that both expected benefits and expected risks have effects on people’s intentions to share. Sometimes the numerous benefits, such as social validation, relationship development, self-presentation [163], and gaining social capital [83], motivate people to share more even at risks to themselves. But, people also used privacy-enhancing strategies to maximize benefit and minimize risk rather than passively accepting it (e.g., adjusting privacy settings) [163]. Other protection strategies people adopt on SNSs include deleting photos and tags [300], providing false information [198], and limiting access to their profiles [32]. While many of these strategies are not applicable to online photo sharing, obfuscation could be a mitigation to the tension between the sharing risk and benefit.

It is interesting that though the privacy warning version does provide some sort of privacy enhancement and increases people’s privacy awareness, due to the lack of protection methods, it appeared to be the least satisfying among the three conditions we studied (see Figure 7.7). Participants seemed to want or expect additional steps “the system identified the risky content by enclosing the identifying features of the subjects in a red box. Does the system then do something to that content? Remove it? Blur it? Log it somewhere?” “there wasn’t any privacy settings. I felt like I had no control over anything. I knew that whatever I posted could and would be seen by others” “what I like most could also be what I like least, too simple and lacking options. Maybe more

complexity would help users feel more safety with privacy.” This indicates that it is not enough for systems to point out privacy issues. For systems to be perceived as effective, they must offer users an easy-to-implement solution that reduces privacy concerns. Our finding is supported by prior work on the impact of privacy nudges on user behavior on Facebook [290]. In this study, researchers evaluated three types of privacy nudge including picture nudge which shows the potential audience for the post, timer nudge which gives users 10 seconds to cancel posting after post an update, and sentiment nudge which shows the potential attitude that viewers may have. Participants perceived privacy nudges effective and most of them changed their privacy settings or edited posts after receiving nudges, which indicates that, indeed, people prefer to take privacy-protective actions after identifying the privacy issues. Another work also pointed out that privacy warnings should not only provide information about data practices but also include control options which could make the information in privacy warning actionable and allow users to set their privacy preferences [243]. However, our warning version does not provide actionable further steps besides identifying sensitive content and people could only choose to cancel their posting. On the other hand, beyond identifying sensitive content, the obfuscation condition provides obfuscation options and shows the photo effect after applying obfuscations.

On the other hand, participants might feel the warning version more effective and satisfying if it could provide detailed information about the sensitive content. Many prior studies in HCI and Human Factors show that warnings that further explain the risks that the user faced and with the options presented result in higher perceived effectiveness and compliance (e.g., [196, 280]). For example, in our case, instead of showing “we’ve highlighted the content we think may be sensitive,” it could be “the highlighted content is potentially sexually inappropriate and it is highly not recommended to share on Facebook. We suggest you cancel posting.” Additionally, personalizing warning to the specific user such as including the user’s name in the notice also enhances effectiveness [77, 243].

### **7.5.3 The Effect of Interpersonal Privacy Concerns and Trust in Facebook**

Regarding the first characteristic variable – interpersonal privacy concerns, we found it has an indirect positive effect on perceived privacy risks and a direct positive effect on system satisfaction (Figure 7.11). In study four of this dissertation about self-censorship, we learn that people with higher levels of privacy consciousness about their personal content are more likely to censor photos.

It is similar to what we've found here – people who have more interpersonal privacy concerns (who might be more privacy-conscious) perceive more privacy risks over their personal content. Hence, they may be more likely to appreciate a privacy-enhancing system and consider it satisfying.

The second characteristic variable – trust in Facebook – has effects on system satisfaction and ease of use. Trust of a commercial site influences personal information disclosed to that site significantly [288]. Many parameters are in play with trust, for instance, users' regard for a company and perceived site privacy protection [197]. Though we do not find a relationship between trust in Facebook and perceived privacy risks, we do see the general effect of trust on willingness to share. Additionally, aligning with prior work on information system evaluation (e.g., [137, 143, 165]), we do see trust in Facebook is positively related to system satisfaction. The other two effects on the usability also indicate that people who trust in Facebook are more likely to look on the bright side of it. For example, even if both privacy warning and obfuscation warning required extra clicks, participants who had high trust still felt the obfuscation system “was easy and intuitive” and they “didn't have to go out of your way to edit the picture and spend all kinds of extra time to change it just to feel comfortable with uploading it”; and the warning system “was very easy and user-friendly.”

## 7.6 Limitation

A possible limitation of our study design is that participants' behavior might be biased by the pre-test questionnaire since they knew from the questions that it would be a study about photo privacy and they might pay special attention to privacy aspects of their photos.

## 7.7 Chapter Conclusion

To summarize, in this study we conducted a two-part experiment to determine whether obfuscation reduces privacy concerns and increases willingness to share while maintaining good usability. The primary contributions of this study are threefold.

First, we create interfaces that integrate obfuscation into SNS photo sharing feature.

Second, through an experiment, we understand that obfuscation does reduce users' privacy concerns about their photos and increase their willingness to share.

Third, though the proposed interfaces require users to perform extra tasks, it is still per-



ceived as usable.

## Chapter 8

# Discussion of all Five Studies of the Dissertation

Increasing awareness of online privacy increases the tension between photo sharing and privacy protection. People actively seek solutions to protect their privacy and self-censorship tends to be a prevalent strategy (see Chapter 6) though it leads to large sharing loss and goes against the nature of SNSs [255]. Researchers have built systems to address photo privacy issues using different approaches. Most of the systems fall into the category of recipient control (e.g., [95, 168, 307]). With recipient control, the entire photo is inaccessible to certain recipients. This creates a sharing loss. Some systems have begun to attempt content control (e.g., [121, 124, 251]). For example, one system obscures sensitive content and shows a partial photo. However, these systems either apply ineffective obfuscations such as blurring or only obscure limited types of sensitive content such as faces. Moreover, many of prior studies are focused on system building, but rarely take users' perceptions into consideration.

This dissertation focused on controlling sensitive content to protect photo privacy. Specifically, the aim of the dissertation was to understanding different aspects of content control – investigating sensitive content, identifying effective and usable obfuscations, understanding photo self-censorship, and studying whether obfuscation can reduce privacy concerns and increase willingness to share. In this chapter, I first describe the contributions of the five studies, then discuss the impacts on privacy research and industry, the considerations during implementation, and the potential

ethical issues of obfuscations.

## 8.1 Contributions

This dissertation consists of three parts – identifying sensitive content in photos, identifying effective and usable obfuscation, and investigating the ability of obfuscation to reduce privacy concerns and increase willingness to share. In the following paragraphs, I discuss the contributions of each part.

First, I created a human-centered taxonomy that describes what content is sensitive based on a much larger data set collected from a larger sample size than prior work. This taxonomy provides a more granular level of detail about sensitive content categories which may be more practical for privacy researchers, computer vision researchers, and practitioners. Additionally, the method we introduced and used for the first time overcomes the limitations of prior machine learning approaches. For example, many photo privacy protection systems were trained using a Flickr data set which is not suitable for photo privacy research (see Chapter Three).

Second, knowing what content is sensitive, we then identified obfuscations that are both effective and provide a good user experience. Obviously masking would be more effective than blurring, but considering the trade-off between effectiveness and utility, would people prefer masking? Are there any obfuscation methods that meet both needs? The second study answers most questions we had for obfuscations. We learned that avatar and inpainting appear to be effective against human recognition and are likable. From the third study, we know that the performance of these two obfuscation methods is equally good when applied to familiar and unfamiliar people in photos.

Third, self-censorship is a common strategy to avoid privacy leakage. Though researchers have explored self-censorship on SNSs in general, photos are a likely, but unexplored, area of self-censorship research. I wanted to understand whether obfuscation has the potential to reduce sharing loss due to self-censorship. Therefore in study four, we quantified the prevalence of self-reported photo self-censorship, and results show that over half of participants have self-censored photos on SNSs. Furthermore, we learned that obfuscations may be useful for helping users achieve photo sharing goals while maintaining privacy and reducing self-censorship. In the last study, we asked participants to make assumptions or predictions about how they would feel about photo sharing with obfuscations. For example, we asked “if you have access to obfuscations will you be willing

to share the previously censored photo?”. On the other hand, in study five, we created interfaces that allow participants to apply various obfuscations to their own photos. The results indicate that the obfuscation does have the ability to reduce people’s privacy concerns about their photos and increases the willingness to share. Though the system requires users to perform extra tasks (e.g., selecting obfuscation), its usability is as good as the original Facebook interface, which does not offer any content control; it only offers recipient control.

To summarize, the contributions of my dissertation are:

- Studying the two privacy parameters in the behavioral privacy model with the focus on the content parameter (e.g., obfuscation)
  - Investigating user-defined sensitive content in photos
  - Learning users’ sharing preferences with different recipients
  - Examining effective and usable obfuscations which can successfully de-identify both familiar and unfamiliar people in photos
  - Understanding how obfuscation might combat photo self-censorship
- Creating effective and usable photo privacy protection interfaces based on the knowledge learned in the series of studies above
  - Create interfaces
  - Evaluating the system’s ability to reduce privacy concerns, the ability to encourage photo sharing, and overall usability

## 8.2 Impact on Privacy Research

In terms of sensitive content, some work in machine learning has tried to identify and classify sensitive content (e.g., [303, 304]), but it has lots of problems (see Sensitive Content subsection in Chapter Two).

Machine learning approaches failed because they did not adopt a human-centered perspective. For example, researchers used a public Flickr dataset to elicit sensitive content. While the set was labeled as “private,” it did not contain users’ most sensitive photos. The label “private” might not mean the same as how people think of “private.” People understand that when they upload

photos, they are sharing them with the organization hosting the photos. Hence, they are likely to censor their most sensitive or “private” photos and not share on Flickr, which means that the photos in this dataset may not represent the most sensitive photos.

Moreover, researchers used a binary classification [304], “private” versus “public,” which we know does not fit with how people think about privacy. This work found that the outdoor or landscape images are generally public and indoor or images with people are private, which obviously is too broad a classification to be useful.

Our work could benefit privacy research which uses the machine learning approach in two ways. First, the photo-elicitation method that we introduced can be a way to supplement existing datasets or to create a new dataset of sensitive photos from scratch. For example, privacy researchers could gather and add new images that contain private content to existing general-purpose image datasets which could then make them useful for image privacy research tasks. Second, we provide a user-defined taxonomy of sensitive content that can be used to compare with and validate other ML generated sensitive content classifications.

In terms of obfuscations, blurring is the most widely adopted obfuscation method, while in my second and third studies, we know that blurring is ineffective in de-identifying both familiar and unfamiliar people. In prior work, researchers consider it as a default or even the only option when building a photo privacy protection system (e.g., [124, 166, 297]). Though the system design might be successful, the flawed obfuscation selection could lead to the failure of systems. Our work can benefit these systems by suggesting substitute obfuscations that are effective and usable – avatar and inpainting instead of blurring.

Our work in this area is already having an important impact. We published a CSCW paper in 2017 in which we identified effective and usable obfuscations. We are pleased to see that privacy researchers have cited this paper and discussed content control and obfuscation options (e.g., [7, 9, 35, 80, 156, 238]). However, some of these researchers only discuss it in the literature review section [9, 35, 80], rather than taking the suggestions to their designs. Moreover, when coming to choosing obfuscations, researchers still often chose an ineffective obfuscation - blurring [9, 80]. Besides the work that cited our paper, some other work about photo privacy protection published later than 2017 used blurring as well without explaining how researchers chose obfuscations and the reasons for choosing blurring (e.g., [138]). On the other hand, a recent study proposed morphing as a privacy protection tool and compared it with blurring and pixelating [172]. However, from the

third study, we know that though morphing is relatively effective, it is not likable as people are skeptical about its concept and it might lead to ethical issues.

The selection of obfuscations limits these systems' practicability when applied in reality. It's possible that the barriers might be related to obfuscation implementation. For example, the implementation of blurring is mature. There are numerous algorithms or applications that can create blurring effects with little human effort. However, avatar and inpainting are not as accessible as blurring. We see privacy researchers are at least discussing obfuscation methods before building systems which we believe is a good starting point. In the future, we hope to see more computer vision researchers understand the importance of effective and usable obfuscations and work on the implementation of them, which could lower the barriers of integrating them into photo privacy protection systems for other privacy researchers.

### 8.3 Impact on Industry

Obfuscation selection is also problematic in the industry. Signal is considered the most secure, privacy-centric messaging application compared to other mainstreaming messaging applications [252]. All messages are end-to-end encrypted and neither Signal nor anyone can read users' messages or listen to their calls. It also hides all of the metadata, including who sent the message. In brief, its selling point is privacy and security. However, we found that Signal uses blurring which is ineffective and easy to be re-identified by both humans and machines. Furthermore, Signal only blurs faces in photos; in our second study, we demonstrated that face obfuscation is generally less effective than body obfuscation. Similarly, another privacy-centric application – Anonymous Camera – allows users to seamlessly “anonymize” photos and videos [13]. All processing is done in real-time, locally on device without uploading to the cloud. However, it again uses face blurring which is ineffective. The other obfuscation option it provides is silhouette which we know from the first study is not likable. Users may trust the ability of these applications to protect their privacy based on their positive publicity, hence these apps could mislead users to share photos with sensitive content, in turn, compromise users' privacy, which goes against the mission of these privacy applications. My dissertation could provide guidance on privacy application design from identifying the sensitive content to choosing effective and usable obfuscations. For managers or designers in the industry, it is important that every design decision that they make should be supported by research evidence.

On the other hand, self-censorship on SNSs has been extensively investigated. Prior work has focused on political self-censorship [120, 157] and post self-censorship [60, 255], while we conducted the first study that investigated photo self-censorship and also provided solutions to combat it. We found that over half of the participants have censored photos due to privacy concerns which might have caused big sharing loss to SNSs and reduced SNSs’ sociability. It should alarm SNSs such as Facebook to put more effort into the photo privacy protection features, which might help increase users’ engagement and their trust in SNSs.

## 8.4 Practical Considerations for System Implementation

Though my thesis is not focused on system building, it is worth mentioning some considerations for system implementation. How to detect each piece of sensitive content in our taxonomy is still an open question. Some categories can be easily addressed. For example, regarding our category “other people,” Hasan et al. created an automated system to identify bystander (vs. subject) in photos [108]. For “online account and password” and “video game” that is shown on screens, they can be identified using the system proposed in [145]. For some common objects in our taxonomy such as “toilet” and “necklace,” Microsoft COCO, which is a large-scale object detection dataset, could be used to detect them [182]. However, there are some complicated categories that are hard to detect. For example, while “naked child” is easy to detect, “child in inappropriate clothes” is very subjective as users may have their own opinions on inappropriate clothes. Another example is “other people’s information.” People’s own to-do list is not sensitive but they consider their friends’ to-do list very sensitive. In such a case, it is difficult to automatically determine whether an object belongs to photo owners or other people. Furthermore, in the last study, one participant expressed concerns that the system “is getting too intelligent detecting private things in your photos.” The other participant said he/she should be able to “move the avatar by myself, or that I can change the size of the avatar.” The qualitative data indicate that people might not want automatic detection. We believe that in the future, with the advances in technology, object detection would not be a barrier anymore and the system can be smart enough to perform fully automated detection. Yet for now, full automation is not very feasible. Hence, using sensitive content taxonomy as suggestions for users to identify sensitive content in their photos might be better than automatic detection at the current stage.

On the other hand, one participant in the last study concerned that “I’m not completely sure I trust a computer program to tell me where the sensitive information is.” Indeed, though artificial intelligence (AI) approaches show great practical success in many domains, its decision making is still in a “BlackBox, ” and people consider it not trustworthy, hence explainable AI is necessary [57, 73]. In our system, we need to provide sufficient information that is interpretable and comprehensible to draw an explanation of why a particular decision is made [73]. Once the system highlights a piece of sensitive content, it should offer explanations of why the system considers it sensitive. For example, “this photo shows you are holding a Bud Light beer can in an office setting. Your workplace may have their own policy on the consumption of alcohol and posting this photo may have negative effects on your career.”

Regarding the implementation of obfuscation, avatar is more accessible than inpainting as it just adds content to images. Inpainting will require more sophisticated image reconstruction techniques to recompose an image after removing the sensitive content [58, 104]. Current commercial applications only have limited functions for inpainting, such as Photoshop’s “content aware patch,” Snapseed’s “Expand[276],” Byebye Camera’s “people removal [46]” features. However, the generated photos are unnatural and appear obviously to have been altered. With the development of new computer vision techniques, the implementation of inpainting is likely to be addressed in the near future. A recent study presents a learning-based method for seamlessly removing obstructions such as windows, fences, and raindrops and recovering clean images [185]. This technique could be used to inpaint sensitive content.

## 8.5 Ethical Issues Related to Obfuscation

Knowing the impacts of my dissertation, the potential ethical issues should not be ignored. Photo manipulation could be risky. Imma, a fashion model from Japan, is very popular on Instagram. She posts her styles daily and has almost 200k followers [125]. However, she is computer generated and completely virtual. Her photos are extremely realistic and her facial features are perfectly rendered. Since people are bad at detecting and locating manipulations within images [206], people without prior knowledge are unlikely to know she is fake. After knowing the truth, people’s attitudes were mostly positive. They were surprised, but they thought it was fun and entertaining. Indeed, on Instagram, no matter whether it is an actual fashion model or virtual person, the use of an entirely



virtual character is unlikely to have an important impact on people’s lives. However, when such AI techniques are applied to people themselves, they have privacy concerns. For example, Deepfakes is a deep-learning system that can produce realistic fake videos by studying photos and videos of a target person from multiple angles and then mimicking its behavior and speech patterns. Deepfakes has been used to generate fake politician videos or celebrities’ pornography. However, Deepfakes is open-sourced and everyone can use it to generate videos of anyone, for example, having porn revenge by swapping porn performers’ faces with ex-girlfriends’ faces [115]. As people are getting more exposure to such video or photo manipulation technologies and understanding these technologies could harm their privacy, they raise concerns around the authenticity and ethic of manipulated photos. For example, participants in our second study stated that “[obfuscations] still provides context that could be misconstrued,” “the photos look disingenuous and photoshopped.” We can clearly tell that people were worried that viewers might misinterpret their obfuscated photos and the photos might look deceptive. These concerns might be roadblocks adoption.

Especially for certain types of obfuscations, the ethical issues are more prominent. Morphing may mislead the viewers into mistaking a morphed person for somebody else, and avatar has the same problem. In terms of inpainting, if a person or an object is removed, the meaning of a photo may dramatically change. For example, a party photo with multiple people may look like a dating photo if only two people remain after other people are removed via inpainting. Though participants in our studies reported liking avatar and inpainting, it is still unknown whether the ethical issue will impede users’ adoption of obfuscation in the long run. A possible solution could be adding indicators in inpainted photos to show that there was somebody or something removed; or in morphed photos, an indicator could inform viewers that the person is being morphed for privacy reasons. In study two, we evaluated “body bar” and “body point-light” obfuscation which are types of indicators, however, they were shown to be not likable, perhaps because they affect the aesthetics and reduce perceived photo quality. Prior work on video surveillance privacy protection introduced dot as an indicator [305] which might be a better option as it does not disturb nor provide any additional information to identify the target. In future studies, researchers should evaluate different variations of indicators and explore approaches to alleviate ethical issues.

# Appendices

## Appendix A Study 1 Online Experiment Questions

### A.1 Demographics

- What is your gender?
  - Male
  - Female
  - Other
  - I prefer not to answer
  
- What is your age?
  - 18-24 years old
  - 25-34 years old
  - 35-44 years old
  - 45-54 years old
  - 55+
  
- What is your ethnicity?
  - White
  - Hispanic or Latino
  - Black or African American
  - Native American or American Indian
  - Asian / Pacific Islander
  - Other
  
- Which of the following best describes your current employer?
  - Government
  - Educational institution
  - Business or industry

- Non-profit organization
  - Other
- What is the highest level of school you have completed or the highest degree you have received?
  - High school incomplete or less
  - High school graduate or GED (includes technical/vocational training that doesn't count towards college credit)
  - Some college (some community college, associate's degree)
  - Four year college degree/bachelor's degree
  - Some postgraduate or professional schooling, no postgraduate degree
  - Postgraduate or professional degree, including master's, doctorate, medical or law degree
  - I prefer not to answer
- Which of these best describes you?
  - Married
  - Living with a partner
  - Divorced
  - Separated
  - Widowed
  - Never been married
  - I prefer not to answer
- About how often do you use Internet either on a computer or on a mobile device like a smartphone or a tablet?
  - Most of the day
  - Several times a day
  - About once a day
  - A few times a week

- A few times a month
  - A few times a year
  - Never
  - I prefer not to answer
- About how often do you visit social media sites such as Facebook, Twitter or LinkedIn?
  - Most of the day
  - Several times a day
  - About once a day
  - A few times a week
  - A few times a month
  - A few times a year
  - Never
  - I prefer not to answer
- About how often do you upload photos to social media sites such as Facebook, Twitter or LinkedIn?
  - Many times a day
  - Several times a day
  - About once a day
  - A few times a week
  - A few times a month
  - A few times a year
  - Never
  - I prefer not to answer

## A.2 Collect Photos That People Do Not Want to Share with Anyone

- Please take out your phone. If you have an iPhone, please go to “Photos” and look at the “Camera Roll” album. If you have an Android phone, please look at the “Gallery.” Approximately how many photos do you have in this album? [open-ended question]
- Are there any photos on your phone that you consider private (those you do not want to share with anyone)?
  - Yes
  - No
- Please look through the photos on your phone and find one of your most private photos.
  - Okay, found one!
- Are you willing to share it with us (the researchers conducting this study)? As a reminder, we won’t share it with anyone.
  - Yes
  - No

If Yes Please upload the photo. [upload button]

If No That’s okay, we understand. But, we’re trying to understand the kinds of photos people consider private. Please tell us about the photo and/or find a photo online which has similar sensitive content. [open-ended question]

- Please upload the photo you found online that has similar sensitive content to your photo. [upload button]
- What content in this photo do you consider sensitive? [open-ended question]
- How likely are you to keep this photo private, meaning not share it with anyone?
  - 1 - Very unlikely
  - 2
  - 3

- 4
  - 5
  - 6
  - 7 - Very likely
  
- How likely are you to share this photo with your significant others (spouse / girlfriend / boyfriend)?
  - 1 - Very unlikely
  - 2
  - 3
  - 4
  - 5
  - 6
  - 7 - Very likely
  - I do not have significant others
  
- How likely are you to share this photo with household members?
  - 1 - Very unlikely
  - 2
  - 3
  - 4
  - 5
  - 6
  - 7 - Very likely
  
- How likely are you to share this photo with relatives who are close with you?
  - 1 - Very unlikely
  - 2

- 3
  - 4
  - 5
  - 6
  - 7 - Very likely
- How likely are you to share this photo with relatives who are NOT close with you?
    - 1 - Very unlikely
    - 2
    - 3
    - 4
    - 5
    - 6
    - 7 - Very likely
- How likely are you to share this photo with friends who are close with you?
    - 1 - Very unlikely
    - 2
    - 3
    - 4
    - 5
    - 6
    - 7 - Very likely
- How likely are you to share this photo with friends who are NOT close with you?
    - 1 - Very unlikely
    - 2
    - 3



- 4
  - 5
  - 6
  - 7 – Very likely
- How likely are you to share this photo with ex-girl/boyfriends?
    - 1 - Very unlikely
    - 2
    - 3
    - 4
    - 5
    - 6
    - 7 – Very likely
    - I do not have ex-girl/boyfriends
- How likely are you to share this photo with colleagues/classmates who are close with you?
    - 1 - Very unlikely
    - 2
    - 3
    - 4
    - 5
    - 6
    - 7 – Very likely
    - I do not have colleagues/classmates who are close with me
- How likely are you to share this photo with colleagues/classmates who are NOT close with you?
    - 1 - Very unlikely

- 2
  - 3
  - 4
  - 5
  - 6
  - 7 – Very likely
  - I do not have colleagues/classmates who are NOT close with me
- How likely are you to share this photo with your supervisor who is close with you?
    - 1 - Very unlikely
    - 2
    - 3
    - 4
    - 5
    - 6
    - 7 – Very likely
    - I do not have a supervisor who is close with me
- How likely are you to share this photo with your supervisor who is NOT close with you?
    - 1 - Very unlikely
    - 2
    - 3
    - 4
    - 5
    - 6
    - 7 – Very likely
    - I do not have a supervisor who is NOT close with me

- How likely are you to share this photo with friends of friends?
  - 1 - Very unlikely
  - 2
  - 3
  - 4
  - 5
  - 6
  - 7 – Very likely
  
- How likely are you to share this photo with people you've only met online?
  - 1 - Very unlikely
  - 2
  - 3
  - 4
  - 5
  - 6
  - 7 – Very likely
  
- How likely are you to share this photo with people you've only met once or twice?
  - 1 - Very unlikely
  - 2
  - 3
  - 4
  - 5
  - 6
  - 7 – Very likely
  
- How likely are you to share this photo with people of your age?

- 1 - Very unlikely
  - 2
  - 3
  - 4
  - 5
  - 6
  - 7 - Very likely
- How likely are you to share this photo with people younger than you?
    - 1 - Very unlikely
    - 2
    - 3
    - 4
    - 5
    - 6
    - 7 - Very likely
- How likely are you to share this photo with people older than you?
    - 1 - Very unlikely
    - 2
    - 3
    - 4
    - 5
    - 6
    - 7 - Very likely
- How likely are you to share this photo with people of the same gender as you?
    - 1 - Very unlikely

- 2
  - 3
  - 4
  - 5
  - 6
  - 7 – Very likely
- How likely are you to share this photo with people of different gender?
    - 1 - Very unlikely
    - 2
    - 3
    - 4
    - 5
    - 6
    - 7 – Very likely

### **A.3 Collect Photos That People Do Not Want to Share with Their Family**

The questions in this section are the same as the ones in the first photo collection section. The only difference is that at the beginning we asked participants to find one photo that they do NOT want to share with at least one of the following groups: significant others, household member, close relatives, relatives who are NOT close with you.

### **A.4 Collect Photos That People Do Not Want to Share with Friends**

The questions in this section are the same as the ones in the first photo collection section. The only difference is that at the beginning we asked participants to find one photo that they do NOT want to share with at least one of the following groups: close friends, friends who are NOT close with you, ex-girl/boyfriends.

## **A.5 Collect Photos That People Do Not Want to Share with Colleagues / Classmates / Supervisors**

The questions in this section are the same as the ones in the first photo collection section. The only difference is that at the beginning we asked participants to find one photo that they do NOT want to share with at least one of the following groups: close colleagues/classmates, colleagues/classmates who are NOT close with you, close supervisors, supervisors who are NOT close with you.

## **A.6 Collect Photos That People Do Not Want to Share with Acquaintances**

The questions in this section are the same as the ones in the first photo collection section. The only difference is that at the beginning we asked participants to find one photo that they do NOT want to share with at least one of the following groups: friends of friends, people you've only met online, people you've only met once or twice.

## Appendix B Study 2 Online Experiment Questions

### B.1 Demographics

- What is your gender?
  - Male
  - Female
  - Other
  - I prefer not to answer
  
- What is your age?
  - 18-24 years old
  - 25-34 years old
  - 35-44 years old
  - 45-54 years old
  - 55+
  
- What is your ethnicity?
  - White
  - Hispanic or Latino
  - Black or African American
  - Native American or American Indian
  - Asian / Pacific Islander
  - Other
  
- Which of the following best describes your current employer?
  - Government
  - Educational institution
  - Business or industry

- Non-profit organization
  - Other
- What is the highest level of school you have completed or the highest degree you have received?
  - High school incomplete or less
  - High school graduate or GED (includes technical/vocational training that doesn't count towards college credit)
  - Some college (some community college, associate's degree)
  - Four year college degree/bachelor's degree
  - Some postgraduate or professional schooling, no postgraduate degree
  - Postgraduate or professional degree, including master's, doctorate, medical or law degree
  - I prefer not to answer
- Which of these best describes you?
  - Married
  - Living with a partner
  - Divorced
  - Separated
  - Widowed
  - Never been married
  - I prefer not to answer
- About how often do you use Internet either on a computer or on a mobile device like a smartphone or a tablet?
  - Most of the day
  - Several times a day
  - About once a day
  - A few times a week



- A few times a month
  - A few times a year
  - Never
  - I prefer not to answer
- About how often do you visit social media sites such as Facebook, Twitter or LinkedIn?
    - Most of the day
    - Several times a day
    - About once a day
    - A few times a week
    - A few times a month
    - A few times a year
    - Never
    - I prefer not to answer

## B.2 Browser Testing

Next two screens will be the browser resizing test.

On the first screen, you need to see if you can view all of the photos, the confidence level question, and the red next page button on your screen at the same time without scrolling. At this point, you do NOT need to answer the two questions.

On the second screen, you will answer the question "Can you view all of the photos, the confidence level question, and the red next page button on your screen at the same time without scrolling on last screen?"

- Please identify the person indicated by the orange arrow. [An example photo was shown. The face of the target person was pixelated.]
  - First ID photo was shown
  - Second ID photo was shown
  - Third ID photo was shown

- None of above [note that the four choices' order is randomized]
- How confident do you feel that you correctly identified the person?
  - Completely unconfident
  - Mostly unconfident
  - Somewhat unconfident
  - Neither unconfident nor confident
  - Somewhat confident
  - Mostly confident
  - Completely confident
- Can you view all of the photos, the confidence level question, and the red next page button on your screen at the same time without scrolling on last screen?
  - Yes
  - No

Next screen is a question example for the browser adjusting. Please zoom your browser to a point where you can see all four photos, the confidence level question, and the red next page button at the same time on your screen.

For Windows users, please press CTRL + PLUS SIGN (+) to zoom in; press CTRL + MINUS SIGN (-) to zoom out. For Mac users, please press COMMAND + PLUS SIGN (+) to zoom in; press COMMAND + MINUS SIGN (-) to zoom out.

After adjusting your browser, please go to next page. At this point, you do NOT need to answer the two questions.

- Please identify the person indicated by the orange arrow. [An example photo was shown. The face of the target person was pixelated.]
  - First ID photo was shown
  - Second ID photo was shown
  - Third ID photo was shown

- None of above [note that the four choices’ order is randomized]
- How confident do you feel that you correctly identified the person?
  - Completely unconfident
  - Mostly unconfident
  - Somewhat unconfident
  - Neither unconfident nor confident
  - Somewhat confident
  - Mostly confident
  - Completely confident

### **B.3 Fourteen obfuscation conditions’ examples**

On next screen, you will see 14 privacy filters. The orange arrow in each photo points to the filter effect area. Please read and understand the description for each privacy filter before going on to the next page.

[Fourteen obfuscation conditions’ examples were shown.]

### **B.4 Training**

In this section, we will teach you about the task you will perform. First, on the next screen, you will see a photo with an orange arrow pointing to an area (person) in a photo. Your job is to identify the person the orange arrow is pointing at.

There are four possible answers. Possible answers include images of three people, and “none of above.” Your job is to identify the person in the obscured photo (pointed out by the orange arrow). If you do not see the person pointed out by the orange arrow, please choose ‘none of above.’

You may be able to identify the person easily. Or, it may be difficult or impossible to identify the person. Either way, please try your best, and choose only one answer.

After you have answered, you will then tell us how confident you are that your answer was correct.

- Please identify the person indicated by the orange arrow. [An example photo was shown. The face of the target person was pixelated.]

- First ID photo was shown
  - Second ID photo was shown
  - Third ID photo was shown
  - None of above [note that the four choices' order is randomized]
- How confident do you feel that you correctly identified the person?
    - Completely unconfident
    - Mostly unconfident
    - Somewhat unconfident
    - Neither unconfident nor confident
    - Somewhat confident
    - Mostly confident
    - Completely confident

[If participants' answers were correct:]

Great job! You correctly identified the person. Notice that in the actual test later, you will not know if your choice is correct or wrong. Next, you will do one more identification training.

[If participants' answers were wrong:]

Your identification is not correct. The correct choice is below: [show the correct ID photo] Notice that in the formal test, you will not know if your choice is correct or wrong. Next, you will do one more identification training.

- Please identify the person indicated by the orange arrow. [An example photo was shown. The face of the target person was blurred.]
  - First ID photo was shown
  - Second ID photo was shown
  - Third ID photo was shown
  - None of above [note that the four choices' order is randomized]
- How confident do you feel that you correctly identified the person?

- Completely unconfident
- Mostly unconfident
- Somewhat unconfident
- Neither unconfident nor confident
- Somewhat confident
- Mostly confident
- Completely confident

[If participants' answers were correct:]

Great job! Your choice is correct. Congratulations! You finished the training. Please go to the next page, and begin the test.

[If participants' answers were wrong:]

The correct choice is “None of above.” Congratulations! You finished the training. Please go to the next page to begin the test.

## B.5 Pre-trials

- Please identify the person indicated by the orange arrow. [An example photo was shown. No obfuscation was applied.]
  - First ID photo was shown
  - Second ID photo was shown
  - Third ID photo was shown
  - None of above [note that the four choices' order is randomized]

Considering the filter used in this photo, please rate the photo on the four statements below.

- The photo is satisfying.
  - Strongly disagree
  - Disagree
  - Somewhat disagree

- Neither agree nor disagree
  - Somewhat agree
  - Agree
  - Strongly agree
- This photo provides sufficient information.
  - Strongly disagree
  - Disagree
  - Somewhat disagree
  - Neither agree nor disagree
  - Somewhat agree
  - Agree
  - Strongly agree
- I enjoy the photo at this moment.
  - Strongly disagree
  - Disagree
  - Somewhat disagree
  - Neither agree nor disagree
  - Somewhat agree
  - Agree
  - Strongly agree
- There was a sense of human contact when I saw the photo.
  - Strongly disagree
  - Disagree
  - Somewhat disagree
  - Neither agree nor disagree

- Somewhat agree
- Agree
- Strongly agree

[Afterward, participants went through the second pre-trial which included the same set of questions and was very similar to the first one above.]

## **B.6 Actual Testing: 14 Trials**

[The actual testing includes 14 trials and participants saw all 14 obfuscation conditions.]

- Please identify the person indicated by the orange arrow. [An example photo was shown. One of the 14 obfuscations was applied.]
  - First ID photo was shown
  - Second ID photo was shown
  - Third ID photo was shown
  - None of above [note that the four choices' order is randomized]
  
- How confident do you feel that you correctly identified the person?
  - Completely unconfident
  - Mostly unconfident
  - Somewhat unconfident
  - Neither unconfident nor confident
  - Somewhat confident
  - Mostly confident
  - Completely confident

Considering the filter used in this photo, please rate the photo on the four statements below.

- The photo is satisfying.
  - Strongly disagree

- Disagree
  - Somewhat disagree
  - Neither agree nor disagree
  - Somewhat agree
  - Agree
  - Strongly agree
- This photo provides sufficient information.
    - Strongly disagree
    - Disagree
    - Somewhat disagree
    - Neither agree nor disagree
    - Somewhat agree
    - Agree
    - Strongly agree
- I enjoy the photo at this moment.
    - Strongly disagree
    - Disagree
    - Somewhat disagree
    - Neither agree nor disagree
    - Somewhat agree
    - Agree
    - Strongly agree
- There was a sense of human contact when I saw the photo.
    - Strongly disagree
    - Disagree



- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Agree
- Strongly agree

[Participants then went through the other 13 trials.]

## B.7 Rating the Likability of Obfuscations

Imagine that online social networks (Facebook etc.) adopted privacy filters so that users could better manage the privacy of their photos. In that case, which privacy filter would you prefer? Please rate your preference for each privacy filter.

- I like the “blurring” privacy filter. [An corresponding example obfuscation photo was shown.]
  - Strongly disagree
  - Disagree
  - Somewhat disagree
  - Neither agree nor disagree
  - Somewhat agree
  - Agree
  - Strongly agree

[Participants then rated the other 13 obfuscations using the same scale as above shows.]

## B.8 Follow-up Questions

- If you could use any of the privacy filters for photos you post on online social networks, which one, if any, would you like to use?
  - As is (no filter)
  - Face blurring

- Face pixelating
- Face silhouette
- Face avatar
- Face masking
- Body blurring
- Body pixelating
- Body silhouette
- Body avatar
- Body point-light
- Body masking
- Body bar
- Body inpainting
- Please tell us the reason [open-ended]
- I am willing to upload photos to online social networks using the filter I selected.
  - Strongly disagree
  - Disagree
  - Somewhat disagree
  - Neither agree nor disagree
  - Somewhat agree
  - Agree
  - Strongly agree
- Have you ever declined to upload a photo to an online social network for privacy reasons?
  - Yes
  - No
  - I don't know

- In the last question, you said you had declined to upload a photo to an online social network for privacy reasons. If you had access to one of the privacy filters here, would you be willing to upload the photo using one of the filters?
  - Yes
  - No
  - I don't know
  
- If you answered "Yes" in above question, which privacy filter would you prefer to use? If you answered "No" or "I don't know" , please select "NA".
  - As is (no filter)
  - Face blurring
  - Face pixelating
  - Face silhouette
  - Face avatar
  - Face masking
  - Body blurring
  - Body pixelating
  - Body silhouette
  - Body avatar
  - Body point-light
  - Body masking
  - Body bar
  - Body inpainting
  
- Please tell us your reasons why you chose this filter.

## B.9 Adoption Willingness

Please rate two statements about your adoption willingness.

- I want online social networks (Facebook etc.) to adopt these privacy filters SO that I can be obscured in certain photos my friends upload (group photo etc.).
  - Strongly disagree
  - Disagree
  - Somewhat disagree
  - Neither agree nor disagree
  - Somewhat agree
  - Agree
  - Strongly agree
- Please tell us the reason [open-ended]
- I want online social networks (Facebook etc.) to adopt these privacy filters SO that some people can be obscured in certain photos I view (group photo etc.).
  - Strongly disagree
  - Disagree
  - Somewhat disagree
  - Neither agree nor disagree
  - Somewhat agree
  - Agree
  - Strongly agree
- Please tell us the reason [open-ended]

## B.10 Attitudes Towards Privacy and Security

Privacy means different things to different people today. In thinking about all of your daily interactions - both online and offline - please tell us how important each of the following are to you:

- Being in control of who can get information about you.
  - Not at all important

- Not very important
  - Somewhat important
  - Very important
  - Don't know
- Being able to share confidential matters with someone you trust.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know
- Not having someone watch you or listen to you without your permission.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know
- Controlling what information is collected about you.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know
- Not being disturbed at home.
  - Not at all important
  - Not very important

- Somewhat important
  - Very important
  - Don't know
- Being in control of who can get information about you.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know
- Being able to have times when you are completely alone, away from anyone else.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know
- Having individuals in social / work situations not ask you things that are highly personal.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know
- Being able to go around in public without always being identified.
  - Not at all important
  - Not very important
  - Somewhat important

- Very important
  - Don't know
- Not being monitored at work.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know

## **B.11 Additional Feedback**

- Do you have any comments, confusions, and suggestions about this survey? [open-ended]

## Appendix C Study 3 and 4 Online Experience Questions

[Please note that the results of Study Three and Four came from the same experiment.]

### C.1 Demographics

- What is your gender?
  - Male
  - Female
  - Other
  - I prefer not to answer
  
- What is your age?
  - 18-24 years old
  - 25-34 years old
  - 35-44 years old
  - 45-54 years old
  - 55+
  
- What is your ethnicity?
  - White
  - Hispanic or Latino
  - Black or African American
  - Native American or American Indian
  - Asian / Pacific Islander
  - Other
  
- Which of the following best describes your current employer?
  - Government
  - Educational institution



- Business or industry
  - Non-profit organization
  - Other
- What is the highest level of school you have completed or the highest degree you have received?
    - High school incomplete or less
    - High school graduate or GED (includes technical/vocational training that doesn't count towards college credit)
    - Some college (some community college, associate's degree)
    - Four year college degree/bachelor's degree
    - Some postgraduate or professional schooling, no postgraduate degree
    - Postgraduate or professional degree, including master's, doctorate, medical or law degree
    - I prefer not to answer
- Which of these best describes you?
    - Married
    - Living with a partner
    - Divorced
    - Separated
    - Widowed
    - Never been married
    - I prefer not to answer
- About how often do you use Internet either on a computer or on a mobile device like a smartphone or a tablet?
    - Most of the day
    - Several times a day
    - About once a day

- A few times a week
  - A few times a month
  - A few times a year
  - Never
  - I prefer not to answer
- About how often do you visit social media sites such as Facebook, Twitter or LinkedIn?
    - Most of the day
    - Several times a day
    - About once a day
    - A few times a week
    - A few times a month
    - A few times a year
    - Never
    - I prefer not to answer

## C.2 Browser Testing

Next two screens will be the browser resizing test.

On the first screen, you need to see if you can view all of the photos, the confidence level question, and the red next page button on your screen at the same time without scrolling. At this point, you do NOT need to answer the two questions.

On the second screen, you will answer the question "Can you view all of the photos, the confidence level question, and the red next page button on your screen at the same time without scrolling on last screen?"

- Please identify the person indicated by the orange arrow. [An example photo was shown. The face of the target person was pixelated.]
  - First ID photo was shown
  - Second ID photo was shown

- Third ID photo was shown
- None of above [note that the four choices' order is randomized]
- How confident do you feel that you correctly identified the person?
  - Completely unconfident
  - Mostly unconfident
  - Somewhat unconfident
  - Neither unconfident nor confident
  - Somewhat confident
  - Mostly confident
  - Completely confident
- Can you view all of the photos, the confidence level question, and the red next page button on your screen at the same time without scrolling on last screen?
  - Yes
  - No

Next screen is a question example for the browser adjusting. Please zoom your browser to a point where you can see all four photos, the confidence level question, and the red next page button at the same time on your screen.

For Windows users, please press CTRL + PLUS SIGN (+) to zoom in; press CTRL + MINUS SIGN (-) to zoom out. For Mac users, please press COMMAND + PLUS SIGN (+) to zoom in; press COMMAND + MINUS SIGN (-) to zoom out.

After adjusting your browser, please go to next page. At this point, you do NOT need to answer the two questions.

- Please identify the person indicated by the orange arrow. [An example photo was shown. The face of the target person was pixelated.]
  - First ID photo was shown
  - Second ID photo was shown

- Third ID photo was shown
- None of above [note that the four choices' order is randomized]
- How confident do you feel that you correctly identified the person?
  - Completely unconfident
  - Mostly unconfident
  - Somewhat unconfident
  - Neither unconfident nor confident
  - Somewhat confident
  - Mostly confident
  - Completely confident

### C.3 Seven Obfuscation Conditions' Examples

On next screen, you will see 7 privacy filters. The orange arrow in each photo points to the filter effect area. Please read and understand the description for each privacy filter before going on to the next page.

[Seven obfuscation conditions' examples were shown.]

### C.4 Training

In this section, we will teach you about the task you will perform. First, on the next screen, you will see a photo with an orange arrow pointing to an area (person) in a photo. Your job is to identify the person the orange arrow is pointing at.

There are four possible answers. Possible answers include images of three people, and “none of above.” Your job is to identify the person in the obscured photo (pointed out by the orange arrow). If you do not see the person pointed out by the orange arrow, please choose ‘none of above.’

You may be able to identify the person easily. Or, it may be difficult or impossible to identify the person. Either way, please try your best, and choose only one answer.

After you have answered, you will then tell us how confident you are that your answer was correct.

- Please identify the person indicated by the orange arrow. [An example photo was shown. The face of the target person was pixelated.]
  - First ID photo was shown
  - Second ID photo was shown
  - Third ID photo was shown
  - None of above [note that the four choices' order is randomized]
  
- How confident do you feel that you correctly identified the person?
  - Completely unconfident
  - Mostly unconfident
  - Somewhat unconfident
  - Neither unconfident nor confident
  - Somewhat confident
  - Mostly confident
  - Completely confident

[If participants' answers were correct:]

Great job! You correctly identified the person. Notice that in the actual test later, you will not know if your choice is correct or wrong. Next, you will do one more identification training.

[If participants' answers were wrong:]

Your identification is not correct. The correct choice is below: [show the correct ID photo] Notice that in the formal test, you will not know if your choice is correct or wrong. Next, you will do one more identification training.

- Please identify the person indicated by the orange arrow. [An example photo was shown. The face of the target person was blurred.]
  - First ID photo was shown
  - Second ID photo was shown
  - Third ID photo was shown

- None of above [note that the four choices' order is randomized]
- How confident do you feel that you correctly identified the person?
  - Completely unconfident
  - Mostly unconfident
  - Somewhat unconfident
  - Neither unconfident nor confident
  - Somewhat confident
  - Mostly confident
  - Completely confident

[If participants' answers were correct:]

Great job! Your choice is correct. Congratulations! You finished the training. Please go to the next page, and begin the test.

[If participants' answers were wrong:]

The correct choice is “None of above.” Congratulations! You finished the training. Please go to the next page to begin the test.

## C.5 Pre-trials

- Please identify the person indicated by the orange arrow. [An example photo was shown. No obfuscation was applied.]
  - First ID photo was shown
  - Second ID photo was shown
  - Third ID photo was shown
  - None of above [note that the four choices' order is randomized]

Considering the filter used in this photo, please rate the photo on the four statements below.

- The photo is satisfying.
  - Strongly disagree

- Disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Agree
- Strongly agree
  
- This photo provides sufficient information.
  - Strongly disagree
  - Disagree
  - Somewhat disagree
  - Neither agree nor disagree
  - Somewhat agree
  - Agree
  - Strongly agree
  
- I enjoy the photo at this moment.
  - Strongly disagree
  - Disagree
  - Somewhat disagree
  - Neither agree nor disagree
  - Somewhat agree
  - Agree
  - Strongly agree
  
- There was a sense of human contact when I saw the photo.
  - Strongly disagree
  - Disagree

- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Agree
- Strongly agree

[Afterward, participants went through the second pre-trial which included the same set of questions and was very similar to the first one above.]

## C.6 Actual Testing: 14 Trials

[The actual testing includes 14 trials and participants saw all seven obfuscation conditions applied to both familiar and unfamiliar targets. They saw 7 familiar and 7 unfamiliar targets.]

- Please identify the person indicated by the orange arrow. [An example photo was shown. One of the obfuscations was applied.]
  - First ID photo was shown
  - Second ID photo was shown
  - Third ID photo was shown
  - None of above [note that the four choices' order is randomized]
- How confident do you feel that you correctly identified the person?
  - Completely unconfident
  - Mostly unconfident
  - Somewhat unconfident
  - Neither unconfident nor confident
  - Somewhat confident
  - Mostly confident
  - Completely confident

Considering the filter used in this photo, please rate the photo on the four statements below.



- The photo is satisfying.
  - Strongly disagree
  - Disagree
  - Somewhat disagree
  - Neither agree nor disagree
  - Somewhat agree
  - Agree
  - Strongly agree
  
- This photo provides sufficient information.
  - Strongly disagree
  - Disagree
  - Somewhat disagree
  - Neither agree nor disagree
  - Somewhat agree
  - Agree
  - Strongly agree
  
- I enjoy the photo at this moment.
  - Strongly disagree
  - Disagree
  - Somewhat disagree
  - Neither agree nor disagree
  - Somewhat agree
  - Agree
  - Strongly agree
  
- There was a sense of human contact when I saw the photo.

- Strongly disagree
- Disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Agree
- Strongly agree

[Participants then went through the other 13 trials.]

## C.7 Rating the Likability of Obfuscations

Imagine that online social networks (Facebook etc.) adopted privacy filters so that users could better manage the privacy of their photos. In that case, which privacy filter would you prefer?

Please rate your preference for each privacy filter.

- I like the “blurring” privacy filter. [An corresponding example obfuscation photo was shown.]
  - Strongly disagree
  - Disagree
  - Somewhat disagree
  - Neither agree nor disagree
  - Somewhat agree
  - Agree
  - Strongly agree

[Participants then rated the other six obfuscations using the same scale as above shows.]

## C.8 Follow-up Questions

- If you could use any of the privacy filters for photos you post on online social networks, which one, if any, would you like to use?

- As is (no filter)
- Blurring
- Morphing
- Silhouette
- Avatar
- Masking
- Inpainting
  
- Please tell us the reason [open-ended]
  
- I am willing to upload photos to online social networks using the filter I selected.
  - Strongly disagree
  - Disagree
  - Somewhat disagree
  - Neither agree nor disagree
  - Somewhat agree
  - Agree
  - Strongly agree
  
- Have you ever declined to upload a photo to an online social network for privacy reasons?
  - Yes
  - No
  - I don't know
  
- In the last question, you said you had declined to upload a photo to an online social network for privacy reasons. If you had access to one of the privacy filters here, would you be willing to upload the photo using one of the filters?
  - Yes
  - No

- I don't know
- If you answered "Yes" in above question, which privacy filter would you prefer to use? If you answered "No" or "I don't know" , please select "NA".
  - As is (no filter)
  - Blurring
  - Morphing
  - Silhouette
  - Avatar
  - Masking
  - Inpainting
- Please tell us your reasons why you chose this filter.

## C.9 Adoption Willingness

Please rate two statements about your adoption willingness.

- I want online social networks (Facebook etc.) to adopt these privacy filters SO that I can be obscured in certain photos my friends upload (group photo etc.).
  - Strongly disagree
  - Disagree
  - Somewhat disagree
  - Neither agree nor disagree
  - Somewhat agree
  - Agree
  - Strongly agree
- Please tell us the reason [open-ended]
- I want online social networks (Facebook etc.) to adopt these privacy filters SO that some people can be obscured in certain photos I view (group photo etc.).

- Strongly disagree
  - Disagree
  - Somewhat disagree
  - Neither agree nor disagree
  - Somewhat agree
  - Agree
  - Strongly agree
- Please tell us the reason [open-ended]

## C.10 Familiarity

Please note that you will be required to write down the names of the people in below images. Please try to answer as accurate as you can. If you cannot remember someone's name accurately, you may write anything that help us understand you are familiar with this person.

[We showed participants seven familiar targets that they saw during the testing and asked them to write down their names and rate the familiarity]

[Familiar target's photo was shown]

- Who is this person? Please write down his/her name. [open-ended]
- How familiar are you with this person?
  - Completely unfamiliar
  - Mostly unfamiliar
  - Somewhat unfamiliar
  - Neither unfamiliar nor familiar
  - Somewhat familiar
  - Mostly familiar
  - Completely familiar

[Participants then answered the same two questions for the rest of the six familiar targets.]

## C.11 Attitudes Towards Privacy and Security

Privacy means different things to different people today. In thinking about all of your daily interactions - both online and offline - please tell us how important each of the following are to you:

- Being in control of who can get information about you.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know
  
- Being able to share confidential matters with someone you trust.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know
  
- Not having someone watch you or listen to you without your permission.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know
  
- Controlling what information is collected about you.
  - Not at all important
  - Not very important
  - Somewhat important

- Very important
  - Don't know
- Not being disturbed at home.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know
- Being in control of who can get information about you.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know
- Being able to have times when you are completely alone, away from anyone else.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know
- Having individuals in social / work situations not ask you things that are highly personal.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important

- Don't know
- Being able to go around in public without always being identified.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know
- Not being monitored at work.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know

## C.12 Additional Feedback

- Do you have any comments, confusions, and suggestions about this survey? [open-ended]



## Appendix D Study 5 Online Experience Questions

This study had two steps – photo collection and system evaluation. The questions during the first step were shown below.

### D.1 Notice

This survey is the first step in a two-step study. Please take this survey ONLY IF you are willing to join the next study a few weeks later. We'll contact you via email prior to the next study.

1) Please read the instruction and questions on each screen carefully. We randomly insert some attention check questions. You will not get paid if you fail more than one attention question. We will check how many attention check questions you fail, and your answer quality before approving.

2) Please do NOT upload the same photo for all the photo uploading questions or have low-quality answers. We will consider those as low-quality answers, and exclude your data WITHOUT any payment.

3) Please do NOT take this survey repeatedly, we will recognize the answers from the same participant and only pay once.

### D.2 Participants Screening

- Please take out your phone. If you have an iPhone, please go to “Photos” and look at the “Camera Roll” album. If you have an Android phone, please look at the “Gallery”. Approximately how many photos do you have in this album? [open-ended]
- Please browse your photo album for at least three minutes. How many photos are there in your album that you would like to upload to an online social network such as Facebook, but have NOT because of privacy concerns. If you have MORE than 5, please choose “5.”

– 1

– 2

– 3

– 4

– 5

- I don't have any
- Are you willing to provide us these photos so that we can use them as your study material in the second step – Facebook study?
  - Yes
  - No
- This photo collection study is the first step in a two-step study. We will invite some participants who provide qualified photos for a second study–Facebook study. We'll send you a reminder via email once the Facebook study is ready for you (may take two or three weeks). Are you available to participate in the Facebook study a few weeks later?
  - Yes
  - No
- About how often do you use or visit Facebook?
  - Most of the day
  - Several times a day
  - About once a day
  - A few times a week
  - A few times a month
  - A few times a year
  - Never
  - I don't have a Facebook account
- About how often do you upload photos to an online social network such as Facebook?
  - Most of the day
  - Several times a day
  - About once a day
  - A few times a week

- A few times a month
  - A few times a year
  - Never
- Have you ever declined to upload a photo to an Online Social Network such as Facebook due to your privacy concerns?
  - Yes
  - No

### **D.3 Demographics**

- What is your gender?
  - Male
  - Female
  - Other
  - I prefer not to answer
- What is your age?
  - 18-24 years old
  - 25-34 years old
  - 35-44 years old
  - 45-54 years old
  - 55+
- What is your ethnicity?
  - White
  - Hispanic or Latino
  - Black or African American
  - Native American or American Indian

- Asian / Pacific Islander
  - Other
- Which of the following best describes your current employer?
  - Government
  - Educational institution
  - Business or industry
  - Non-profit organization
  - Other
- What is the highest level of school you have completed or the highest degree you have received?
  - High school incomplete or less
  - High school graduate or GED (includes technical/vocational training that doesn't count towards college credit)
  - Some college (some community college, associate's degree)
  - Four year college degree/bachelor's degree
  - Some postgraduate or professional schooling, no postgraduate degree
  - Postgraduate or professional degree, including master's, doctorate, medical or law degree
  - I prefer not to answer
- Which of these best describes you?
  - Married
  - Living with a partner
  - Divorced
  - Separated
  - Widowed
  - Never been married
  - I prefer not to answer

- About how often do you use Internet either on a computer or on a mobile device like a smartphone or a tablet?
  - Most of the day
  - Several times a day
  - About once a day
  - A few times a week
  - A few times a month
  - A few times a year
  - Never
  - I prefer not to answer
  
- About how often do you visit social media sites such as Facebook, Twitter or LinkedIn?
  - Most of the day
  - Several times a day
  - About once a day
  - A few times a week
  - A few times a month
  - A few times a year
  - Never
  - I prefer not to answer

#### **D.4 Contact**

- What is your email address that we can reach out for the second step study? [open-ended]
  
- What is your Facebook profile name? [open-ended]

## D.5 Upload Instruction

Please read the instructions below carefully:

In the next few questions, you will provide us [the number of photos participants chose in screening questions] photo(s) that you would like to upload to an online social network such as Facebook, but have NOT because of privacy concerns.

We STRONGLY recommend that you choose a photo you took with the camera on your mobile phone. The photo you choose will be shown to you during the second step – Facebook study. Please note that only you and the researchers will see your photos.

## D.6 Photo Collection

- Please browse your photo album for at least two minutes and find a photo that you would like to post on your Facebook but have NOT because of privacy concerns.

– Okay, found one!

- Are you willing to share it with us (the researchers conducting this study)? As a reminder, we won't share it with anyone.

– Yes

– No

[If participants answered “Yes”:]

- Please upload the photo. [Photo uploading button]

[If participants answered “No”:]

- That's okay, we understand. If you feel uncomfortable to upload the original photo, please upload a photo you found online that has similar sensitive content to this photo. [Photo uploading button]

- What content prevents you from posting this photo? (Minimum 100 Characters Required)  
[open-ended question]

- What about this photo that makes you want to upload if you didn't have privacy concerns? (Minimum 100 Characters Required) [open-ended question]

[If participants indicated that they had more photos, they would keep looping, provide more photos, and answer the same set of questions.]

## D.7 Attitudes Towards Privacy and Security

Privacy means different things to different people today. In thinking about all of your daily interactions - both online and offline - please tell us how important each of the following are to you:

- Being in control of who can get information about you.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know
  
- Being able to share confidential matters with someone you trust.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know
  
- Not having someone watch you or listen to you without your permission.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know
  
- Controlling what information is collected about you.

- Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know
- Not being disturbed at home.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know
- Being in control of who can get information about you.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know
- Being able to have times when you are completely alone, away from anyone else.
  - Not at all important
  - Not very important
  - Somewhat important
  - Very important
  - Don't know
- Having individuals in social / work situations not ask you things that are highly personal.
  - Not at all important



- Not very important
  - Somewhat important
  - Very important
  - Don't know
- Being able to go around in public without always being identified.
    - Not at all important
    - Not very important
    - Somewhat important
    - Very important
    - Don't know
  - Not being monitored at work.
    - Not at all important
    - Not very important
    - Somewhat important
    - Very important
    - Don't know

[The questions for the second-step study are shown below.]

## **D.8 Confirming Email**

- What is your email address that I contacted for this study? We will use your email address as a reference for your response. [open-ended]
- Please confirm your Facebook profile name. [open-ended]

## **D.9 Pre-test Questions: Trust in SNSs**

- I trust that Facebook will not use my personal information for any other purpose.
  - 1 - Strongly Disagree

- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

- I feel that the privacy of my personal information is protected by Facebook.

- 1 - Strongly Disagree
- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

- I believe most of the profiles I view on Facebook are exaggerated to make the personal look more appealing.

- 1 - Strongly Disagree
- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

- I worry that I will be embarrassed by wrong information others post about me on Facebook.

- 1 - Strongly Disagree

- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

### D.10 Pre-test Questions: Interpersonal Privacy Concerns

- It usually bothers me when people ask me something personal.

- 1 - Strongly Disagree
- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

- I will tell people anything they want to know about me.

- 1 - Strongly Disagree
- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

- I have nothing to hide from other people.

- 1 - Strongly Disagree

- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

- I am concerned that people know too many personal things about me.

- 1 - Strongly Disagree
- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

- To me, it is the most important thing to keep things private from others.

- 1 - Strongly Disagree
- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

- When people ask me something personal, I sometimes think twice before telling them.

- 1 - Strongly Disagree
- 2

- 3
- 4
- 5
- 6
- 7 - Strongly Agree

- I think it is risky to tell people personal things about myself.

- 1 - Strongly Disagree
- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

- I feel safe telling people personal things about me.

- 1 - Strongly Disagree
- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

- I feel comfortable sharing my private thoughts and feelings with others.

- 1 - Strongly Disagree
- 2
- 3

- 4
- 5
- 6
- 7 - Strongly Agree

## D.11 Pre-test Questions: Perceived Privacy Risks

In the invitation email we sent to you, you have seen a photo that you provided us about a month ago. (If you forget which photo it is, please check your email)

Regarding that photo, please answer the following questions.

- How risky would you say it would be to post this photo on Facebook?
  - 1 - Not risky at all
  - 2
  - 3
  - 4
  - 5
  - 6
  - 7 - Very risky
  
- Posting this photo on Facebook would be risky.
  - 1 - Strongly Disagree
  - 2
  - 3
  - 4
  - 5
  - 6
  - 7 - Strongly Agree
  
- Posting this photo on Facebook is dangerous.

- 1 - Strongly Disagree
- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

- Sharing this photo on my Facebook would add great uncertainty to my privacy.

- 1 - Strongly Disagree
- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

- Sharing this photo on my Facebook exposes me to an overall risk.

- 1 - Strongly Disagree
- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

## D.12 Pre-test Questions: Willingness to Share

- I am willing to share this photo on my Facebook.
  - 1 - Strongly Disagree
  - 2
  - 3
  - 4
  - 5
  - 6
  - 7 - Strongly Agree

## D.13 Interacting with Prototype

- You will use a Facebook prototype we create (you will get paid ONLY IF you go to the link below and actually interact with the prototype, we'll check the backend data). Please read the instruction carefully when first landing on the prototype.
- After successfully finishing the task, you should see a random code that proves you've finished the task. Please paste the code to the field below.
- Now using your laptop or PC (NOT the phone), please open a new tab, copy this link (do NOT click on it or you will exit your survey): <https://clemsontphotos.sites.clemson.edu/login.php>  
Open the site, you'll log in using the email address and password in the invitation email.
- After using the prototype and pasting the code, please continue to finish the remaining questions. [open-ended question]

## D.14 Post-test Questions: Perceived Privacy Risks

Regarding the photo you saw in the prototype, please answer the following questions.

- How risky would you say it would be to post this photo on Facebook?
  - 1 - Not risky at all
  - 2



- 3
- 4
- 5
- 6
- 7 - Very risky

- Posting this photo on Facebook would be risky.

- 1 - Strongly Disagree
- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

- Posting this photo on Facebook is dangerous.

- 1 - Strongly Disagree
- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

- Sharing this photo on my Facebook would add great uncertainty to my privacy.

- 1 - Strongly Disagree
- 2
- 3

- 4
  - 5
  - 6
  - 7 - Strongly Agree
- Sharing this photo on my Facebook exposes me to an overall risk.
    - 1 - Strongly Disagree
    - 2
    - 3
    - 4
    - 5
    - 6
    - 7 - Strongly Agree

#### **D.15 Post-test Questions: Willingness to Share**

- I am willing to share this photo on my Facebook.
  - 1 - Strongly Disagree
  - 2
  - 3
  - 4
  - 5
  - 6
  - 7 - Strongly Agree

#### **D.16 Post-test Questions: Perceived System Effectiveness**

Regarding the Facebook prototype that you just used, please answer the following questions.

- This system has no real benefit for me.

- 1 - Strongly Disagree
  - 2
  - 3
  - 4
  - 5
  - 6
  - 7 - Strongly Agree
  
- This system is useful.
  - 1 - Strongly Disagree
  - 2
  - 3
  - 4
  - 5
  - 6
  - 7 - Strongly Agree
  
- I can protect my privacy better using this system.
  - 1 - Strongly Disagree
  - 2
  - 3
  - 4
  - 5
  - 6
  - 7 - Strongly Agree
  
- I can protect my privacy better using other approaches without the help of this system.
  - 1 - Strongly Disagree

- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

### D.17 Post-test Questions: Usage Efforts

- The system is convenient.
  - 1 - Strongly Disagree
  - 2
  - 3
  - 4
  - 5
  - 6
  - 7 - Strongly Agree
  
- I do not have to invest a lot of effort in the system.
  - 1 - Strongly Disagree
  - 2
  - 3
  - 4
  - 5
  - 6
  - 7 - Strongly Agree
  
- It takes many mouse-clicks to use the system.
  - 1 - Strongly Disagree

- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

### D.18 Post-test Questions: System Satisfaction

- I am very satisfied when using the system.

- 1 - Strongly Disagree
- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

- I am very pleased when using the system.

- 1 - Strongly Disagree
- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

- Using this system made me contented.

- 1 - Strongly Disagree

- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

- I feel delighted when using this system.

- 1 - Strongly Disagree
- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

- I will strongly recommend it to my friends.

- 1 - Strongly Disagree
- 2
- 3
- 4
- 5
- 6
- 7 - Strongly Agree

- I will most likely use this system again.

- 1 - Strongly Disagree
- 2

- 3
- 4
- 5
- 6
- 7 - Strongly Agree

### **D.19 Qualitative Feedback**

- What did you like the most about using this system? Please tell us the reasons. (Minimum 100 characters required) [open-ended question]
- What did you like the least about using this system? Please tell us the reasons. (Minimum 100 characters required) [open-ended question]
- Do you have any other comments or suggestions on this system? (Minimum 100 characters required) [open-ended question]

# Bibliography

- [1] James D Abbey and Margaret G Meloy. Attention by design: Using attention checks to detect inattentive respondents and improve data quality. Journal of Operations Management, 53:63–70, 2017.
- [2] Patricia Sanchez Abril. A (my) space of one’s own: on privacy and online social networks. Nw. J. Tech. & Intell. Prop., 6:73, 2007.
- [3] Anne Adams, Sally Jo Cunningham, and Masood Masoodian. Sharing, privacy and trust issues for photo collections. Working paper, 2007.
- [4] Adobe Support. Content-aware patch and move, February 2017. Retrieved April 23, 2017 from <https://helpx.adobe.com/photoshop/using/content-aware-patch-move.html>.
- [5] Prachi Agrawal. De-Identification for Privacy Protection in Surveillance Videos. PhD thesis, International Institute of Information Technology Hyderabad, India, 2010.
- [6] Shane Ahern, Dean Eckles, Nathaniel S Good, Simon King, Mor Naaman, and Rahul Nair. Over-exposed?: privacy patterns and considerations in online and mobile photo sharing. In Proceedings of the SIGCHI conference on Human factors in Computing Systems, pages 357–366. ACM, 2007.
- [7] 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), 2020.
- [8] Abdullah Al Hasib. Threats of online social networks. IJCSNS International Journal of Computer Science and Network Security, 9(11):288–93, 2009.
- [9] Rawan Alharbi, Mariam Tolba, Lucia C Petito, Josiah Hester, and Nabil Alshurafa. To mask or not to mask? balancing privacy with visual confirmation utility in activity-oriented wearable cameras. Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies, 3(3):1–29, 2019.
- [10] David S Allison and Miriam AM Capretz. Furthering the growth of cloud computing by providing privacy as a service. In International Conference on Information and Communication on Technology, pages 64–78. Springer, 2011.
- [11] Irwin Altman. The environment and social behavior: Privacy, personal space, territory, and crowding. 1975.
- [12] 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 79–95, 2020.



- [13] Anonymous Camera. Anonymous camera, 2020. Retrieved July 5, 2020 from <https://apps.apple.com/us/app/anonymous-camera/id1504102584>.
- [14] Salman Aslam. Snapchat by the numbers: Stats, demographics & fun facts, 2018. Retrieved December 11, 2018 from <https://www.omnicoreagency.com/snapchat-statistics/>.
- [15] Tuomas Aura, Thomas A Kuhn, and Michael Roe. Scanning electronic documents for personally identifiable information. In Proceedings of the 5th ACM workshop on Privacy in electronic society. ACM, 2006.
- [16] Karla Badillo-Urquiola, Yaxing Yao, Oshrat Ayalon, Bart Knijnenurg, Xinru Page, Eran Toch, Yang Wang, and Pamela J Wisniewski. Privacy in context: Critically engaging with theory to guide privacy research and design. In Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing, pages 425–431. ACM, 2018.
- [17] Gaurav Bansal, David Gefen, et al. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. Decision support systems, 49(2):138–150, 2010.
- [18] Daniel Bar-Tal. Self-censorship as a socio-political-psychological phenomenon: Conception and research. Political Psychology, 38:37–65, 2017.
- [19] Susan B Barnes. A privacy paradox: Social networking in the united states. First Monday, 11(9), 2006.
- [20] Eric P.S. Baumer, Xiaotong Xu, Christine Chu, Shion Guha, and Geri K. Gay. When subjects interpret the data: Social media non-use as a case for adapting the delphi method to cscw. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW '17, pages 1527–1543, New York, NY, USA, 2017. ACM.
- [21] Kathy Baxter, Catherine Courage, and Kelly Caine. Understanding your users: A practical guide to user research methods. Morgan Kaufmann, 2015.
- [22] Gary Bente, Sabine Rüggenberg, Nicole C Krämer, and Felix Eschenburg. Avatar-mediated networking: Increasing social presence and interpersonal trust in net-based collaborations. Human communication research, 34(2):287–318, 2008.
- [23] Michael S Bernstein, Eytan Bakshy, Moira Burke, and Brian Karrer. Quantifying the invisible audience in social networks. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pages 21–30. ACM, 2013.
- [24] Andrew Besmer and Heather Lipford. Tagged photos: concerns, perceptions, and protections. In CHI'09 Extended Abstracts on Human Factors in Computing Systems, pages 4585–4590. ACM, 2009.
- [25] 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. ACM, 2010.
- [26] Leyla Bilge, Thorsten Strufe, Davide Balzarotti, and Engin Kirda. All your contacts are belong to us: automated identity theft attacks on social networks. In Proceedings of the 18th international conference on World wide web, pages 551–560. ACM, 2009.
- [27] Markus Bindemann, Janice Attard, Amy Leach, and Robert A Johnston. The effect of image pixelation on unfamiliar-face matching. Applied Cognitive Psychology, 27(6):707–717, 2013.

- [28] Jens Binder, Andrew Howes, and Alistair Sutcliffe. The problem of conflicting social spheres: effects of network structure on experienced tension in social network sites. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pages 965–974. ACM, 2009.
- [29] 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.
- [30] Bitmoji. Your own personal emoji, 2016. Retrieved April 23, 2017 from <https://www.bitmoji.com/>.
- [31] Danah Boyd. Networked privacy. Surveillance & Society, 10(3/4):348, 2012.
- [32] Danah Boyd and Eszter Hargittai. Facebook privacy settings: Who cares? First Monday, 15(8), 2010.
- [33] Michael Boyle, Christopher Edwards, and Saul Greenberg. The effects of filtered video on awareness and privacy. In Proceedings of the 2000 ACM conference on Computer supported cooperative work, pages 1–10. ACM, 2000.
- [34] Petter Bae Brandtzæg, Marika Lüders, and Jan Håvard Skjetne. Too many facebook “friends”? content sharing and sociability versus the need for privacy in social network sites. Intl. Journal of Human-Computer Interaction, 26(11-12):1006–1030, 2010.
- [35] Kieran Browne, Ben Swift, and Terhi Nurmikko-Fuller. Camera adversaria. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pages 1–9, 2020.
- [36] Vicki Bruce, Zoë Henderson, Craig Newman, and A Mike Burton. Matching identities of familiar and unfamiliar faces caught on cctv images. Journal of Experimental Psychology: Applied, 7(3):207, 2001.
- [37] Vicki Bruce and Tim Valentine. Identity priming in the recognition of familiar faces. British Journal of Psychology, 76(3):373–383, 1985.
- [38] Vicki Bruce and Andy Young. Understanding face recognition. British journal of psychology, 77(3):305–327, 1986.
- [39] Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. Amazon’s mechanical turk: A new source of inexpensive, yet high-quality, data? Perspectives on psychological science, 6(1):3–5, 2011.
- [40] United States Census Bureau. Geography division, 2014. Retrieved April 21, 2017 from [https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us\\_regdiv.pdf](https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf).
- [41] U.S. Census Bureau. American factfinder - race results, 2010.
- [42] Thorben Burghardt, Andreas Walter, Erik Buchmann, and Klemens Böhm. Primo-towards privacy aware image sharing. In 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, volume 3, pages 21–24. IEEE, 2008.
- [43] Judee K Burgoon, Roxanne Parrott, Beth A Le Poire, Douglas L Kelley, Joseph B Walther, and Denise Perry. Maintaining and restoring privacy through communication in different types of relationships. Journal of Social and Personal Relationships, 6(2):131–158, 1989.
- [44] A Mike Burton, Vicki Bruce, and Robert A Johnston. Understanding face recognition with an interactive activation model. British Journal of Psychology, 81(3):361–380, 1990.

- [45] Daniel Buschek, Moritz Bader, Emanuel von Zezschwitz, and Alexander De Luca. Automatic privacy classification of personal photos. In Human-Computer Interaction, pages 428–435. Springer, 2015.
- [46] Byebye Camera. Byebye camera, 2020. Retrieved July 5, 2020 from <https://apps.apple.com/us/app/bye-bye-camera/id1467903179>.
- [47] Kelly Caine. Linking studies of privacy in hci to psychological theories of privacy, 2008.
- [48] Kelly Caine, Lorraine G Kisselburgh, and Louise Lareau. Audience visualization influences disclosures in online social networks. In CHI’11 Extended Abstracts on Human Factors in Computing Systems, pages 1663–1668. ACM, 2011.
- [49] Kelly E Caine, Wendy A Rogers, and Arthur D Fisk. Privacy perceptions of an aware home with visual sensing devices. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, volume 49, pages 1856–1858. SAGE Publications Sage CA: Los Angeles, CA, 2005.
- [50] Kelly Erinn Caine. Exploring everyday privacy behaviors and misclosures. PhD thesis, Georgia Institute of Technology, 2009.
- [51] Krista Casler, Lydia Bickel, and Elizabeth Hackett. Separate but equal? a comparison of participants and data gathered via amazon’s mturk, social media, and face-to-face behavioral testing. Computers in Human Behavior, 29(6):2156–2160, 2013.
- [52] Datong Chen, Yi Chang, Rong Yan, and Jie Yang. Protecting personal identification in video. In Protecting Privacy in Video Surveillance, pages 115–128. Springer, 2009.
- [53] Eun Kyoung Choe, Jaeyeon Jung, Bongshin Lee, and Kristie Fisher. Nudging people away from privacy-invasive mobile apps through visual framing. In IFIP Conference on Human-Computer Interaction, pages 74–91. Springer, 2013.
- [54] Edward J Clarke, Mar Preston, Jo Raksin, and Vern L Bengtson. Types of conflicts and tensions between older parents and adult children. The Gerontologist, 39(3):261–270, 1999.
- [55] Sunny Consolvo, Ian E Smith, Tara Matthews, Anthony LaMarca, Jason Tabert, and Pauline Powledge. Location disclosure to social relations: why, when, & what people want to share. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems. ACM, 2005.
- [56] 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. ACM, 2011.
- [57] Mark G Core, H Chad Lane, Michael Van Lent, Dave Gomboc, Steve Solomon, and Milton Rosenberg. Building explainable artificial intelligence systems. In AAAI, pages 1766–1773, 2006.
- [58] David Crandall and Noah Snavely. Modeling people and places with internet photo collections. Communications of the ACM, 55(6):52–60, 2012.
- [59] Dianne Cyr, Milena Head, Hector Larios, and Bing Pan. Exploring human images in website design: a multi-method approach. MIS quarterly, pages 539–566, 2009.
- [60] Sauvik Das and Adam Kramer. Self-censorship on facebook. In Proceedings of the International Conference on Weblogs and Social Media (ICWSM), pages 120–127, 2013.

- [61] Katie Davis and Carrie James. Tweens' conceptions of privacy online: implications for educators. Learning, Media and Technology, 38(1):4–25, 2013.
- [62] Patricia C de Souza and Cristiano Maciel. Legal issues and user experience in ubiquitous systems from a privacy perspective. In International Conference on Human Aspects of Information Security, Privacy, and Trust, pages 449–460. Springer, 2015.
- [63] 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.
- [64] Judith DeCew. Privacy. In Edward N. Zalta, editor, The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, spring 2018 edition, 2018.
- [65] Jelle Demanet, Kristof Dhont, Lies Notebaert, Sven Pattyn, and André Vandierendonck. Pixelating familiar people in the media: Should masking be taken at face value? Psychologica belgica, 47(4), 2007.
- [66] Tamara Denning, Zakariya Dehlawi, and Tadayoshi Kohno. In situ with bystanders of augmented reality glasses: Perspectives on recording and privacy-mediating technologies. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 2377–2386. ACM, 2014.
- [67] Amandeep Dhir, Torbjørn Torsheim, Ståle Pallesen, and Cecilie S Andreassen. Do online privacy concerns predict selfie behavior among adolescents, young adults and adults? Frontiers in psychology, 8:815, 2017.
- [68] Digital Media Law. Barrow county school district v. payne, 2010. Retrieved October 5, 2018 from [http://www.dmlp.org/threats/barrow-county-school-district-v-payne#node\\_legal\\_threat\\_full\\_group\\_description](http://www.dmlp.org/threats/barrow-county-school-district-v-payne#node_legal_threat_full_group_description).
- [69] 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.
- [70] Christine DiStefano, Min Zhu, and Diana Mindrila. Understanding and using factor scores: Considerations for the applied researcher. Practical Assessment, Research, and Evaluation, 14(1):20, 2009.
- [71] J. B. Djoko, J. Lange, and A. 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, 2019.
- [72] William J Doll and Gholamreza Torkzadeh. The measurement of end-user computing satisfaction. MIS quarterly, pages 259–274, 1988.
- [73] Derek Doran, Sarah Schulz, and Tarek R Besold. What does explainable ai really mean? a new conceptualization of perspectives. arXiv preprint arXiv:1710.00794, 2017.
- [74] Marcie D Dorethy, Martin S Fiebert, and Christopher R Warren. Examining social networking site behaviors: Photo sharing and impression management on facebook. International Review of Social Sciences and Humanities, 6(2):111–116, 2014.
- [75] Steven Dow, Blair MacIntyre, Jaemin Lee, Christopher Oezbek, Jay David Bolter, and Mari-beth Gandy. Wizard of oz support throughout an iterative design process. IEEE Pervasive Computing, 4(4):18–26, 2005.

- [76] Liang Du, Meng Yi, Erik Blasch, and Haibin Ling. Garp-face: Balancing privacy protection and utility preservation in face de-identification. In Biometrics (IJCB), 2014 IEEE International Joint Conference on, pages 1–8. IEEE, 2014.
- [77] Emília Duarte, Francisco Rebelo, Júlia Teles, and Michael S Wogalter. A personalized speech warning facilitates compliance in an immersive virtual environment. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, volume 56, pages 2045–2049. SAGE Publications Sage CA: Los Angeles, CA, 2012.
- [78] Catherine Dwyer, Starr Hiltz, and Katia Passerini. Trust and privacy concern within social networking sites: A comparison of facebook and myspace. AMCIS 2007 proceedings, page 339, 2007.
- [79] Malin Eiband, Mohamed Khamis, Emanuel Von Zezschwitz, Heinrich Hussmann, and Florian Alt. Understanding shoulder surfing in the wild: Stories from users and observers. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pages 4254–4265. ACM, 2017.
- [80] Passant Elagroudy, Mohamed Khamis, Florian Mathis, Diana Irmscher, Andreas Bulling, and Albrecht Schmidt. Can privacy-aware lifelogs alter our memories? In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, pages 1–6, 2019.
- [81] Hadyn D Ellis, John W Shepherd, and Graham M Davies. Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. Perception, 8(4):431–439, 1979.
- [82] Nicole B Ellison, Charles Steinfield, and Cliff Lampe. The benefits of facebook “friends:” social capital and college students’ use of online social network sites. Journal of Computer-Mediated Communication, 12(4):1143–1168, 2007.
- [83] Nicole B Ellison, Jessica Vitak, Charles Steinfield, Rebecca Gray, and Cliff Lampe. Negotiating privacy concerns and social capital needs in a social media environment. In Privacy online, pages 19–32. Springer, 2011.
- [84] Facebook. How do i edit the privacy settings for my photo albums?, 2017. Retrieved January 2, 2018 from [https://www.facebook.com/help/215496745135618?helpref=about\\_content](https://www.facebook.com/help/215496745135618?helpref=about_content).
- [85] Facebook Help Center. Approve or remove tags, 2018. Retrieved October 5, 2018 from [https://www.facebook.com/help/267689476916031/?helpref=hc\\_fnav](https://www.facebook.com/help/267689476916031/?helpref=hc_fnav).
- [86] Facebook Help Center. Who can see what you post, 2018. Retrieved October 5, 2018 from [https://www.facebook.com/help/1297502253597210/?helpref=hc\\_fnav](https://www.facebook.com/help/1297502253597210/?helpref=hc_fnav).
- [87] Liang Fang, Lihua Yin, Qiaoduo Zhang, Fenghua Li, and Binxing Fang. Who is visible: Resolving access policy conflicts in online social networks. In GLOBECOM 2017-2017 IEEE Global Communications Conference, pages 1–6. IEEE, 2017.
- [88] Lujun Fang and Kristen LeFevre. Privacy wizards for social networking sites. In Proceedings of the 19th international conference on World wide web, pages 351–360, 2010.
- [89] Martha Farah, GW Humphreys, and HR Rodman. Object and face recognition. Fundamental neuroscience, ed. MJ Zigmond, FE Bloom, SC Landis, JL Roberts & LR Squire. Academic Press.[aFVDV], 1999.
- [90] Mauricio S Featherman and Paul A Pavlou. Predicting e-services adoption: a perceived risk facets perspective. International journal of human-computer studies, 59(4):451–474, 2003.

- [91] Kraig Finstad. Response interpolation and scale sensitivity: Evidence against 5-point scales. Journal of Usability Studies, 5(3):104–110, 2010.
- [92] Michael Fire, Roy Goldschmidt, and Yuval Elovici. Online social networks: threats and solutions. IEEE Communications Surveys & Tutorials, 16(4):2019–2036, 2014.
- [93] Simone Fischer-Hubner, Chris Hoofnagle, Ioannis Krontiris, Kai Rannenberg, and Michael Waidner. Online privacy: Towards informational self-determination on the internet. 2011.
- [94] Andrew J Flanagin and Miriam J Metzger. Internet use in the contemporary media environment. Human communication research, 27(1):153–181, 2001.
- [95] Ricard L Fogues, Jose M Such, Agustin Espinosa, and Ana Garcia-Fornes. Exploring the viability of tie strength and tags in access controls for photo sharing. In Proceedings of the Symposium on Applied Computing, pages 1082–1085. ACM, 2017.
- [96] Forbes. Celebrity 100, 2017. <https://www.forbes.com/celebrities/#1775ffe95947>.
- [97] Jesse Fox and Megan A Vendemia. Selective self-presentation and social comparison through photographs on social networking sites. Cyberpsychology, behavior, and social networking, 19(10):593–600, 2016.
- [98] Andrea Frome, German Cheung, Ahmad Abdulkader, Marco Zennaro, Bo Wu, Alessandro Bissacco, Hartwig Adam, Hartmut Neven, and Luc Vincent. Large-scale privacy protection in google street view. In Computer Vision, 2009 IEEE 12th International Conference on, pages 2373–2380. IEEE, 2009.
- [99] Genevieve Gebhart and Tadayoshi Kohno. Internet censorship in thailand: User practices and potential threats. In Security and Privacy (EuroS&P), 2017 IEEE European Symposium on, pages 417–432. IEEE, 2017.
- [100] Liqiang Geng, Larry Korba, Xin Wang, Yunli Wang, Hongyu Liu, and Yonghua You. Using data mining methods to predict personally identifiable information in emails. In International Conference on Advanced Data Mining and Applications, pages 272–281. Springer, 2008.
- [101] Kambiz Ghazinour and John Ponchak. Hidden privacy risks in sharing pictures on social media. Procedia Computer Science, 113:267–272, 2017.
- [102] Eric Gilbert and Karrie Karahalios. Predicting tie strength with social media. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pages 211–220. ACM, 2009.
- [103] Google Street View. Image acceptance and privacy policies, 2018. Retrieved May 7, 2018 from <https://www.google.com/streetview/privacy/>.
- [104] Pulkit Goyal, Sapan Diwakar, et al. Fast and enhanced algorithm for exemplar based image inpainting. In Image and Video Technology (PSIVT), 2010 Fourth Pacific-Rim Symposium on, pages 325–330. IEEE, 2010.
- [105] Steven Greenhouse and Michael Barbaro. Wal-mart memo suggests ways to cut employee benefit costs, 2005. Retrieved September 8, 2018 from <https://www.nytimes.com/2005/10/26/business/walmart-memo-suggests-ways-to-cut-employee-benefit-costs.html>.
- [106] Ralph Gross, Edoardo Airoldi, Bradley Malin, and Latanya Sweeney. Integrating utility into face de-identification. In International Workshop on Privacy Enhancing Technologies, pages 227–242. Springer, 2005.

- [107] Rebecca Gulotta, Haakon Faste, and Jennifer Mankoff. Curation, provocation, and digital identity: risks and motivations for sharing provocative images online. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 387–390, 2012.
- [108] Rakibul Hasan, David Crandall, Mario Fritz, and Apu Kapadia. Automatically detecting bystanders in photos to reduce privacy risks. 2020.
- [109] 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, page 47. ACM, 2018.
- [110] 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.
- [111] 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. IEEE, 2016.
- [112] Amelia Heathman. Fabook’s snapchat-style stories and effects are live. here’s how to use them, March 2017. Retrieved April 23, 2017 from <http://www.wired.co.uk/article/facebook-filters-like-snapchat>.
- [113] Benjamin Henne, Maximilian Koch, and Matthew Smith. On the awareness, control and privacy of shared photo metadata. In International Conference on Financial Cryptography and Data Security, pages 77–88. Springer, 2014.
- [114] Benjamin Henne, Christian Szongott, and Matthew Smith. Snapme if you can: privacy threats of other peoples’ geo-tagged media and what we can do about it. In Proceedings of the sixth ACM conference on Security and privacy in wireless and mobile networks, pages 95–106, 2013.
- [115] Nicola Henry, Anastasia Powell, and Asher Flynn. Ai can now create fake porn, making revenge porn even more complicated. The Conversation, 28, 2018.
- [116] Heroes Rising. 15 year old girl commits suicide after nude photos posted on facebook, 2012. Retrieved October 5, 2018 from <https://www.causes.com/causes/580526-heroes-rising/updates/616046-15-year-old-girl-commits-suicide-after-nude-photos-posted-on-facebook>.
- [117] Susan Herring and Ashley Dainas. “nice picture comment!” graphicons in facebook comment threads. In Proceedings of the 50th Hawaii International Conference on System Sciences, 2017.
- [118] Anne Hewitt and Andrea Forte. Crossing boundaries: Identity management and student/faculty relationships on the facebook. Poster presented at CSCW, Banff, Alberta, pages 1–2, 2006.
- [119] E Tory Higgins. Self-discrepancy: a theory relating self and affect. Psychological review, 94(3):319, 1987.
- [120] Christian Pieter Hoffmann and Christoph Lutz. Spiral of silence 2.0: Political self-censorship among young facebook users. In Proceedings of the 8th International Conference on Social Media & Society, pages 1–12, 2017.

- [121] 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, pages 1645–1648. ACM, 2015.
- [122] 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, pages 571–582. ACM, 2014.
- [123] Hongxin Hu, Gail-Joon Ahn, and Jan Jorgensen. Multiparty access control for online social networks: model and mechanisms. IEEE transactions on knowledge and data engineering, 25(7):1614–1627, 2013.
- [124] Panagiotis Iliia, 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. ACM, 2015.
- [125] Instagram. Imma instagram, 2020. Retrieved July 5, 2020 from <https://www.instagram.com/imma.gram/?hl=en>.
- [126] Danesh Irani, Steve Webb, Kang Li, and Calton Pu. Modeling unintended personal-information leakage from multiple online social networks. IEEE Internet Computing, 15(3):13–19, 2011.
- [127] Suman Jana, Arvind Narayanan, and Vitaly Shmatikov. A scanner darkly: Protecting user privacy from perceptual applications. In Security and Privacy (SP), 2013 IEEE Symposium on, pages 349–363. IEEE, 2013.
- [128] Samantha Jaroszewski, Danielle Lottridge, Oliver L Haimson, and Katie Quehl. Genderfluid or attack helicopter: Responsible hci research practice with non-binary gender variation in online communities. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, page 307. ACM, 2018.
- [129] Jobvite. Social recruiting survey results, 2014. Retrieved October 5, 2018 from [https://timedotcom.files.wordpress.com/2014/09/jobvite\\_socialrecruiting\\_survey2014.pdf](https://timedotcom.files.wordpress.com/2014/09/jobvite_socialrecruiting_survey2014.pdf).
- [130] Rob Johns. Likert items and scales. Survey Question Bank: Methods Fact Sheet, 1:1–11, 2010.
- [131] Robert A Johnston and Andrew J Edmonds. Familiar and unfamiliar face recognition: A review. Memory, 17(5):577–596, 2009.
- [132] Harvey Jones and José Hiram Soltren. Facebook: Threats to privacy. Project MAC: MIT Project on Mathematics and Computing, 1:1–76, 2005.
- [133] Holy Juan. Facebook redacting, 2010. Retrieved April 23, 2017 from <http://www.holyjuan.com/2010/12/facebook-redacting.html>.
- [134] Aleksandra Kacperczyk. Social isolation in the workplace: A cross-national and longitudinal analysis. 2011.
- [135] Sanjay Kairam, Mike Brzozowski, David Huffaker, and Ed Chi. Talking in circles: selective sharing in google+. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pages 1065–1074. ACM, 2012.



- [136] Sanjay Kairam, Joseph'Jofish' Kaye, John Alexis Guerra-Gomez, and David A Shamma. Snap decisions?: How users, content, and aesthetics interact to shape photo sharing behaviors. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pages 113–124. ACM, 2016.
- [137] Dan J Kim, Donald L Ferrin, and H Raghav Rao. Trust and satisfaction, two stepping stones for successful e-commerce relationships: A longitudinal exploration. Information systems research, 20(2):237–257, 2009.
- [138] Myeung Un Kim, Harim Lee, Hyun Jong Yang, and Michael S Ryoo. Privacy-preserving robot vision with anonymized faces by extreme low resolution. In 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 462–467. IEEE, 2019.
- [139] Amanda M Kimbrough, Rosanna E Guadagno, Nicole L Muscanell, and Janeann Dill. Gender differences in mediated communication: Women connect more than do men. Computers in Human Behavior, 29(3):896–900, 2013.
- [140] Peter Klemperer, Yuan Liang, Michelle Mazurek, Manya Sleeper, Blase Ur, Lujo Bauer, Lorie Faith Cranor, Nitin Gupta, and Michael Reiter. Tag, you can see it!: Using tags for access control in photo sharing. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 377–386. ACM, 2012.
- [141] Bart P Knijnenburg. Privacy? i can't even! making a case for user-tailored privacy. IEEE Security & Privacy, 15(4):62–67, 2017.
- [142] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. Explaining the user experience of recommender systems. User Modeling and User-Adapted Interaction, 22(4-5):441–504, 2012.
- [143] Bart Piet Knijnenburg. A user-tailored approach to privacy decision support. PhD thesis, UC Irvine, 2015.
- [144] Bart Piet Knijnenburg and Alfred Kobsa. Increasing sharing tendency without reducing satisfaction: finding the best privacy-settings user interface for social networks. 2014.
- [145] Mohammed Korayem, Robert Templeman, Dennis Chen, David Crandall, and Apu Kapadia. Enhancing lifelogging privacy by detecting screens. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pages 4309–4314. ACM, 2016.
- [146] Pavel Korshunov, Claudia Araimo, Francesca De Simone, Carmelo Velardo, J-L Dugelay, and Touradj Ebrahimi. Subjective study of privacy filters in video surveillance. In Multimedia Signal Processing (MMSP), 2012 IEEE 14th International Workshop on, pages 378–382. Ieee, 2012.
- [147] Pavel Korshunov and Touradj Ebrahimi. Using face morphing to protect privacy. In Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference on, pages 208–213. IEEE, 2013.
- [148] Pavel Korshunov, Andrea Melle, Jean-Luc Dugelay, and Touradj Ebrahimi. Framework for objective evaluation of privacy filters. In Applications of Digital Image Processing XXXVI, volume 8856, page 88560T. International Society for Optics and Photonics, 2013.
- [149] Takashi Koshimizu, Tomoji Toriyama, and Noboru Babaguchi. Factors on the sense of privacy in video surveillance. In Proceedings of the 3rd ACM workshop on Continuous archival and retrieval of personal experiences, pages 35–44. ACM, 2006.

- [150] Sokol Kosta, Andrius Aucinas, Pan Hui, Richard Mortier, and Xinwen Zhang. Thinkair: Dynamic resource allocation and parallel execution in the cloud for mobile code offloading. In Infocom, 2012 Proceedings IEEE, pages 945–953. IEEE, 2012.
- [151] Hanna Krasnova, Oliver Günther, Sarah Spiekermann, and Ksenia Koroleva. Privacy concerns and identity in online social networks. Identity in the Information Society, 2(1):39–63, 2009.
- [152] William Kruskal and Frederick Mosteller. Representative sampling, ii: Scientific literature, excluding statistics. International Statistical Review/Revue Internationale de Statistique, pages 111–127, 1979.
- [153] Abhishek Kumar, Subham Kumar Gupta, Animesh Kumar Rai, and Sapna Sinha. Social networking sites and their security issues. International Journal of Scientific and Research Publications, 3(4):1–5, 2013.
- [154] Nanda Kumar and Izak Benbasat. Research note: the influence of recommendations and consumer reviews on evaluations of websites. Information Systems Research, 17(4):425–439, 2006.
- [155] Priya Kumar and Sarita Schoenebeck. The modern day baby book: Enacting good mothering and stewarding privacy on facebook. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, pages 1302–1312. ACM, 2015.
- [156] Priya C Kumar, Marshini Chetty, Tamara L Clegg, and Jessica Vitak. Privacy and security considerations for digital technology use in elementary schools. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pages 1–13, 2019.
- [157] K Hazel Kwon, Shin-II Moon, and Michael A Stefanone. Unspeaking on facebook? testing network effects on self-censorship of political expressions in social network sites. Quality & Quantity, 49(4):1417–1435, 2015.
- [158] Karen Lander, Vicki Bruce, and Harry Hill. Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. Applied Cognitive Psychology, 15(1):101–116, 2001.
- [159] Caroline Lang and Hannah Barton. Just untag it: Exploring the management of undesirable facebook photos. Computers in Human Behavior, 43:147–155, 2015.
- [160] Judith H Langlois and Lori A Roggman. Attractive faces are only average. Psychological science, 1(2):115–121, 1990.
- [161] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. arXiv preprint arXiv:1609.04802, 2016.
- [162] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In CVPR, volume 2, page 4, 2017.
- [163] Haein Lee, Hyejin Park, and Jinwoo Kim. Why do people share their context information on social network services? a qualitative study and an experimental study on users’ behavior of balancing perceived benefit and risk. International Journal of Human-Computer Studies, 71(9):862–877, 2013.

- [164] Jerry W Lee, Patricia S Jones, Yoshimitsu Mineyama, and Xinwei Esther Zhang. Cultural differences in responses to a likert scale. Research in nursing & health, 25(4):295–306, 2002.
- [165] Kun Chang Lee and Namho Chung. Understanding factors affecting trust in and satisfaction with mobile banking in korea: A modified delone and mclean’s model perspective. Interacting with computers, 21(5-6):385–392, 2009.
- [166] Ang Li, Qinghua Li, and Wei Gao. Privacycamera: Cooperative privacy-aware photographing with mobile phones. In 2016 13th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), pages 1–9. IEEE, 2016.
- [167] Congcong Li, Alexander C Loui, and Tsuhan Chen. Towards aesthetics: A photo quality assessment and photo selection system. In Proceedings of the 18th ACM international conference on Multimedia, pages 827–830. ACM, 2010.
- [168] Fenghua Li, Zhe Sun, Ang Li, Ben Niu, Hui Li, and Guohong Cao. Hideme: privacy-preserving photo sharing on social networks. In IEEE INFOCOM 2019-IEEE Conference on Computer Communications, pages 154–162. IEEE, 2019.
- [169] Fenghua Li, Zhe Sun, Ben Niu, Jin Cao, and Hui Li. An extended control framework for privacy-preserving photo sharing across different social networks. In 2019 International Conference on Computing, Networking and Communications (ICNC), pages 390–394. IEEE, 2019.
- [170] Na Li, Nan Zhang, and Sajal Das. Preserving relation privacy in online social network data. IEEE Internet Computing, 15(3):35–42, 2011.
- [171] Nan Li and Guanling Chen. Sharing location in online social networks. IEEE network, 24(5), 2010.
- [172] Tao Li and Lei Lin. Anonymousnet: Natural face de-identification with measurable privacy. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 0–0, 2019.
- [173] 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.
- [174] Yao Li, Alfred Kobsa, Bart P Knijnenburg, and MH Carolyn Nguyen. Cross-cultural privacy prediction. Proceedings on Privacy Enhancing Technologies, 2017(2):113–132, 2017.
- [175] Yifang Li, Wyatt Troutman, Bart P Knijnenburg, and Kelly Caine. Human perceptions of sensitive content in photos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 1590–1596, 2018.
- [176] 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.
- [177] Yifang Li, Nishant Vishwamitra, Hongxin Hu, Bart P. Knijnenburg, and Kelly Caine. Effectiveness and users’ experience of face blurring as a privacy protection for sharing photos via online social networks. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, volume 61. SAGE, 2017.

- [178] Yifang Li, Nishant Vishwamitra, Hongxin Hu, Bart P Knijnenburg, and Kelly Caine. Effectiveness and users' experience of face blurring as a privacy protection for sharing photos via online social networks. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, volume 61, pages 803–807. SAGE Publications Sage CA: Los Angeles, CA, 2017.
- [179] Yifang Li, Nishant Vishwamitra, Bart P Knijnenburg, Hongxin Hu, and Kelly Caine. Blur vs. block: Investigating the effectiveness of privacy-enhancing obfuscation for images. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on, pages 1343–1351. IEEE, 2017.
- [180] 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(2), 2017.
- [181] Kaitai Liang, Joseph K Liu, Rongxing Lu, and Duncan S Wong. Privacy concerns for photo sharing in online social networks. IEEE Internet Computing, 19(2):58–63, 2015.
- [182] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In European conference on computer vision, pages 740–755. Springer, 2014.
- [183] Heather Richter Lipford, Gordon Hull, Celine Latulipe, Andrew Besmer, and Jason Watson. Visible flows: Contextual integrity and the design of privacy mechanisms on social network sites. In Computational Science and Engineering, 2009. CSE'09. International Conference on, volume 4, pages 985–989. IEEE, 2009.
- [184] Eden Litt and Eszter Hargittai. Smile, snap, and share? a nuanced approach to privacy and online photo-sharing. Poetics, 42:1–21, 2014.
- [185] Yu-Lun Liu, Wei-Sheng Lai, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. Learning to see through obstructions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 14215–14224, 2020.
- [186] Mary Madden and Lee Rainie. Americans' attitudes about privacy, security and surveillance, May 2015.
- [187] Shah Mahmood. Online social networks: Privacy threats and defenses. In Security and Privacy Preserving in Social Networks, pages 47–71. Springer, 2013.
- [188] Naresh K Malhotra, Sung S Kim, and James Agarwal. Internet users' information privacy concerns (iupc): The construct, the scale, and a causal model. Information systems research, 15(4):336–355, 2004.
- [189] Aqdas Malik, Amandeep Dhir, and Marko Nieminen. Uses and gratifications of digital photo sharing on facebook. Telematics and Informatics, 33(1):129–138, 2016.
- [190] Banu Manav. Color-emotion associations and color preferences: A case study for residences. Color Research & Application, 32(2):144–150, 2007.
- [191] Alice E Marwick and Danah Boyd. I tweet honestly, i tweet passionately: Twitter users, context collapse, and the imagined audience. New media & society, 13(1):114–133, 2011.
- [192] Alice E Marwick and Danah Boyd. Networked privacy: How teenagers negotiate context in social media. New media & society, 16(7):1051–1067, 2014.
- [193] Winter Mason and Duncan J Watts. Financial incentives and the performance of crowds. ACM SigKDD Explorations Newsletter, 11(2):100–108, 2010.

- [194] Jennifer Dodorico McDonald. Measuring personality constructs: The advantages and disadvantages of self-reports, informant reports and behavioural assessments. Enquire, 1(1):1–19, 2008.
- [195] Richard McPherson, Reza Shokri, and Vitaly Shmatikov. Defeating image obfuscation with deep learning. arXiv preprint arXiv:1609.00408, 2016.
- [196] Jeremy Mendel, Christopher B Mayhorn, Jefferson B Hardee, Ryan T West, and Richard Pak. The effect of warning design and personalization on user compliance in computer security dialogs. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, volume 54, pages 1961–1965. SAGE Publications Sage CA: Los Angeles, CA, 2010.
- [197] Miriam J Metzger. Privacy, trust, and disclosure: Exploring barriers to electronic commerce. Journal of computer-mediated communication, 9(4):JCMC942, 2004.
- [198] Caroline Lancelot Miltgen and Dominique Peyrat-Guillard. Cultural and generational influences on privacy concerns: a qualitative study in seven european countries. European Journal of Information Systems, 23(2):103–125, 2014.
- [199] Kimberly J Mitchell, David Finkelhor, and Janis Wolak. Risk factors for and impact of online sexual solicitation of youth. Jama, 285(23):3011–3014, 2001.
- [200] Adam D Moore. Toward informational privacy rights. San Diego L. Rev., 44:809, 2007.
- [201] Tyler Moore, Richard Clayton, and Ross Anderson. The economics of online crime. Journal of Economic Perspectives, 23(3), 2009.
- [202] Masahiro Mori, Karl F MacDorman, and Norri Kageki. The uncanny valley [from the field]. IEEE Robotics & Automation Magazine, 19(2):98–100, 2012.
- [203] Kyle B Murray and Gerald Häubl. Freedom of choice, ease of use, and the formation of interface preferences. 2010.
- [204] Nicole L Muscanell and Rosanna E Guadagno. Make new friends or keep the old: Gender and personality differences in social networking use. Computers in Human Behavior, 28(1):107–112, 2012.
- [205] Sucheta Nadkarni and Reetika Gupta. A task-based model of perceived website complexity. Mis Quarterly, pages 501–524, 2007.
- [206] Sophie J Nightingale, Kimberley A Wade, and Derrick G Watson. Can people identify original and manipulated photos of real-world scenes? Cognitive research: principles and implications, 2(1):30, 2017.
- [207] Helen Nissenbaum. Privacy as contextual integrity. Wash. L. Rev., 79:119, 2004.
- [208] Angelo Nodari, Marco Vanetti, and Ignazio Gallo. Digital privacy: Replacing pedestrians from google street view images. In Pattern Recognition (ICPR), 2012 21st International Conference on, pages 2889–2893. IEEE, 2012.
- [209] Patricia A Norberg, Daniel R Horne, and David A Horne. The privacy paradox: Personal information disclosure intentions versus behaviors. Journal of consumer affairs, 41(1):100–126, 2007.
- [210] Glen J Nowak and Joseph Phelps. Understanding privacy concerns. an assessment of consumers’ information-related knowledge and beliefs. Journal of Direct Marketing, 6(4):28–39, 1992.

- [211] Anne Oeldorf-Hirsch and S Shyam Sundar. Online photo sharing as mediated communication. In annual conference of the International Communication Association, Singapore, 2010.
- [212] 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.
- [213] Judith S Olson, Jonathan Grudin, and Eric Horvitz. A study of preferences for sharing and privacy. In CHI’05 extended abstracts on Human Factors in Computing Systems, pages 1985–1988. ACM, 2005.
- [214] Tribhuvanesh Orekondy, Mario Fritz, and Bernt Schiele. Connecting pixels to privacy and utility: Automatic redaction of private information in images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 8466–8475, 2018.
- [215] Tribhuvanesh Orekondy, Bernt Schiele, and Mario Fritz. Towards a visual privacy advisor: Understanding and predicting privacy risks in images. In 2017 IEEE International Conference on Computer Vision (ICCV), pages 3706–3715. IEEE, 2017.
- [216] José Ramón Padilla-López, Alexandros Andre Chaaraoui, Feng Gu, and Francisco Flórez-Revuelta. Visual privacy by context: proposal and evaluation of a level-based visualisation scheme. Sensors, 15(6):12959–12982, 2015.
- [217] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10):1345–1359, 2009.
- [218] Zizi Papacharissi. The virtual geographies of social networks: a comparative analysis of facebook, linkedin and asmallworld. New media & society, 11(1-2):199–220, 2009.
- [219] Zizi Papacharissi and Paige L Gibson. Fifteen minutes of privacy: Privacy, sociality, and publicity on social network sites. In Privacy online, pages 75–89. Springer, 2011.
- [220] Omkar M Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. In BMVC, volume 1, page 6, 2015.
- [221] Delroy L Paulhus. Measurement and control of response bias. 1991.
- [222] Delroy L Paulhus and Simine Vazire. The self-report method. Handbook of research methods in personality psychology, 1:224–239, 2007.
- [223] Eyal Peer, Joachim Vosgerau, and Alessandro Acquisti. Reputation as a sufficient condition for data quality on amazon mechanical turk. Behavior research methods, 46(4):1023–1031, 2014.
- [224] Tiffany A Pempek, Yevdokiya A Yermolayeva, and Sandra L Calvert. College students’ social networking experiences on facebook. Journal of applied developmental psychology, 30(3):227–238, 2009.
- [225] Andrew Perrin and Jingjing Jiang. About a quarter of u.s. adults say they are ‘almost constantly’ online, 2018. Retrieved February, 2019 from <http://www.pewresearch.org/fact-tank/2018/03/14/about-a-quarter-of-americans-report-going-online-almost-constantly/>.
- [226] Sandra Petronio. Boundaries of privacy: Dialectics of disclosure. Suny Press, 2012.

- [227] Pew Research Center. Pew research center demographic question, 2015. Retrieved April 28, 2018 from <http://assets.pewresearch.org/wp-content/uploads/sites/12/2015/03/Demographic-Questions-Web-and-Mail-English-3-20-2015.pdf>.
- [228] Pew Research Center. Internet/broadband fact sheet, 2018. Retrieved February, 2019 from <http://www.pewinternet.org/fact-sheet/internet-broadband/>.
- [229] Joseph Phelps, Glen Nowak, and Elizabeth Ferrell. Privacy concerns and consumer willingness to provide personal information. Journal of public policy & marketing, 19(1):27–41, 2000.
- [230] Mark R Phillips, Bradley D McAuliff, Margaret Bull Kovera, and Brian L Cutler. Double-blind photoarray administration as a safeguard against investigator bias. Journal of Applied Psychology, 84(6):940, 1999.
- [231] Susie Poppick. 10 social media blunders that cost a millennial a job — or worse, 2014. Retrieved October 5, 2018 from <http://time.com/money/3019899/10-facebook-twitter-mistakes-lost-job-millennials-viral/>.
- [232] 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 Proceedings of the Fourteenth USENIX Conference on Usable Privacy and Security, pages 143–157. USENIX Association, 2018.
- [233] Joseph P Redden. Reducing satiation: The role of categorization level. Journal of Consumer Research, 34(5):624–634, 2008.
- [234] Elissa M Redmiles, Sean Kross, Alisha Pradhan, and Michelle L Mazurek. How well do my results generalize? comparing security and privacy survey results from mturk and web panels to the us. Technical report, 2017.
- [235] Chi-Hyoung Rhee and C LEE. Cartoon-like avatar generation using facial component matching. Int. J. of Multimedia and Ubiquitous Engineering, 8(4):69–78, 2013.
- [236] Victoria Richards. Paedophile websites steal half their photos from social media sites like facebook, 2015. Retrieved October 5, 2018 from <https://www.independent.co.uk/news/world/australasia/paedophile-websites-steal-half-their-photos-from-social-media-sites-like-facebook-a6673191.html>.
- [237] Jessica Ringrose, Laura Harvey, Rosalind Gill, and Sonia Livingstone. Teen girls, sexual double standards and ‘sexting’: Gendered value in digital image exchange. Feminist theory, 14(3):305–323, 2013.
- [238] Hoyle Roberto, Luke Stark, Ismail Qatrunnada, David Crandall, Apu Kapadia, and Denise Anthony. Privacy norms and preferences for photos posted online.
- [239] Joel Ross, Lilly Irani, M Silberman, Andrew Zaldivar, and Bill Tomlinson. Who are the crowdworkers?: shifting demographics in mechanical turk. In CHI’10 extended abstracts on Human Factors in Computing Systems, pages 2863–2872. ACM, 2010.
- [240] Joel Ross, Andrew Zaldivar, Lilly Irani, and Bill Tomlinson. Who are the turkers? worker demographics in amazon mechanical turk. Department of Informatics, University of California, Irvine, USA, Tech. Rep, 2009.

- [241] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 115(3), 2015.
- [242] Mukesh Saini, Pradeep K Atrey, Sharad Mehrotra, and Mohan Kankanhalli. W3-privacy: understanding what, when, and where inference channels in multi-camera surveillance video. Multimedia Tools and Applications, 68(1):135–158, 2014.
- [243] Florian Schaub, Rebecca Balebako, Adam L Durity, and Lorrie Faith Cranor. A design space for effective privacy notices. In Eleventh Symposium On Usable Privacy and Security ({SOUPS} 2015), pages 1–17, 2015.
- [244] Stefan R Schweinberger, Esther C Pickering, A Mike Burton, and Jürgen M Kaufmann. Human brain potential correlates of repetition priming in face and name recognition. Neuropsychologia, 40(12):2057–2073, 2002.
- [245] Stefan R Schweinberger, Esther C Pickering, Ines Jentzsch, A Mike Burton, and Jürgen M Kaufmann. Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. Cognitive Brain Research, 14(3):398–409, 2002.
- [246] Christien Marie Seamon. Self-esteem, sex differences, and self-disclosure: A study of the closeness of relationships. 2003.
- [247] Peter Seddon and Min-Yen Kiew. A partial test and development of delone and mclean’s model of is success. Australasian Journal of Information Systems, 4(1), 1996.
- [248] Sadia Shamma and Md Yusuf Sarwar Uddin. Towards privacy-aware photo sharing using mobile phones. In Electrical and Computer Engineering (ICECE), 2014 International Conference on, pages 449–452. IEEE, 2014.
- [249] Robert Sheridan. Malingering: Yes, it may get you fired, 2014. Retrieved September 19, 2018 from <https://www.mintz.com/insights-center/viewpoints/2014-05-malingering-yes-it-may-get-you-fired>.
- [250] Yoshinari Shirai, Yasue Kishino, Takayuki Suyama, and Shin Mizutani. Pasnic: a thermal based privacy-aware sensor node for image capturing. In Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers, pages 202–205. ACM, 2019.
- [251] 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. ACM, 2018.
- [252] Signal. Why use signal?, 2020. Retrieved July 5, 2020 from <https://signal.org/en/>.
- [253] M Six Silberman, Bill Tomlinson, Rochelle LaPlante, Joel Ross, Lilly Irani, and Andrew Zaldivar. Responsible research with crowds: pay crowdworkers at least minimum wage. Commun. ACM, 61(3):39–41, 2018.
- [254] Meredith M Skeels and Jonathan Grudin. When social networks cross boundaries: a case study of workplace use of facebook and linkedin. In Proceedings of the ACM 2009 international conference on Supporting group work, pages 95–104. ACM, 2009.
- [255] Manya Sleeper, Rebecca Balebako, Sauvik Das, Amber Lynn McConahy, Jason Wiese, and Lorrie Faith Cranor. The post that wasn’t: exploring self-censorship on facebook. In Proceedings of the 2013 conference on Computer supported cooperative work, pages 793–802. ACM, 2013.



- [256] Aaron Smith and Monica Anderson. Social media use in 2018, 2018. Retrieved December 11, 2018 from <http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/>.
- [257] Kit Smith. 47 incredible facebook statistics and facts, 2018. Retrieved December 11, 2018 from <https://www.brandwatch.com/blog/47-facebook-statistics/>.
- [258] Snapeditor. Snapchat effects: how to use lenses & filters for face effects, January 2017. Retrieved April 23, 2017 from <https://snapchat.photography/snapchat-effects/>.
- [259] Donna Spencer. Card sorting: Designing usable categories. Rosenfeld Media, 2009.
- [260] Donna Spencer and Todd Warfel. Card sorting: a definitive guide. Boxes and Arrows, 2, 2004.
- [261] Julie Spencer-Rodgers, Melissa J Williams, David L Hamilton, Kaiping Peng, and Lei Wang. Culture and group perception: Dispositional and stereotypic inferences about novel and national groups. Journal of personality and social psychology, 93(4):525, 2007.
- [262] Nadia Spock. Snapchat gives a voice to survivors of sexual abuse. Blog, July 2016. Retrieved December 17, 2018 from <http://news.wgbh.org/2016/07/21/snapchat-gives-voice-survivors-sexual-abuse>.
- [263] Anna Cinzia Squicciarini, Dan Lin, Smitha Sundareswaran, and Joshua Wede. Privacy policy inference of user-uploaded images on content sharing sites. IEEE transactions on knowledge and data engineering, 27(1):193–206, 2015.
- [264] Agrima Srivastava and G Geethakumari. A privacy settings recommender system for online social networks. In Recent Advances and Innovations in Engineering (ICRAIE), 2014, pages 1–6. IEEE, 2014.
- [265] Eugene F Stone, Hal G Gueutal, Donald G Gardner, and Stephen McClure. A field experiment comparing information-privacy values, beliefs, and attitudes across several types of organizations. Journal of applied psychology, 68(3):459, 1983.
- [266] Katherine Strater and Heather Richter Lipford. Strategies and struggles with privacy in an online social networking community. In Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction-Volume 1, pages 111–119. British Computer Society, 2008.
- [267] David L Streiner. Finding our way: an introduction to path analysis. The Canadian Journal of Psychiatry, 50(2):115–122, 2005.
- [268] Fred Stutzman and Jacob Kramer-Duffield. Friends only: examining a privacy-enhancing behavior in facebook. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pages 1553–1562. ACM, 2010.
- [269] Jose M Such, Joel Porter, Sören Preibusch, and Adam Joinson. Photo privacy conflicts in social media: A large-scale empirical study. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pages 3821–3832. ACM, 2017.
- [270] Weiwei Sun, Jiantao Zhou, Ran Lyu, and Shuyuan Zhu. Processing-aware privacy-preserving photo sharing over online social networks. In Proceedings of the 24th ACM international conference on Multimedia, pages 581–585, 2016.
- [271] John A Swets. Signal detection and recognition in human observers: Contemporary readings. 1964.

- [272] Yasuhiro Tanaka, Akihisa Kodate, Yu Ichifuji, and Noboru Sonehara. Relationship between willingness to share photos and preferred level of photo blurring for privacy protection. In Proceedings of the ASE BigData & SocialInformatics 2015, page 33. ACM, 2015.
- [273] Kurt Thomas, Chris Grier, and David M Nicol. unfriendly: Multi-party privacy risks in social networks. In International Symposium on Privacy Enhancing Technologies Symposium, pages 236–252. Springer, 2010.
- [274] Sara E Thomas. “what should i do?”: Young women’s reported dilemmas with nude photographs. Sexuality Research and Social Policy, 15(2):192–207, 2018.
- [275] Matt Tierney, Ian Spiro, Christoph Bregler, and Lakshminarayanan Subramanian. Cryptagram: photo privacy for online social media. In Proceedings of the first ACM conference on Online social networks, pages 75–88. ACM, 2013.
- [276] Cody Toombs. Snapseed v2.17 adds tools to make wild facial adjustments, create double exposures, and more. Blog, March 2017. <http://www.androidpolice.com/2017/03/22/snapseed-v2-17-adds-tools-to-make-wild-facial-adjustments-create-double-exposures-and-more-apk-download/>.
- [277] Lisa Torrey and Jude Shavlik. Transfer learning. In Handbook of research on machine learning applications and trends: algorithms, methods, and techniques, pages 242–264. IGI Global, 2010.
- [278] Lam Tran, Deguang Kong, Hongxia Jin, and Ji Liu. Privacy-cnnet: A framework to detect photo privacy with convolutional neural network using hierarchical features. In AAAI, pages 1317–1323, 2016.
- [279] Tom Tullis and Larry Wood. How many users are enough for a card-sorting study. In Proceedings UPA, volume 2004, 2004.
- [280] Joanne R Ullman and N Clayton Silver. Perceived effectiveness of potential music piracy warnings. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, volume 62, pages 1353–1357. SAGE Publications Sage CA: Los Angeles, CA, 2018.
- [281] United States Census Bureau. U.s. census bureau quickfacts selected: United states, July 2016. <https://www.census.gov/quickfacts/fact/table/US/PST045216>.
- [282] Evgeniy Upenik, Pinar Akyazi, Mehmet Tuzmen, and Touradj Ebrahimi. Inpainting in omnidirectional images for privacy protection. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 2487–2491. IEEE, 2019.
- [283] Carmen Ruiz Vicente, Dario Freni, Claudio Bettini, and Christian S Jensen. Location-related privacy in geo-social networks. IEEE Internet Computing, 15(3):20–27, 2011.
- [284] Nishant Vishwamitra, Yifang Li, Kevin Wang, Hongxin Hu, Kelly Caine, and Gail-Joon Ahn. Towards pii-based multiparty access control for photo sharing in online social networks. In Proceedings of the 22nd ACM on Symposium on Access Control Models and Technologies, pages 155–166. ACM, 2017.
- [285] Jessica Vitak and Jinyoung Kim. You can’t block people offline: Examining how facebook’s affordances shape the disclosure process. In Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing, pages 461–474. ACM, 2014.

- [286] Emanuel von Zezschwitz, Alexander De Luca, and Heinrich Hussmann. Filter selection and evaluation. 2015.
- [287] Tim Wafa. How the lack of prescriptive technical granularity in hipaa has compromised patient privacy. *N. Ill. UL Rev.*, 30:531, 2009.
- [288] Robin Wakefield. The influence of user affect in online information disclosure. *The Journal of Strategic Information Systems*, 22(2):157–174, 2013.
- [289] 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, 2009.
- [290] Yang Wang, Pedro Giovanni Leon, Kevin Scott, Xiaoxuan Chen, Alessandro Acquisti, and Lorrie Faith Cranor. Privacy nudges for social media: an exploratory facebook study. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 763–770, 2013.
- [291] Yang Wang, Gregory Norice, and Lorrie Faith Cranor. Who is concerned about what? a study of american, chinese and indian users’ privacy concerns on social network sites. In *International Conference on Trust and Trustworthy Computing*, pages 146–153. Springer, 2011.
- [292] Samuel D Warren and Louis D Brandeis. The right to privacy. *Harvard law review*, pages 193–220, 1890.
- [293] Alan F Westin. Privacy and freedom. *Washington and Lee Law Review*, 25(1):166, 1968.
- [294] Wikipedia. Forbes celebrity 100, September 2017. [https://en.wikipedia.org/wiki/Forbes\\_Celebrity\\_100](https://en.wikipedia.org/wiki/Forbes_Celebrity_100).
- [295] Heng Xu. The effects of self-construal and perceived control on privacy concerns. *ICIS 2007 proceedings*, page 125, 2007.
- [296] Heng Xu, Na Wang, and Jens Grossklags. Privacy by redesign: Alleviating privacy concerns for third-party apps. 2012.
- [297] 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, 2017.
- [298] George Yee and Larry Korba. Comparing and matching privacy policies using community consensus. In *Proceedings, 16th IRMA International Conference, San Diego, California*. Citeseer, 2005.
- [299] Andrew W Yip and Pawan Sinha. Contribution of color to face recognition. *Perception*, 31(8):995–1003, 2002.
- [300] Alyson Leigh Young and Anabel Quan-Haase. Privacy protection strategies on facebook: The internet privacy paradox revisited. *Information, Communication & Society*, 16(4):479–500, 2013.
- [301] YouTube. Face blurring: when footage requires anonymity, 2012. Retrieved December, 2018 from <https://youtube.googleblog.com/2012/07/face-blurring-when-footage-requires.html>.
- [302] Hyunwoo Yu, Jaemin Lim, Kiyeon Kim, and Suk-Bok Lee. Pinto: enabling video privacy for commodity iot cameras. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pages 1089–1101, 2018.

- [303] Jun Yu, Zhenzhong Kuang, Zhou Yu, Dan Lin, and Jianping Fan. Privacy setting recommendation for image sharing. In Machine Learning and Applications (ICMLA), 2017 16th IEEE International Conference on, pages 726–730. IEEE, 2017.
- [304] Jun Yu, Zhenzhong Kuang, Baopeng Zhang, Wei Zhang, Dan Lin, and Jianping Fan. Leveraging content sensitiveness and user trustworthiness to recommend fine-grained privacy settings for social image sharing. IEEE Transactions on Information Forensics and Security, 13(5):1317–1332, 2018.
- [305] Xiaoyi Yu, Kenta Chinomi, Takashi Koshimizu, Naoko Nitta, Yoshimichi Ito, and Noboru Babaguchi. Privacy protecting visual processing for secure video surveillance. In Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on, pages 1672–1675. IEEE, 2008.
- [306] Lin Yuan, Pavel Korshunov, and Touradj Ebrahimi. Secure jpeg scrambling enabling privacy in photo sharing. In Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on, volume 4, pages 1–6. IEEE, 2015.
- [307] Sergej Zerr, Stefan Siersdorfer, and Jonathon Hare. Picalert! a system for privacy-aware image classification and retrieval. In Proceedings of the 21st ACM international conference on Information and knowledge management, pages 2710–2712, 2012.
- [308] Michael T Zimmer. Personal information and the design of vehicle safety communication technologies: An application of privacy as contextual integrity. Proceedings of AAAS Science & Technology in Society, pages 222–226, 2005.