# Towards A Taxonomy of Content Sensitivity and Sharing Preferences for Photos

**Yifang Li, Nishant Vishwamitra, Hongxin Hu, Kelly Caine**
Clemson University, Clemson, SC, USA
{yifang2, nvishwa}@g.clemson.edu {hongxih, caine}@clemson.edu

## ABSTRACT

Determining which photos are sensitive is difficult. Although emerging computer vision systems can label content items, previous attempts to distinguish private or sensitive content fall short. There is no human-centered taxonomy that describes what content is sensitive or how sharing preferences for content differs across recipients. To fill this gap, we introduce a new sensitive content elicitation method which surmounts limitations of previous approaches, and, using this new method, collected sensitive content from 116 participants. We also recorded participants' sharing preferences with 20 recipient groups. Next, we conducted a card sort 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 recipients. This taxonomy can serve as a framework for understanding photo privacy, which can, in turn, inform new photo privacy protection mechanisms.

### Author Keywords

Privacy, security, photo privacy, sensitive content

### CCS Concepts

•**Security and privacy** → **Human and societal aspects of security and privacy;**

## 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 [67]. 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 [38]. 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 [23, 55], 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 1).
- 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.


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

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376498

## BACKGROUND

### Photo Sharing Privacy Protection

Photo shared on SNSs usually expose rich information about the people in the photo, their activity, objects, or the environment. Some content can be sensitive (e.g., an embarrassing facial expression [3]), or even introduce significant safety risks (e.g., location leakage [36]). In light of these risks, both SNSs and researchers in privacy and security have sought to develop photo privacy protection mechanisms. Existing mechanisms are in line with a simplified privacy framework, which identified two fundamental elements at play in privacy: **information content** and **recipient** [18]. Adjusting either element will affect photo privacy. This simplified model complements contextual integrity by triangulating from a focused, empirical perspective [9]. Most existing SNSs protect users' privacy via controlling the groups of recipients (e.g., Facebook, Flickr). Concerning the other element–content–researchers have developed photo privacy protection systems that make part of the photo content invisible to viewers [5, 38, 67, 73].

To develop an effective photo privacy protection mechanism, for the first element–content–we must know 1) what is the sensitive content to be obfuscated in a photo, 2) which obfuscation is effective and usable. For the second element–photo recipient–we must understand 3) what are users' sharing preferences in terms of different recipients. While existing work explores the second aspect of the first element–obfuscation methods–we focus on the other two prerequisite aspects, determining what content is sensitive and investigating users' sharing preferences with various recipient groups.

### First Element: Content

#### *Sensitive Content*

We have some hints from prior work about the kinds of content in a photo that people consider sensitive (summarized in Table 1). 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, 10, 45]. Sensitive features extracted from participants' photos and users' comments via machine learning indicate that people, landscape, and certain places and events are sensitive [16]. Certain objects, backgrounds [4], and phone screens [36] are also common concerns. When cameras are ubiquitous, such as in life-logging, monitor screens and irrelevant persons in photos lead to privacy concerns [36]. People are also concerned about revealing photos that contain text such as their address, organizational affiliation, and email address [7, 29].

The most comprehensive study to date using a ML approach examining content sensitivity is work that claims to identify 268 privacy-sensitive object classes [86, 87]. 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 a 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 [70]. 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 objece classes present in existing photo datasets, such as ImageNet [65] and MS COCO [50], 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 [31].

#### *Obfuscation Methods*

There is adequate knowledge on the second aspect of content–obfuscation methods [47, 48, 49]. Researchers introduced two promising obfuscations, avatar (replacing a person with an avatar that hides the identity but preserves facial expression and gesture) and inpainting (completely removing a person), that overcome the trade-off between the effectiveness and the utility of obfuscation [49]. While this work only focuses on people in photos, other work applies obfuscations on scene elements finding that silhouette works well on objects [33].

### Second Element: Recipients

Aside from the first element–sensitive content and obfuscation method–the second element is photo recipient. People have different levels of privacy preference for various groups of

| Category | Sensitive content & Citation | Research method (sample size) |
|---|---|---|
| Identity | Photo owner [10, 16] | Focus group (14); Machine learning on photos participants identify as private (16) |
| | Family members [3] | Interview (37) |
| | Children [3, 41] | Interview (37); Analyzing the automatically-generated tags for photos in Flickr data sets |
| Nudity | [41, 56] | Analyzing the automatically-generated tags for photos in Flickr data sets; EU Data Protection Directive 95/46/EC, EUS Privacy Act of 1974, SNSs rules (N/A) |
| Factors that harm impression management | Unflattering/embarrassing shots [3] | Interview (37) |
| | Activity that may be misinterpreted [3] | Interview (37) |
| | Presentation management [36] | Field deployment and survey (36) |
| | Environment [4, 16] | Interview (15); Machine learning on photos participants identify as private (16) |
| | Event [4, 10, 16] | Interview (15); Focus group (14); Machine learning on photos participants identify as private (16) |
| Factors that reveal personal information | Monitor screen [36, 41] | Field deployment and survey (36); Analyzing the automatically-generated tags for photos in Flickr data sets; |
| | Location [36] | Field deployment and survey (36) |
| | Written information [36] | Field deployment and survey (36) |
| Illegal | Illegal activity [10] | Focus group (14) |
| | Copyright [3] | Interview (37) |
| Photo quality | Technically flawed photo [3, 36, 41] | Interview (37); field deployment and survey (36) |
| No need to share | Irrelevant to viewers [3] | Interview (37) |

**Table 1. Categories of sensitive content in photos from prior work. The number in the parenthesis in the Research Method column is the sample size from the study identifying each content category. Notably, the sample size reported in this paper is three times the largest sample size of any prior work.**

| Category | Recipients |
|---|---|
| Private | Only me |
| Family | Spouse/significant others [17, 22] |
| | Household members [17, 23, 22, 30, 43, 59] |
| | Relatives [17, 59] |
| Friends | Close friends [17, 30, 43, 59, 79] |
| | Normal friends [17, 23, 22, 59, 85] |
| Colleagues & Classmates | Colleagues, co-workers [17, 22, 43] |
| | Classmates [79] |
| | Supervisors [17, 22] |
| Acquaintances | SNS friends that haven't met offline [85] |
| | Acquaintances [43, 59] |
| | Loose acquaintances [30] |

**Table 2. Summary of categories of recipients from prior literature**

recipients [23, 43, 55]. For example, people may be unwilling to share party photos with their supervisor or co-worker, but they feel comfortable sharing with friends and family. Hence, identifying sensitive content considering each recipient group is important. We summarize different recipient groups from prior literature in Table 2 and adapt them in our method design.

## METHOD

### 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.

#### *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 [61, 60] 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 [64]. Additionally, MTurk recruitment results in a more diverse sample compared to standard Internet sampling and college sampling [15]. The data in studies using MTurk are as reliable as those obtained via other recruitment methods [19]. Moreover, MTurk is commonly used successfully for conducting privacy research [20, 62, 83]. We paid participants $4.00 to complete the 30-minute session which is in line with the recommendation in [68] 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 [58]. 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 [44] 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 [61, 60]. 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.

*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). These recipients were developed based on prior work (Table 2) with additional granularity in the form of close and not close as suggested by [43]. 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 [81].

*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 3.

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), colleagues/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.

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

**Table 3. Recipient groups used in our study**

### 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 [72, 8]. 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 [8].

*Participants*

We recruited 14 participants (in line with the sample size recommended in [78]) 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., [71]) there was no overlap in participants between study one, where participants provided content and study two, where participants sorted content.

*Procedure*

Each participant first saw digital cards in XSort, a computer program designed to collect card sort data. All 181 cards were

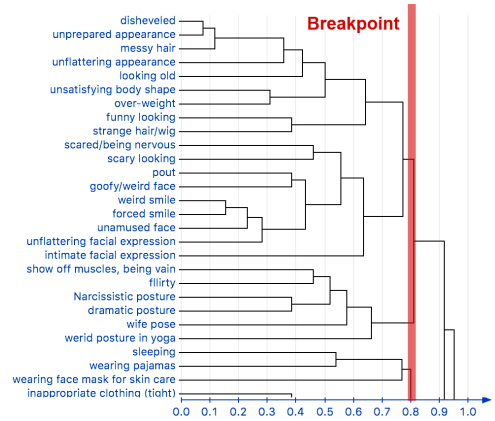


**Figure 1. A part of the dendrogram. All items in this sort are listed vertically. Items placed next to each other vertically are more similar. The horizontal line from each item joins other items vertically, showing where items are grouped at higher levels of relationship [8].**

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.

## 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 4). 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.

### 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 [8]. 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 [8]?" The results are visualized in a dendrogram. Due to space limitations, Figure 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 4). 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 1, our results are much more granular in detail, and therefore simultaneously expand and refine those categories. For example, whereas prior work [41, 56] 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: while prior work [3, 41] 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.

### 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 2 which illustrates the overall likelihood to share

| Category | Example |
|---|---|
| Nudity / Sexual (113) | - Genitals; naked person; butt crack; naked buttocks; breasts; same-sex (as the photo owner) nudity; cleavage; bare back; shirtless; masturbation; sexual action; erotic online photo; sexualized objects; sexual motion with a statue; suggestive posture |
| Mitigated (10) | - Breastfeeding; bent over showing behind; kissing |
| Close up (6) | - Unflattering close up of body parts |
| Irresponsible to child / pet (8) | - Child in a dangerous situation; child in inappropriate clothes; naked child; delinquent pet owner |
| Bad characters / unlawful / criminal (27) | - Infidelity/cheating; photo owner in a dangerous situation; illegal drugs; being physically abused; mug shot/getting 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; being flirty; Narcissistic posture; dramatic posture; photo owner's wife's pose (no sexual meaning); weird posture in yoga |
| Not professional at work (9) | - Activities that break work rules; negative attitude towards work; looking unprofessional at work; co-workers kissing |
| Sleep and grooming (5) | - Sleeping; wearing pajamas; wearing face mask for skin care |
| Clothing (33) | - Tight clothing; revealing clothing (swimsuit; underwear); wearing body-shaping corset; changing clothes; unfashionable outfit; tacky outfit; wearing bib for dining; cross-dressing; wearing a 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 / memes; vulgar / explicit texts / memes; politically incorrect texts / memes; racist texts/memes; violation of religious dogma |
| Other people (74) | - Grandparents; family members; significant other; step-parents; step-children; young family members; older children; friends; family member who passed away; photo owner's children; estranged people; ex-significant other; people who are unacceptable by photo owner's family |
| Personal moment (14) | - Affectionate moment with significant other; affectionate moment with friends |
| Event (5) | - Family event / party; children's beauty pageant; funeral |
| Photo owner (18) | - Photo owner's non-sensitive body parts; photo owner by 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 to be private; saving others' photos without permission |
| Personal identifiable information (24) | - Vehicle license plate; driver's 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 4. 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?".**

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

[1] $d$ represents Cohen's d.

### 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 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.

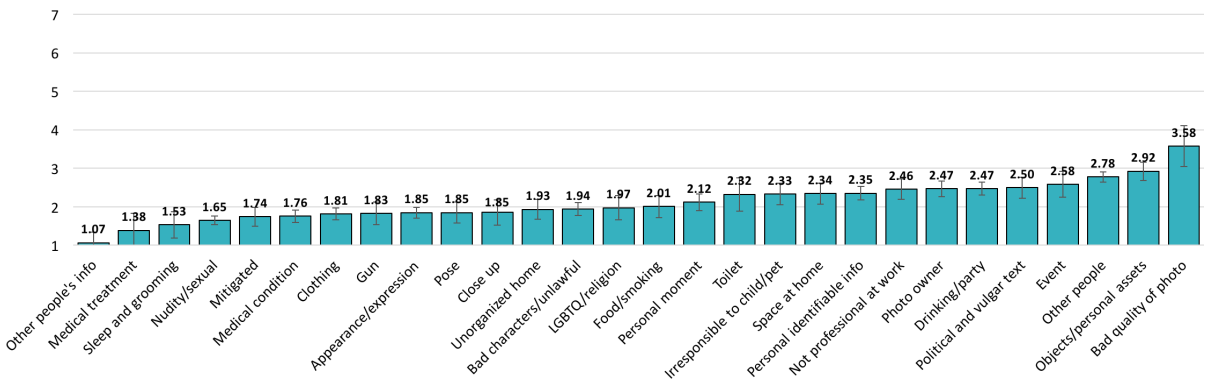


**Figure 2. Participants' likelihood to share each sensitive content category across all recipient groups.**

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), 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 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 2 and 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 $p > .05$).

## DISCUSSION

### 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 [55]. 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 [40]. 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 [21]

Though sharing information with work-related recipients on SNSs is prevalent because of the specific sharing needs for workplace SNS use [12, 69], 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

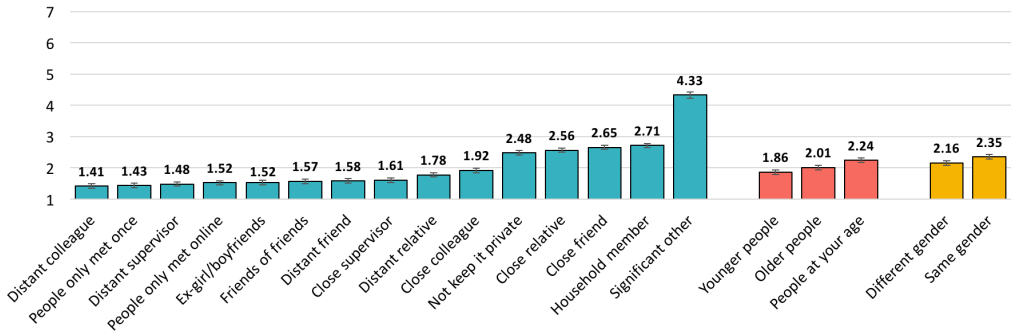


**Figure 3. Participants' likelihood to share with each recipient across all sensitive content categories.**

social isolation in the workplace suggesting that people find it difficult to establish friendships with their colleagues or supervisors [39]. Moreover, some content may be sensitive because it has the potential to reveal "white lies" or remove "plausible deniability" at work [66]. 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' [40], such as *distant friends or relatives*, *friends of friends*, or *people they only met once or online*.

### 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 percieved 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 [26]. It also suggests that people try to avoid social tension caused by unauthorized sharing and saving of others' photos [13]. People also generally respect others' privacy concerns [10]. 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 [74]. Existing privacy controls on SNS are unable to protect a user from content leakage by their friends [74], hence emerging work has developed multi-party privacy control mechanisms to alleviate this problem [37]. 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 [56, 87] 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 [75]. However, leakage of these photos damages people's impression management and reputation, and even leads to social ostracism, depression, and suicide [63]. Second, sharing sexually suggestive photos may become a potential threat to physical safety via off-line contact, [51]. Last, disseminating other people's nude photos violates the law and may get photo uploaders into legal trouble.

Aligning with previous work [54], *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 [32]. 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 [25, 35], 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 2, 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 [11, 53].

### 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 [52], it may be subjective and dependent on complex social, cultural, and historical factors [24, 28, 52]. 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 [24]. What some people are comfortable sharing others might consider a threat to the privacy [80]. 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 [77]. 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) [6]. Some other commonly identified categories of private items in personal photos include human faces, sensitive text, and objects such as cars and animals) [34].

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 [70]. Moreover, we know that there is a desire to use machine learning approaches to find consistencies regarding sensitive content [86, 87]. 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 [84], 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.

### 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 [86, 87]. 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 [86, 87]. 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 [57, 76]. 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 [86, 87]. 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 [86, 87]. 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.

### 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 [27]. 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 [70]. 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 making it difficult for them to sort through all potential recipients and make decisions about desirable recipients every time they upload a photo [14]. 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 [42].

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 [49]) that when applied, would prevent some viewers from seeing the sensitive content as shown in Figure 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.

## LIMITATIONS AND DIRECTIONS FOR 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. participants 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., [46, 82]) 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. However, reviewing Table 4, most items seem intuitively to fit within each category.

Another limitation is that we were only able to collect sensitive content that participants would identify in one of three ways:

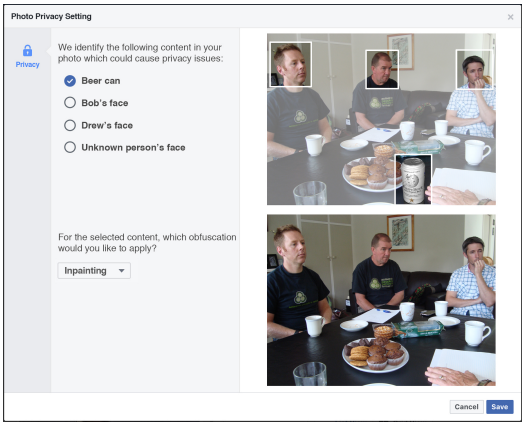


**Figure 4. Example interface of content detection and obfuscation.**

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 [86, 87].

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.

## 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.

## ACKNOWLEDGEMENT

The research was supported by the National Science Foundation under grant no.1527421. We thank Dr. Bart Knijnenburg and colleagues from the HATlab for the suggestions that improved the study design and data analysis and the reviewers for suggestions that improved the final paper. Finally, we are grateful to the participants of our online study.

## REFERENCES


[1] James D Abbey and Margaret G Meloy. 2017. Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management* 53 (2017), 63–70.

[2] Patricia Sanchez Abril. 2007. A (My) space of one's own: on privacy and online social networks. *Nw. J. Tech. & Intell. Prop.* 6 (2007), 73.

[3] Anne Adams, Sally Jo Cunningham, and Masood Masoodian. 2007. Sharing, privacy and trust issues for photo collections. (2007).

[4] Shane Ahern, Dean Eckles, Nathaniel S Good, Simon King, Mor Naaman, and Rahul Nair. 2007. Over-exposed?: privacy patterns and considerations in online and mobile photo sharing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM.

[5] Rawan Alharbi, Mariam Tolba, Lucia C Petito, Josiah Hester, and Nabil Alshurafa. 2019. 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 (2019), 72.

[6] David S Allison and Miriam AM Capretz. 2011. Furthering the growth of cloud computing by providing privacy as a service. In *International Conference on Information and Communication on Technology*. Springer, 64–78.

[7] Tuomas Aura, Thomas A Kuhn, and Michael Roe. 2006. Scanning electronic documents for personally identifiable information. In *Proceedings of the 5th ACM workshop on Privacy in electronic society*. ACM.

[8] Kathy Baxter, Catherine Courage, and Kelly Caine. 2015. *Understanding your users: A practical guide to user research methods*. Morgan Kaufmann.

[9] Sebastian Benthall, Seda Gürses, Helen Nissenbaum, Cornell Tech, and NYU Steinhardt MCC. 2017. *Contextual integrity through the lens of computer science*. Now Publishers.

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

[11] Leyla Bilge, Thorsten Strufe, Davide Balzarotti, and Engin Kirda. 2009. 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*. ACM, 551–560.

[12] Jens Binder, Andrew Howes, and Alistair Sutcliffe. 2009. 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*. ACM, 965–974.

[13] Jens F Binder, Andrew Howes, and Daniel Smart. 2012. Harmony and tension on social network sites: Side-effects of increasing online interconnectivity. *Information, Communication & Society* 15, 9 (2012), 1279–1297.

[14] Petter Bae Brandtzæg, Marika Lüders, and Jan Håvard Skjetne. 2010. Too many Facebook "friends"? Content sharing and sociability versus the need for privacy in social network sites. *Intl. Journal of Human–Computer Interaction* 26, 11-12 (2010), 1006–1030.

[15] Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. 2011. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on psychological science* 6, 1 (2011), 3–5.

[16] Daniel Buschek, Moritz Bader, Emanuel von Zezschwitz, and Alexander De Luca. 2015. Automatic privacy classification of personal photos. In *Human-Computer Interaction*. Springer, 428–435.

[17] Kelly Caine. 2008. Linking studies of privacy in HCI to psychological theories of privacy. (2008).

[18] Kelly Erinn Caine. 2009. *Exploring everyday privacy behaviors and misclosures*. Ph.D. Dissertation. Georgia Institute of Technology.

[19] Krista Casler, Lydia Bickel, and Elizabeth Hackett. 2013. 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 (2013), 2156–2160.

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

[21] Edward J Clarke, Mar Preston, Jo Raksin, and Vern L Bengtson. 1999. Types of conflicts and tensions between older parents and adult children. *The Gerontologist* 39, 3 (1999), 261–270.

[22] Sunny Consolvo, Ian E Smith, Tara Matthews, Anthony LaMarca, Jason Tabert, and Pauline Powledge. 2005. 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.

[23] Eric C Cook and Stephanie D Teasley. 2011. Beyond promotion and protection: creators, audiences and common ground in user-generated media. In *Proceedings of the 2011 iConference*. ACM, 41–47.

[24] Patricia C de Souza and Cristiano Maciel. 2015. Legal Issues and User Experience in Ubiquitous Systems from a Privacy Perspective. In *International Conference on Human Aspects of Information Security, Privacy, and Trust*. Springer, 449–460.

[25] Marcie D Dorethy, Martin S Fiebert, and Christopher R Warren. 2014. Examining social networking site behaviors: Photo sharing and impression management on Facebook. *International Review of Social Sciences and Humanities* 6, 2 (2014), 111–116.

[26] Malin Eiband, Mohamed Khamis, Emanuel Von Zezschwitz, Heinrich Hussmann, and Florian Alt. 2017. Understanding shoulder surfing in the wild: Stories from users and observers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 4254–4265.

[27] Facebook. 2017. 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`.

[28] Simone Fischer-Hubner, Chris Hoofnagle, Ioannis Krontiris, Kai Rannenberg, and Michael Waidner. 2011. Online Privacy: Towards Informational Self-Determination on the Internet. (2011).

[29] Liqiang Geng, Larry Korba, Xin Wang, Yunli Wang, Hongyu Liu, and Yonghua You. 2008. Using data mining methods to predict personally identifiable information in emails. In *International Conference on Advanced Data Mining and Applications*. Springer, 272–281.

[30] Eric Gilbert and Karrie Karahalios. 2009. Predicting tie strength with social media. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 211–220.

[31] Google Street View. 2018. Image Acceptance and Privacy Policies. (2018). Retrieved May 7, 2018 from `https://www.google.com/streetview/privacy/`.

[32] Steven Greenhouse and Michael Barbaro. 2005. 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`.

[33] Rakibul Hasan, Eman Hassan, Yifang Li, Kelly Caine, David J Crandall, Roberto Hoyle, and Apu Kapadia. 2018. Viewer experience of obscuring scene elements in photos to enhance privacy. In *ACM CHI Conference on Human Factors in Computing Systems (CHI)*.

[34] Jianping He, Bin Liu, Deguang Kong, Xuan Bao, Na Wang, Hongxia Jin, and George Kesidis. 2016. Puppies: Transformation-supported personalized privacy preserving partial image sharing. In *2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. IEEE, 359–370.

[35] E Tory Higgins. 1987. Self-discrepancy: a theory relating self and affect. *Psychological review* 94, 3 (1987), 319.

[36] Roberto Hoyle, Robert Templeman, Denise Anthony, David Crandall, and Apu Kapadia. 2015. Sensitive lifelogs: A privacy analysis of photos from wearable cameras. In *Proceedings of the 33rd Annual ACM conference on human factors in computing systems*. ACM, 1645–1648.

[37] Hongxin Hu, Gail-Joon Ahn, and Jan Jorgensen. 2013. Multiparty access control for online social networks: model and mechanisms. *ieee transactions on knowledge and data engineering* 25, 7 (2013), 1614–1627.

[38] Panagiotis Ilia, Iasonas Polakis, Elias Athanasopoulos, Federico Maggi, and Sotiris Ioannidis. 2015. Face/off: Preventing privacy leakage from photos in social networks. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. ACM, 781–792.

[39] Aleksandra Kacperczyk. 2011. Social isolation in the workplace: A cross-national and longitudinal analysis. (2011).

[40] Sanjay Kairam, Mike Brzozowski, David Huffaker, and Ed Chi. 2012. Talking in circles: selective sharing in google+. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1065–1074.

[41] Sanjay Kairam, Joseph Kaye, John Alexis Guerra-Gomez, and David A Shamma. 2016. 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*. ACM.

[42] Bart P Knijnenburg. 2017. Privacy? I Can't Even! making a case for user-tailored privacy. *IEEE Security & Privacy* 15, 4 (2017), 62–67.

[43] Bart Piet Knijnenburg and Alfred Kobsa. 2014. Increasing sharing tendency without reducing satisfaction: finding the best privacy-settings user interface for social networks. (2014).

[44] William Kruskal and Frederick Mosteller. 1979. Representative sampling, II: Scientific literature, excluding statistics. *International Statistical Review/Revue Internationale de Statistique* (1979), 111–127.

[45] Priya Kumar and Sarita Schoenebeck. 2015. 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*. ACM, 1302–1312.

[46] Yao Li, Alfred Kobsa, Bart P Knijnenburg, and MH Carolyn Nguyen. 2017a. Cross-cultural privacy prediction. *Proceedings on Privacy Enhancing Technologies* 2017, 2 (2017), 113–132.

[47] Yifang Li, Nishant Vishwamitra, Hongxin Hu, Bart P Knijnenburg, and Kelly Caine. 2017b. 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*, Vol. 61. SAGE Publications Sage CA: Los Angeles, CA, 803–807.

[48] Yifang Li, Nishant Vishwamitra, Bart P Knijnenburg, Hongxin Hu, and Kelly Caine. 2017c. Blur vs. block: Investigating the effectiveness of privacy-enhancing obfuscation for images. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 1343–1351.

[49] Yifang Li, Nishant Vishwamitra, Bart P. Knijnenburg, Hongxin Hu, and Kelly Caine. 2017d. 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).

[50] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*. Springer, 740–755.

[51] Kimberly J Mitchell, David Finkelhor, and Janis Wolak. 2001. Risk factors for and impact of online sexual solicitation of youth. *Jama* 285, 23 (2001), 3011–3014.

[52] Adam D Moore. 2007. Toward informational privacy rights. *San Diego L. Rev.* 44 (2007), 809.

[53] Tyler Moore, Richard Clayton, and Ross Anderson. 2009. The economics of online crime. *Journal of Economic Perspectives* 23, 3 (2009).

[54] Glen J Nowak and Joseph Phelps. 1992. Understanding privacy concerns. An assessment of consumers' information-related knowledge and beliefs. *Journal of Direct Marketing* 6, 4 (1992), 28–39.

[55] Judith S Olson, Jonathan Grudin, and Eric Horvitz. 2005. A study of preferences for sharing and privacy. In *CHI'05 extended abstracts on Human factors in computing systems*. ACM, 1985–1988.

[56] Tribhuvanesh Orekondy, Bernt Schiele, and Mario Fritz. 2017. Towards a visual privacy advisor: Understanding and predicting privacy risks in images. In *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 3706–3715.

[57] Sinno Jialin Pan and Qiang Yang. 2009. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22, 10 (2009), 1345–1359.

[58] Eyal Peer, Joachim Vosgerau, and Alessandro Acquisti. 2014. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior research methods* 46, 4 (2014), 1023–1031.

[59] Tiffany A Pempek, Yevdokiya A Yermolayeva, and Sandra L Calvert. 2009. College students' social networking experiences on Facebook. *Journal of applied developmental psychology* 30, 3 (2009), 227–238.

[60] Andrew Perrin and Jingjing Jiang. 2018. 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/`.

[61] Pew Research Center. 2018. Internet/Broadband Fact Sheet. (2018). Retrieved February, 2019 from `http://www.pewinternet.org/fact-sheet/internet-broadband/`.

[62] Elissa M Redmiles, Sean Kross, Alisha Pradhan, and Michelle L Mazurek. 2017. *How well do my results generalize? Comparing security and privacy survey results from MTurk and web panels to the US*. Technical Report.

[63] Jessica Ringrose, Laura Harvey, Rosalind Gill, and Sonia Livingstone. 2013. Teen girls, sexual double standards and 'sexting': Gendered value in digital image exchange. *Feminist theory* 14, 3 (2013), 305–323.

[64] Joel Ross, Andrew Zaldivar, Lilly Irani, and Bill Tomlinson. 2009. Who are the turkers? worker demographics in amazon mechanical turk. *Department of Informatics, University of California, Irvine, USA, Tech. Rep* (2009).

[65] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander Berg, and Fei-Fei Li. 2015. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision* 115, 3 (2015).

[66] Robert Sheridan. 2014. 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`.

[67] Yoshinari Shirai, Yasue Kishino, Takayuki Suyama, and Shin Mizutani. 2019. 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*. ACM, 202–205.

[68] M Six Silberman, Bill Tomlinson, Rochelle LaPlante, Joel Ross, Lilly Irani, and Andrew Zaldivar. 2018. Responsible research with crowds: pay crowdworkers at least minimum wage. *Commun. ACM* 61, 3 (2018), 39–41.

[69] Meredith M Skeels and Jonathan Grudin. 2009. 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*. ACM, 95–104.

[70] Manya Sleeper, Rebecca Balebako, Sauvik Das, Amber Lynn McConahy, Jason Wiese, and Lorrie Faith Cranor. 2013. The post that wasn't: exploring self-censorship on facebook. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 793–802.

[71] Donna Spencer. 2009. *Card sorting: Designing usable categories*. Rosenfeld Media.

[72] Donna Spencer and Todd Warfel. 2004. Card sorting: a definitive guide. *Boxes and Arrows* 2 (2004).

[73] Qianru Sun, Liqian Ma, Seong Joon Oh, Luc Van Gool, Bernt Schiele, and Mario Fritz. 2017. Natural and Effective Obfuscation by Head Inpainting. *arXiv preprint arXiv:1711.09001* (2017).

[74] Kurt Thomas, Chris Grier, and David M Nicol. 2010. unfriendly: Multi-party privacy risks in social networks. In *International Symposium on Privacy Enhancing Technologies Symposium*. Springer, 236–252.

[75] Sara E Thomas. 2018. "What Should I Do?": Young Women's Reported Dilemmas with Nude Photographs. *Sexuality Research and Social Policy* 15, 2 (2018), 192–207.

[76] Lisa Torrey and Jude Shavlik. 2010. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI Global, 242–264.

[77] Lam Tran, Deguang Kong, Hongxia Jin, and Ji Liu. 2016. Privacy-CNH: A Framework to Detect Photo Privacy with Convolutional Neural Network using Hierarchical Features.. In *AAAI*. 1317–1323.

[78] Tom Tullis and Larry Wood. 2004. How many users are enough for a card-sorting study. In *Proceedings UPA*, Vol. 2004.

[79] Jessica Vitak and Jinyoung Kim. 2014. 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*. ACM, 461–474.

[80] Tim Wafa. 2009. How the lack of prescriptive technical granularity in HIPAA has compromised patient privacy. *N. Ill. UL Rev.* 30 (2009), 531.

[81] Wenbo Wang, Hean Tat Keh, and Lisa E Bolton. 2009. Lay theories of medicine and a healthy lifestyle. *Journal of Consumer Research* 37, 1 (2009), 80–97.

[82] Yang Wang, Gregory Norice, and Lorrie Faith Cranor. 2011. 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*. Springer, 146–153.

[83] Heng Xu, Na Wang, and Jens Grossklags. 2012. Privacy by redesign: Alleviating privacy concerns for third-party apps. (2012).

[84] George Yee and Larry Korba. 2005. Comparing and matching privacy policies using community consensus. In *Proceedings, 16th IRMA International Conference, San Diego, California*. Citeseer.

[85] Alyson Leigh Young and Anabel Quan-Haase. 2013. Privacy protection strategies on Facebook: The Internet privacy paradox revisited. *Information, Communication & Society* 16, 4 (2013), 479–500.

[86] Jun Yu, Zhenzhong Kuang, Zhou Yu, Dan Lin, and Jianping Fan. 2017. Privacy Setting Recommendation for Image Sharing. In *Machine Learning and Applications (ICMLA), 2017 16th IEEE International Conference on*. IEEE, 726–730.

[87] Jun Yu, Zhenzhong Kuang, Baopeng Zhang, Wei Zhang, Dan Lin, and Jianping Fan. 2018. 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 (2018), 1317–1332.