Over-Exposed? Privacy Patterns and Considerations in Online and Mobile Photo Sharing

Shane Ahern, Dean Eckles*, Nathan Good, Simon King, Mor Naaman, Rahul Nair
Yahoo! Research Berkeley
{sahern, ngood, simonk, rnair, mor}@yahoo-inc.com, *deaneckles@yahoo.com

ABSTRACT
As sharing personal media online becomes easier and widely spread, new privacy concerns emerge – especially when the persistent nature of the media and associated context reveals details about the physical and social context in which the media items were created. In a first-of-its-kind study, we use context-aware camerephone devices to examine privacy decisions in mobile and online photo sharing. Through data analysis on a corpus of privacy decisions and associated context data from a real-world system, we identify relationships between location of photo capture and photo privacy settings. Our data analysis leads to further questions which we investigate through a set of interviews with 15 users. The interviews reveal common themes in privacy considerations: *security, social disclosure, identity* and *convenience*. Finally, we highlight several implications and opportunities for design of media sharing applications, including using past privacy patterns to prevent oversights and errors.

Author Keywords
Privacy, online content, photo sharing, social software, location-aware, context-aware, photos.

ACM Classification Keywords
H.1.2 User/Machine Systems: Human factors.

INTRODUCTION
The growing amount of online personal content exposes users to a new set of privacy concerns [1,2,20,21]. Digital cameras, and lately, a new class of cameraphone applications that can upload photos or video content directly to the web, make publishing of personal content increasingly easy. Privacy concerns are especially acute in the case of these multimedia collections, as they could reveal much of the user's personal and social environment. The persistent nature of such online media could expose rich aggregate information about the owner, and subjects, of the content. The considerations made by users during the content sharing process are crucial for the design of systems that support the creation of such content.

CHI 2007, April 28–May 3, 2007, San Jose, California, USA.

In this work, we examine how users of Flickr [8], a popular photo-sharing web site, manage their privacy policies for photographic content. The users we studied upload photos to the Flickr web site using ZoneTag, a mobile application running on high-resolution, location-aware cameraphones. Concentrating on these users and the existence of contextual data that is associated with their actions puts us in a unique position to explore critical aspects of privacy, including:

- Users' considerations in making privacy decisions about online content.
- The content- and context-based patterns of privacy decisions in an online photo sharing environment.
- Ways in which different people make privacy policy decisions "in the moment", and their strategy of dealing with such decisions in mobile settings.
- User behavior regarding location disclosure [7] and systems that maintain, and sometimes expose, long-term and persistent information about their location.

Our study consists of both qualitative and quantitative analysis. In the quantitative analysis, we offer a study of a real-world, large-scale system and its regular usage, analyzing previously unavailable usage data such as capture location. The findings of the data analysis inform a series of interviews with ZoneTag users to extract qualitative information about privacy decisions and considerations.

We discuss a taxonomy of privacy considerations that was surfaced by our study. These considerations can be classified according to four main themes: *security, social disclosure, identity,* and *convenience*. For each of these themes, users may consider implications for themselves or for others. We expand on this taxonomy and demonstrate how different users' privacy considerations fall within it. In addition, we show initial evidence that many users have content- and context-derived patterns in making privacy decisions. For example, patterns of "location-based privacy" emerged, showing that, as one user phrased it, *"some locations are more private than others"*.

While our study focused on a specific online sharing system (Flickr) and a specific device and capture software (cameraphones and ZoneTag), we believe that implications of this study are broad. As an online community, Flickr has

many different foci (artistic expression being the most high-profile, but not the most common). However, we focus on the use of Flickr as an online tool for personal sharing and archival. As ZoneTag runs on phones with high quality cameras, its usage is somewhat different than cameraphone usage as described in [19]: as the images are often of archival quality, usage is often similar to that of regular digital cameras. The implications, therefore, extend beyond the specifics of Flickr and ZoneTag to other systems of web-based video, photo, and content sharing as well as to different systems of mobile capture, for example, digital cameras and other mobile devices.

RELATED WORK

Studying users' privacy concerns is notoriously difficult, and accurate measures of user behavior are in some cases unattainable [1,15,23]. Even the meaning of the term "privacy" varies considerably between people and contexts [27], and people's stated preferences often don't match their actions [23]. Performing privacy studies in the world of mobile computing is even more difficult, as getting information about users' concerns at the moment they occur can be cumbersome and unreliable. Some research efforts rely on diary studies [7,13,16], surveys [14,23] and interviews [19,26]. Recently, Iachello et al [15] have looked into a novel technique called "paratypes" as a method for eliciting user feedback. Paratypes employs specific privacy-based scenarios, similar to *critical incident* techniques in workplace psychology. We take an alternative approach in this work, examining privacy decisions as they were applied in practice, in a real world mobile (and online) application. From observed data and user reports we try to understand user motivations and extract privacy patterns.

In HCI research, especially in the field of ubiquitous computing, feedback, control, and transparency have emerged as primary methods of dealing with privacy issues [5,6,20,21,22]. The privacy issues with mobile and networked devices have been explored for networked desktops [6,10], wireless devices [11,17], mobile phones [3] as well as sensor networks [12].

Several approaches were developed to help users mitigate privacy concerns when disclosing information. Of these, privacy of location information is of particular interest to our work. Varying the degree of "vagueness" of location information is one approach described in the work of [7,15,20]. Consolvo et al [7] describe a formative study, where they examine disclosure of location information to social cohorts. In their study, the researchers contacted mobile users with hypothetical periodic queries for their location throughout the day. Findings indicated that the identity of the hypothetical requester was a main factor in deciding about disclosure of location information; when granted, the disclosure was given with full granularity. In other studies [15,17,18,25] vague location information was shown not to alleviate privacy concerns. For example, knowledge that a person is one state as opposed to another for a business trip could be as damaging as revealing as their home address. In contrast to some of these studies, our observations in this work are grounded in a real deployed and active system.

Systems that persist personal context and content information, such as MyLifeBits [9] raise privacy considerations similar to those of our system. The MyLifeBits project uses cameras and audio recording devices to continuously record and categorize every moment of a person's life. Zonetag is similar to MyLifeBits in that it enables users to capture a context-aware record of their daily lives, and it uploads, archives and categorizes this information. Unlike MyLifeBits, ZoneTag and Flickr are also designed to enable sharing of photos and the associated metadata with friends, family, and the general public, making the privacy considerations in the system more complex.

Flickr is perhaps similar to existing "social-network" sites, enabling its users to share, organize and comment on their mutual photo collections. Consequently, many of the privacy and identity issues that arise in social network sites such as Facebook [2,24] and MySpace [4,24], exist in ZoneTag and Flickr as well. Privacy and disclosure factors in those systems have not yet been studied in depth. In addition, by extending a user's social network into the mobile space with real time image and context capture as well as context capture, ZoneTag and Flickr raise additional concerns that are not reflected in other social network sites.

SYSTEM DESCRIPTION

This section briefly describes the key privacy-related features on Flickr and ZoneTag. In particular, we discuss how privacy can be controlled by the user at capture time (using ZoneTag) and later, on the Flickr interface. We also describe how user content is made available and findable on Flickr.

Flickr

Flickr is a popular online photo organization and sharing service with over five million users who have uploaded more than 250 million images. Flickr gives users control over how their photos are shared with others, primarily by allowing users to select which groups (or classes) of people can view and find each photo. Flickr has five privacy levels: private, family-only, friends-only, friends-and-family, and public[1]. A user can change the privacy settings for any of their photos at any time via Flickr's web

[1] The privacy settings can be grouped into two basic classes: *public* (any visitor to the Flickr website can find and view the photo) and *non-public* (visible only to the photo owner, or extended to users the owner designates as 'friends' or 'family'); for the purpose of the data analysis, we often do not make the distinction between different types of non-public photos.

interface. Similarly, the user can designate other Flickr users as friends or family at any time. A 'friend', for example, can access all the photos from the contributing user that are marked as available for friends, regardless of the time the photo was uploaded or the time the viewing user was designated as a friend.

Photos on Flickr are found or discovered in a variety of ways, for example:

- Public photos appear in search results for terms matching text from the photo's title, caption or tags (textual labels).
- For each Flickr user there is an easy-to-find page displaying their photos, sorted by recency.
- A Flickr user's "contacts" page shows recent photos uploaded by all of their contacts, family and friends that the user has permission to view.

ZoneTag

ZoneTag is a mobile phone application, available as a public prototype, that supports cameraphone photo sharing and organization via Flickr. ZoneTag is available for high-end "smartphones" from Nokia and Motorola and is designed to reduce barriers to photo upload and annotation.

ZoneTag is designed for ease of content publishing. After the user captures a photo, the ZoneTag application prompts them to upload the newly captured image to Flickr. If they choose to upload the photo, users can upload photos with the previous photo's settings, requiring minimal interaction on the mobile device. Alternatively, users can change any of the photo's settings before upload. The settings available include selecting one of the five privacy options for the photo, as described above.

In addition to applying privacy settings, ZoneTag allows users to select tags that will be associated with the photo on Flickr. ZoneTag employs a number of techniques, such as tag suggestions and quick text entry, to encourage users to add tags to each photo. Tags often suggest the content of the image; we use this fact in our data analysis section.

ZoneTag uses cell-tower information to expose the capture location for each photo via the Flickr interface. The system converts the phone's cell-tower information to human-readable location labels (i.e. city, state, country, zip/postal code) that are added as tags to the photo's page on Flickr together with the set of user-provided tags. This feature of ZoneTag exposes the location where a photo was taken to any user that has permission to see the photo on Flickr.

Location data is particularly interesting for a number of reasons. First, location is highly indicative of life patterns and significant contexts of the users' daily lives. Second, location data is increasingly available in various consumer devices. The usefulness of location in many applications (such as photo organization) will make more location-annotated consumer content available online.

In summary, ZoneTag combines features that make daily life recording and sharing through digital photos possible even for non-technical people. ZoneTag brings together elements from both digital cameras and traditional cameraphones — ready-to-hand capture and a high quality camera.

We now turn our focus to an analysis of the ZoneTag usage logs; findings from this data analysis lead to questions we explore in interviews with individual ZoneTag users.

DATA ANALYSIS

At the time of writing, ZoneTag had been deployed as a publicly available prototype for over 5 months. Most of the users of ZoneTag are self-selected, early adopters of technology. In total, over the months, ZoneTag was used by more than 350 people who uploaded a total of over 44,000 photos to Flickr. We will focus our data analysis on 81 users who have uploaded at least 40 photos, accounting for 36,915 photos – an average of 455 photos per user (stddev=878.8). As expected, the number of photos per user follows a power law distribution. We chose to focus on users with at least 40 photos so that we could examine variation within a user's behavior over time. Furthermore, users with fewer photos have not used the system enough to establish recognizable behavior patterns.

During deployment, we collected detailed data regarding the usage of the system. This data includes automatically-captured metadata (time and cell ID-based location), the settings (privacy and tags) applied to images using ZoneTag on the phone before upload, and subsequent changes made to these settings via Flickr's web interface.

The data analysis attempts to answer the following research questions:

RQ1) Is location (as approximated by cell ID) a reasonable predictor of privacy settings?

RQ2) Is the content of photos (as approximated by tags) a good predictor of privacy settings?

RQ3) Do users revisit the privacy choices they made while mobile, and how frequently?

RQ4) Are users generally willing to expose the location of their photos?

It is important to note that our analysis is limited by the extent of our data capture. From the data, we cannot tell how often users chose not to upload a photo, or modified their photo-taking behavior to protect their privacy or the privacy of others. Also, for simplicity, we do not distinguish between the various non-public privacy settings. We address some of these deficiencies in the user interviews.

Does Location Predict Privacy Decisions?

To examine RQ1 we tested two hypotheses:

H1) There are some locations where each user is more likely to make photos public, compared to their overall behavior across all photos. Similarly, in some locations, a user is more likely to set photos as non-public.

H2) Users' privacy settings are likely to differ between locations they photograph frequently (e.g. home, work) and locations they photograph infrequently. We expect more frequently photographed locations to be more private.

To test H1, we examined privacy decisions made by users in each location (as determined by cell ID) where they took photos. While using cell tower-based location is "fuzzy" to some extent, we found that patterns still emerge from the data. Presumably, access to more fine-grained location information would allow even better predictions (though any analysis would still likely be based on grouping locations into somewhat arbitrary clusters.)

For each user, we grouped locations into three categories by comparing the ratio of public photos to total photos for each location, to the same overall ratio for all photos (across all locations) from that user. If the location-specific ratio was within 0.1 either side of the overall ratio the location was classified as *typical*. For example, if a user's overall public ratio is 0.5, and in a certain location they have public photo ratio of 0.42, this location is classified as *typical*. When the ratio was less than the overall public photo ratio for that user by between 0.1 and 0.25, the location was classified as *private*. When the ratio differed by more than 0.25, the location was classified as *very private*. Equivalently, location-specific public ratios greater than the overall user ratio led to locations classified as *public* or *very public*. Of the 81 users we examined 5 had only private photos, and 14 user had only public photos. For the remaining 62 "mixed-privacy" users, location-privacy sensitivity varied, with about half (30) showing privacy settings to be quite sensitive to location (fewer than half their photos were taken in *typical* locations.) 19 of these 30 users' privacy settings were highly sensitive to location, with at least half their photos taken in *very private* or *very public* locations.

To examine hypothesis H2, that photos from frequently-photographed locations are more likely to be private, we looked at privacy decision as a function of how many photos were taken in a location by a user. In Figure 1, we grouped such user-location pairs by the number of photos per pair. For example, there were 2697 locations where some user took a single photo. In another example, there were 29 instances of locations in which some user took 20 photos (accounting for a total of 580 photos). Next, we computed the ratio of public photos to total photos for each group, shown in Figure 1. Figure 1 shows photos per user per location (grouped into buckets of size 5, e.g., all instances of locations where one user took between 1-5 photos are group together). The Y-axis represents the ratio of public photos in each group (data beyond 210 photos per user per location is removed for clarity – only several such locations occurred). For example, examine the left-most point in the graph – this point represents all instances where a single user took between one to five photos in some cell. Roughly 60% of these photos are public. In particular, we found a significant negative correlation between the ratio of public photos for the group and the number of photos per user-location pair ($r(118) = -.213$, $p < .05$). That is, users are indeed less likely to make photos public in locations they frequently photograph, more likely to take public photos in locations they photograph infrequently.

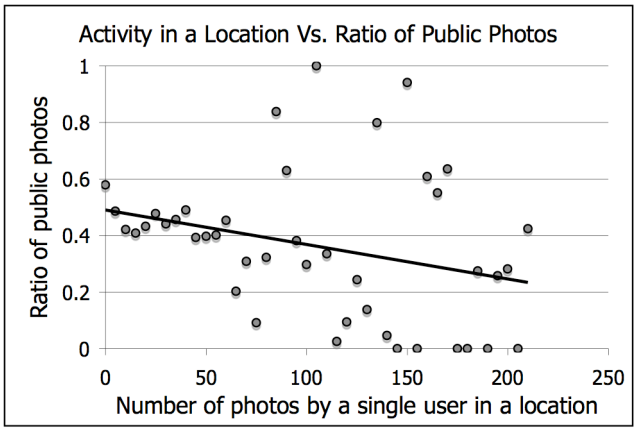

Figure 1. Do users tend to make photos private in frequently-photographed locations?

In summary, it appears that location (even as approximated by cell ID) could be used a predictor of likely privacy settings. Specifically, in response to H1: a significant portion of users have some set of locations in which they are more likely to take private photos, and some in which they are more likely to take public photos. As for H2, it seems that users are indeed more likely to make photos private in frequently photographed locations.

Does Content Predict Privacy Decisions?

To examine the relationship between content and privacy, we utilized user-supplied tags that were associated with many photographs, as a rough descriptor of the photo's content. We hand-classified the tags into six categories, selected subjectively by identifying major themes in the set of all tags: *Person, Location, Place, Object, Event, and Activity*. Then we associated photos with a each category, according to the tags attached to the photo, and observed privacy differences between photos in the different categories. Note that since each photo may have multiple tags from different categories associated with it, a photo may be counted in multiple categories.

To simplify the task of hand-classification, we only classified frequently recurring tags: the top-fifth most frequently used tags for each of the 81 users, resulting in 1538 distinct tags. The tags were classified according to their text, without examining any images; for example, the tags `Mom' or `Marc' were both categorized as *Person*. Three members of our team classified about 500 tags each, with the option to flag a tag as "difficult to categorize". The "difficult" tags were discussed as a group; if a consensus could not be reached, the tag was left as uncategorized, leaving 1295 categorized tags. The ratio of public photos to non-public photos for each tag category can be seen in Figure 2. For example, of photos that had *Person* tags, 72% were marked as private. For each category, the number of

corresponding public and private photos is also shown in the figure (for instance, there were 3063 public photos with a *Person* tag).

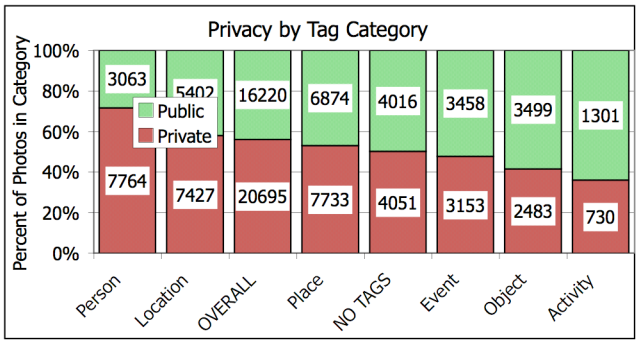

Figure 2. Ratio of public to private photos by tag category.

Further analysis reveals that the differences visually apparent in Figure 2 between public photo ratio for *Person* and all other categories except *Location* is significant ($p<.05$); we also found moderately significant differences ($p<.1$) between *Person* and the overall ratio (*Person* photos are more private). Significance values were calculated by dividing the difference between two samples by its standard error to obtain the z-scores. As reported below, in the qualitative study we have explored, re-affirmed and partially explain this *Person* privacy finding. In addition to the significance of *Person* privacy, we found moderately significant differences between *Activity* (more public) and the overall ratio.

Are Capture-time Privacy Decisions 'Good'?

We looked at how often users changed privacy settings on Flickr, as an indication that they may have regretted their original mobile privacy decision. The privacy settings of approximately 7% of photos were changed after upload; either from public to non-public (2.4%) or non-public to public (4.6%). These numbers indicate that users' capture-time privacy decisions were not, in general, regretted strongly enough to result in a change to the privacy settings. The changes from public to non-public, though relatively few in number, do represent a serious potential problem for users, perhaps indicating that users inadvertently temporarily exposed private or sensitive photos. The fact that the ZoneTag interface defaults to the last photo's privacy settings (unless explicitly changed) could be the cause of this issue (in 95% of the cases where a photo was uploaded as public and later made non-public, the previous photo uploaded by the user had also been public). Another possible cause for change of privacy settings is that users noticed, when viewing on a larger screen, that a photo was less interesting or of lesser quality (or, conversely, better) than they believed when they uploaded, leading to the privacy switch. We examine some of these issues in further detail in the user interviews.

Are Users Concerned About Exposing Their Location?

The ZoneTag client application allows users to suppress the automatic location tags associated with a photo. Only 2% of the photos (767) have location information suppressed, possibly indicating that users were not, overall, concerned about exposing zip/postal-code-level location with their photos. Only 18 of the 81 users ever suppressed location tags, with only three of these users suppressing location on more than 10% of their photos. Note that the default interface option is for the location to be included; therefore, this short examination is not conclusive and served as an interesting point to be re-examined in the interviews.

INTERVIEWS

Our qualitative study, consisting of a series of user interviews, was designed to explore themes that cannot be extracted from the data: considerations and user attitudes regarding privacy and privacy decisions in the system. The interviews were also extended to explore themes that were exposed in the preliminary data analysis, such as privacy of location and content, and its underlying human factors. In particular, our qualitative study's goals included answering the following questions:

RQ5) What are the major issues that users consider when making decisions about privacy of photos?

RQ6) What is the effect of photo content on privacy decisions?

RQ7) How do users make "in the moment" decisions about photo privacy, and do they ever consider revisiting these decisions?

RQ8) What are the human factors leading to the observed correlation between photo location and the privacy settings applied by the user?

RQ9) Are users generally willing to expose the location of their photos, or are they concerned about such disclosure?

To this end, we recruited, equipped and interviewed 15 participants during four weeks in summer 2006. Out of our 15 users, 11 had no technical background. The participants ranged in age from early twenties to early fifties: three participants fell in the 20-25 age group; eight were between the ages of 30-35, three were 35-40, and one user was just over 50 years old. Eight participants were male, and seven female. The users were not recruited through Flickr; most of them were not Flickr users, and signed up for Flickr and ZoneTag only when they were recruited for our experiment. Recruiting external participants allowed us to investigate patterns that are relevant to a broader population, avoiding the self-selection bias of recruiting active Flickr users.

Our interview participants were provided with a Nokia cameraphone (a 2-megapixel Nokia N70 or a 1.3-megapixel Nokia 6682). The ZoneTag software was installed on each device. To ensure that participants carry and use the phone, we allowed them to use it as their personal phone by installing their own SIM card. We paid for each subject to purchase a unlimited data-access plan from their cellular carrier, so that they were able to upload a virtually unlimited number of photos to Flickr using ZoneTag. We also supplied the users with Flickr Pro accounts for the

duration of the study, again to prevent any limitation in data capture and access.

Recruiting Participants

The participants were recruited through fliers, mailing lists, and social connections to existing ZoneTag users. In recruiting participants, we explicitly searched for social groups of non-technical people who would be interested in sharing photos within their group. A second desirable quality of the user base was broad coverage of personal and social motivations and goals. For this reason, our recruiting effort focused on different types of groups: groups of parents with young children, groups of coworkers, and groups of young adults in their early twenties.

We aimed for this variation in background and social interaction to make sure we covered various scenarios of sharing, privacy and security scenarios. With families, we recruited three pairs of spouses (overall, nine of our interview participants had young children). Many of these spouses often took pictures of their children to share with family and friends. With co-workers, our group consisted of three coworkers from the same department, working on the same floor. Other people in their department, and many others in the company, were users of Flickr. In addition to families and coworkers, we recruited a group of young friends. The group consisted of a triad of two males and one female, all in their early twenties. There were a number of connections between users in the different groups. To summarize, we recruited users in a way designed to elicit varied scenarios of privacy and sharing behavior.

Post-Usage Interviews

The main portion of our qualitative study consisted of post-usage interviews. We conducted the interviews after each subject had used ZoneTag for two or more weeks. The structured interviews included several methods of reviewing the participant's activity during the study, in an attempt to draw out additional information about their privacy decision-making process, attitudes and actions.

In particular, the interviews consisted of the following portions. First, the interviewer informally queried about the participant's usage of ZoneTag during the experiment, including an overview of the occasions and situations in which the participant used the cameraphone.

In the second part of the interview, we performed a more grounded investigation of the participant's activity. We used the photo elicitation method from Van House et al [27]. In photo elicitation, the interviewer discusses with the participants specific photos taken during the study. This method allowed us to capture more accurately the considerations that came into play for each user, as the discussion was pointed and geared towards a specific occurrence. To explore and select photos, we used a visualization tool that displayed a timeline view of the user's photos taken during the study, split according to the photos' privacy settings. We then asked the participants to use the tool to select five photos from their collection. We discussed each photo in turn, focusing on the privacy decisions and considerations regarding that photo, and how those were related to the social, temporal, and location context of the photo. We queried specifically about how comfortable users feel about the level of location disclosure for the photo. We also discussed the motivation for taking the photo, and whether it was taken in order to be shared with other users. Of course, we discussed how the specific privacy setting was selected for that photo, and why.

In the last part of the interview, we explored the privacy concerns of the participant in a directed discussion that allowed participants to address issues that were not encountered in the more-structured portion. We asked about other conditions that influence the participants' privacy decisions; about specific cases, not covered earlier, when the participant have been concerned about people accessing their photos on Flickr; and about their level of comfort with their current level of exposure on Flickr. Each interview was recorded and transcribed. Interviews lasted between 45 and 75 minutes.

DISCUSSION

Our interviews surfaced several concerns and considerations for users who upload and share photos online. In the process of reviewing of the set of interviews and transcriptions, we evaluated the set of concerns and considerations mentioned by each of the participants. Based on the review, we constructed a taxonomy of the main topics that repeated in the interviews. The taxonomy classifies the considerations in two dimensions. One dimension captures the *theme* of the consideration: security, social identity, social disclosure and convenience. The second dimension describes the *object* of the consideration: self or other. The taxonomy is described in Figure 3.

Object	Security	Identity	Social Disclosure	Convenience
Self	Exposing self to security hazards	Managing own on-line identity	Exposing socially sensitive information to contacts	Difficulty of sharing and viewing
other	Exposing other to security hazarads	Influencing other's on-line identity	Exposing other's socially-sensitive information	Other's ability to share and view

Figure 3. Taxonomy of privacy considerations.

It is important to note that not all of the different considerations that are represented in our taxonomy were made or expressed by every participant. However, most participants considered *several* of these when making photo-related privacy decisions. Unfortunately, competing considerations often generated conflicts and forced users to make compromises in one or more areas.

The *object* dimension captures the object of consideration: *self* means that the user is thinking about how a certain privacy theme will affect them. The *other* category captures considerations regarding the privacy, identity, social disclosure and convenience of other individuals that may

appear in the photo or are related to the photo in any way. Often, the participants held these photos and considerations to a higher standard than photos whose object of consideration was the user herself. One participant noted that *"I'm a little more cautious if it is other people and I don't know how they would feel about being online"*. We give more examples of considerations in both *self* and *other* categories from our user interviews below, in our discussion of the *theme* dimension of the taxonomy. It is worth mentioning that the *object* dimension can be thought of as a continuum, ranging from personal concerns regarding one's own privacy, to concerns that pertain to close contacts such as friends or family that appear in the photos, to concerns regarding complete strangers.

Next, we describe the different categories in the *theme* dimension of the privacy considerations taxonomy. For each theme, we provide examples and supply quotes from our user interviews.

Security. Personal security (including physical security, property security and so forth) was, as expected, one of the themes that participants raised in the interviews. People with family and children seemed especially concerned about publishing photos and photo privacy. A participant commented on a photo that was marked public that it *"slipped by me... I need to change that... Half naked pictures of my daughter open to the public right now"*. Another user was not comfortable with her kid's photos online - referring to this idea as a *"roadmap for sexual predators"*. A third mother said, *"Pictures of the kids... I don't know if I'd mark public... I don't know who is out there."* The availability of location information with these photos was also a factor for some users, exposing (potentially in real-time) location information about their whereabouts, or their children's whereabouts. This theme of privacy concerns was not limited to parents: a younger, male participant noted, *"If I did something to upset somebody somehow... and they knew exactly where I lived by looking at my Flickr photos, that would bother me"*. The granularity of the location was of course a factor in this consideration. One of our participants expressed a worry that burglars could see the contents and exact location of their apartment. However, we also found that not all participants had specific concerns about security. As one participant said, *"There is so much stuff out there, I don't think I'm really a prime target."*

Participants also considered the security of other people that appeared in their photos. For instances, one participant noted about a photo, *"It was somebody else's kid so I made it private."*

These types of expressed concerns about security seem to confirm the data findings regarding the tag- and location-based privacy decisions patterns that were found in the data analysis. Recurrent themes were security of children and locations like 'home'; the data analysis had suggested that person tags are especially sensitive, and that many users have some locations that are "more private" than others.

Identity. A major part of the interaction on Flickr, as on other social network and content-sharing web sites, involves crafting and presenting one's identity. Photos that a user shares can reflect on their identity formation in two ways. First, the content of the image may be displaying the user or their environment in an unflattering way: *"You've heard about all these random stories about people on MySpace/YouTube and a HR department will search for people and find stuff."* A second way in which the identity of the user can be damaged is when the photo exposes some of the user's interests that they may try to keep private. A younger user was worried about what his friends would think: *"I might [upload] as family-only if it was a family event that I didn't want my friends knowing I was at, embarrassing pictures of me and my aunt"*. Another user made sure her conservative family was not able to see her photos taken at a gay pride parade.

Again, some participants were aware and concerned about other people's identity considerations; as one said, *"I really wouldn't make any of my pictures of my friends or people I know public to the whole Internet... that way I don't have to worry if someone doesn't want their photo online"*.

Social Disclosure. Perhaps more immediate and explicit than identity concerns, users raised issues of disclosure of their activity and whereabouts to known people. Such considerations naturally arise in systems with immediate upload, an always-with device, and policy-based sharing (as opposed to sharing with particular individuals). Some participants found this disclosure to be a positive factor on occasion (*"I added the tag '[bar name]' so he'd know where I was..."*). Others were more worried; in one example, a user stated, *"if I went somewhere and I didn't invite someone and I didn't want them to see, I might want to make that 'only family'"*.

Users make similar social disclosure considerations regarding other people. In one instance, a user noted that *"When I'm hanging out with my musician friends, when they're doing their, uh, 'musician things', I might not want to take any pictures of that... don't need any incriminating evidence of anyone"*. This self-censorship is similar to that discussed by Bellotti et al. [5], in which the most effective means of privacy control was found to be covering a camera lens when potentially embarrassing behavior is taking place in view of a video camera. We saw similar behavior with some of our participants, opting for not taking a photo at all under some circumstances.

Convenience. The consideration of how easy it would be for other people to find, view and discover the images online was a recurring theme. Limiting photos for friends- or family-only viewing means that people who are interested in viewing these images have to sign up to the service and be marked as friends by the user. This type of inclusion was not always possible; as a result, participants

often found themselves in a position that forced them to make photos public if they wanted certain individuals to see them. One participant said *"I made [a set] of my photos public so I wouldn't have to worry when I gave a friend the URL [to the set of photos] so they could browse my pictures freely"*. In the *Other* dimension, the 'Convenience' consideration for other people that appear in the user's photos means making the photos visible for the benefits of others; for example, making sure a friend of the photographed person can view the image. Again, this consideration for others was often given more weight than the convenience of the users themselves.

Before we summarize the findings of the interview, we go beyond the taxonomy to look at our participants' attitudes and practices regarding location disclosure.

Location Privacy

As mentioned above, the interviews included specific probes about the attitudes and level of concern of our participants had towards exposure of location data. We asked the participants about their aggregate data and specific photos: were the users concerned about the fact that the system collects and often exposes location data?

We showed participants aggregate location data which revealed information about the times and locations of their photos. This information was derived from their ZoneTag/Flickr photos, and in many cases (for participants with many public photos) was publicly available on the web in aggregate form. On the whole, though, participants expressed little or no concern. One user's comment was typical: *"I don't care [if my photos show my location], I'm not trying to hide my location from anyone. And if I was, I wouldn't be taking pictures and putting them online."* For most users, the fact that their location history is exposed did not seem to add additional concerns. Two users expressed specific concerns about advertisers that might start using their disclosed location information, combined with the content of their photos. One participant commented *"If marketers from Flickr or another site looked at people's photos and say 'Oh hey they like cycling, let's look at where they live and then send them advertisements', I wouldn't like that very much. If that's the case, I would really be selective about what I make public."*

The granularity at which the users expect the location data associated with the photos to be exposed may have an effect on the privacy and location disclosure decisions. We found that different users are comfortable with different levels of location granularity. One user commented that *"city information is okay, but I'm not sure about zip codes"*, while most other users had no problem with exposing the zip code even for their home, but were averse to exposing the exact address.

It should be noted that for many participants, applied privacy settings and attitudes differed considerably from their stated position in a pre-study interview and survey. When asked about exposing the zip code of the photo to Friends, 17% of the users stated they would never share the zip code, while 50% said they would never share the zip code except for special circumstances. In reality, all users shared zip code level information, and made no effort to configure the location settings to conceal this information. In the post interview, when asked about zip code level information, most of the users were comfortable with that level of disclosure for friends and family.

Summary of Qualitative Findings

The qualitative examination of considerations and behaviors surfaced four major themes in privacy considerations about the self and others (*security*, *identity*, *social disclosure*, and *convenience*). This preliminary taxonomy emphasizes the complexity and potential for conflict in the factors behind privacy choices and offers a vocabulary for thinking and communicating about this difficult landscape.

Figure 4 shows the taxonomy originally presented in Figure 3, now with the number of interview participants that expressed a specific concern for each type of consideration. The total number of subjects expressing a concern appears together with (in brackets) the number of parents that expressed the concern (recall that we interviewed nine parents and six non-parents.) For example, in 13 out of 15 interviews, the participant raised the *identity* consideration regarding themselves. The same type of consideration was raised by 8 out of 9 parents.

Object	Theme			
	Security	Identity	Social Disclosure	Convenience
self	3 [2]	13 [8]	8 [3]	6 [2]
other	8 [8]	14 [9]	12 [8]	7 [3]

Figure 4. Breakdown of user concerns.

While these numbers are not necessarily indicative of overall trends, they provide an initial look into the breakdown of concerns expressed in our study. We observe that *security* of others (their children, presumably) is an overwhelming concern for parents, while the *security* theme is only mentioned by a single non-parent. Overall, identity was a consideration for virtually all interviewees, with concern for exposing photos of others voiced even more often than concern with managing one's own identity.

Several additional factors that influence the privacy decisions, but are not covered in our taxonomy, appear below. These factors pertain to deficiencies and limitations of the decision-making process, rather than specific themes that directly influence the privacy decisions. For this reason, we decided to list these factors separately. In particular, we refer to the implication of making privacy decisions at capture time—with limits on attention, long term perspective, and knowledge.

Choice under uncertainty — Users are uncertain about the content of, audience for, and norms regarding particular disclosures. This uncertainty limits users' ability to make the best decision at capture time. For example, at the time of capture and potential upload, users do not know how the photo will look to those viewing it on the web, or they may be uncertain about the preferences of the photo's subject or other people related to the photo's content.

Dealing with complexity — Making the best available disclosure choice is often difficult and demanding of attention and time, sometimes prohibiting careful decisions for each photo in the moment. Users may regret a decision shortly after making it or just mistakenly over- or under-disclose information. Users sometimes adopt strategies for reducing the complexity of the decision (e.g. choosing the same setting for all photos uploaded).

Compromises and dissatisfaction — Disclosure decision-making can involve significant compromises, as multiple factors and preferences provide reasons for conflicting decisions. Unsatisfactory decisions are much more frequent than regretted decisions; that is, users often do not prefer other available options but are unhappy with the chosen option because some reasons speak against it.

IMPLICATIONS

The above findings have actionable implications that identify both choices for system designers, and topics for future research. There are opportunities to support and influence users' privacy decision-making process by changing available settings and providing information, simulations, and recommendations. Specifically, we identify five directions suggested by our work.

Preventing mistakes or reducing their impact — Through the study of patterns in disclosure behavior, systems may be able to helpfully warn users when disclosing following potentially significant change in context, perhaps reducing potential for mistakes. As we found that privacy decisions are often correlated with the context of capture and the content of the photo as indicated by user-specified tags, it could be feasible to use these patterns for prediction or recommendation of privacy settings. In addition, providing an optional "staging area" before disclosure actually takes place and an easy way to review recent disclosures may reduce the immediate consequences of quickly regretted or accidental disclosure decisions.

Increasing awareness of information aggregation — Users may not realize what aggregations of disclosed information, such as photo locations over time, can reveal to the system, or to people examining the user's online activity. One way to increase awareness and change behavior is by presenting personalized sample output of such aggregations.

Impactful information and feedback – Systems can reduce user uncertainty about factors important to disclosure choices. For example, systems may be able to estimate the audience for a particular disclosure at decision-time, thereby reducing uncertainty and influencing user choices. Systems could use social comparison, such as decisions made by friends or other users in similar context, to reduce uncertainty about relevant norms for disclosure. Finally, tools for viewing photo "disclosures" in ways similar to how others will view these photos could help users understand the content and appearance of their disclosures.

Discoverability and the convenience of disclosure — Work on further decoupling the visibility and discoverability of media could maintain the convenience of public settings while decreasing potential for unexpected viewing.

Decoupling photo and location information visibility — Concerns about disclosing location information conflict with the appeal of location information for photo organization. Allowing users to control the disclosure of location independently of the privacy of the photo could resolve this conflict. Flickr now supports such decoupling, where users can set a policy regarding whom can see the location data associated with their photos. An extension of this feature will allow users to set varying granularity in which viewers can see their location data.

Discouraging blanket strategies — Use of a single privacy setting for all photos is often undesirable for system designers. For example, a user who makes all photos public may be self-censoring (and not using use the system for personal organization and private sharing) or over-disclosing (and making regret likely). On the other hand, public photos can create value for system owners, so potentially-public photos marked non-public may lose value. Future work should explore strategies to encourage use of a range of privacy settings.

CONCLUSIONS

Issues of online privacy have long been of concern in the HCI community, and are of growing concern for the general public as an increasing amount of personal content is becoming available online.

We have conducted a qualitative and quantitative analysis of privacy in a real-world photo-sharing mobile and online application. Using context-aware cameraphones as capture devices allowed us to explore patterns of privacy in a way that was previously unavailable. Our findings, and design implications, are relevant to researchers and designers of content-sharing systems as well as mobile capture devices.

Are users over-exposed? For now, it is a matter of taste; but while the potential for disaster exists, some users remain unconcerned. We are hoping to keep investigating the topic to get a more detailed look at patterns across a longer time period, and perhaps in different cultures.

ACKNOWLEDGEMENTS

We would like to thank our participants who have graciously contributed their time and input to this work. We would like to thank Nancy van House and Vlad Kaplun for allowing us to use the Flickr-based Photo Elicitation tool.

Finally, we thank the anonymous reviewers for their thoughtful and extensive comments.

REFERENCES

1. Acquisti, A. and Grossklags, J. Privacy and Rationality: Preliminary Evidence from Pilot Data. In Proc. WEIS 2004.
2. Acquisti, A. and Gross, R.. Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook. Privacy Enhancing Technologies (2006).
3. Barkhuus, L., Dey, A.K. Location-based services for mobile telephony: a study of users' privacy concerns. In Proc. Interact 2003, (2003) 709-712.
4. Barnes, S. A privacy paradox: Social networking in the United States. First Monday 10, 9 (2006).
5. Bellotti, V. and Sellen, A. Design for Privacy in Ubiquitous Computing Environments. In Proc. ECSCW 1993, Kluwer (1993), 77-92.
6. Bellotti, V. What You Don't Know Can Hurt You: Privacy in Collaborative Computing. In Proc. HCI on People and Computers XI, ACM Press (1996), 241-261.
7. Consolvo, S., Smith, I. E., Matthews, T., LaMarca, A., Tabert, J., and Powledge, P. Location disclosure to social relations: why, when, & what people want to share. In Proc. CHI 2005, ACM Press (2005), 81-90.
8. Flickr – http://flickr.com
9. Gemmell, J., Williams, L., Wood, K., Lueder, R., and Bell, G. Passive capture and ensuing issues for a personal lifetime store. In Proc. CARPE 2004, ACM Press (2004), 48-55.
10. Good, N. S. and Krekelberg, A. Usability and privacy: a study of Kazaa P2P file-sharing. In Proc. CHI '03. ACM Press (2003), 137-144
11. Gruteser, M. and Grunwald, D. Enhancing location privacy in wireless LAN through disposable interface identifiers: a quantitative analysis. Mob. Netw. Appl. 10, 3 (2005), 315-325.
12. Gruteser, M., Schelle, G., Jain, A., and Grunwald. D. Privacy-aware location sensor networks. In Proc. HotOS 2003, (2003).
13. Halderman, J. A., Waters, B., and Felten, E. W. Privacy management for portable recording devices. In Proc. WPES 2004, ACM Press (2004), 16-24.
14. Hawkey, K. and Inkpen, K. M. Keeping up appearances: understanding the dimensions of incidental information privacy. In Proc. CHI 2006, ACM Press, 821-830.
15. Iachello, G., Truong, K. N., Abowd, G. D., Hayes, G. R., and Stevens, M. Prototyping and sampling experience to evaluate ubiquitous computing privacy in the real world. In Proc. CHI 2006, ACM Press, 1009-1018.
16. Ito, M., Okabe, D., and Matsuda, M. Personal, Portable, Pedestrian: Mobile Phones in Japanese Life. MIT Press (2005).
17. Hong, J. and Landay, J. An architecture for privacy-sensitive ubiquitous computing. In Proc. Conference on Mobile Systems, Applications, and Services, 2004.
18. Hull, R., Kumar, B., Lieuwen, D., Patel-Schneider, P. F., Sahuguet, A., Varadarajan, S. and Vyas A. Enabling Context-Aware and Privacy-Conscious User Data Sharing. In Proc. MDM 2004, IEEE (2004), 187-198.
19. Kindberg, T., Spasojevic, M., Fleck, R. and Sellen, A. The Ubiquitous Camera: An In-depth Study of Cameraphone Use. IEEE Pervasive Comp. 4, 2 (2005), 42-50.
20. Lederer, S., Hong, I., Dey, K. and Landay, A. Personal privacy through understanding and action: five pitfalls for designers. Personal Ubiquitous Computing. 8, 6 (Nov. 2004), 440-454.
21. Palen, L. and Dourish, P. Unpacking "privacy" for a networked world. In Proc. CHI 2003, ACM Press (2003), 129-136.
22. Patil, S. and Lai, J. Who gets to know what when: configuring privacy permissions in an awareness application. In Proc. CHI 2005, ACM Press (2005), 101-110.
23. Spiekermann, S., Grossklags, J., Berendt, B. Stated Privacy Preferences versus Actual Behaviour in EC environments: a Reality Check. In Proc. WI-IF 2001.
24. Stutzman, F. An Evaluation of Identity-Sharing Behavior in Social Network Communities. In Proc. iDMAa and IMS Code Conference, 2006.
25. Tang, K. P., Keyani, P., Fogarty, J., and Hong, J. I. Putting people in their place: an anonymous and privacy-sensitive approach to collecting sensed data in location-based applications. In Proc. CHI 2006, ACM Press (2006), 93-102.
26. van House, N., Davis, M., Ames, M., Finn, M., and Viswanathan, V. The uses of personal networked digital imaging: an empirical study of cameraphone photos and sharing. In Ext. Abstracts CHI 2006, ACM Press (2006), 1853-1856.
27. Varian, H. Economic Aspects of personal privacy. In Privacy and Self-Regulation in the Information Age, NTIA report (1996).