

# The Post that Wasn't: Exploring Self-Censorship on Facebook

Manya Sleeper, Rebecca Balebako, Sauvik Das,  
Amber Lynn McConahy, Jason Wiese, Lorrie Faith Cranor

Carnegie Mellon University  
Pittsburgh, PA  
{msleeper, balebako, sauvik, amberlynn, wiese, lorrie}@cmu.edu

## ABSTRACT

Social networking site users must decide what content to share and with whom. Many social networks, including Facebook, provide tools that allow users to selectively share content or block people from viewing content. However, sometimes instead of targeting a particular audience, users will self-censor, or choose not to share. We report the results from an 18-participant user study designed to explore self-censorship behavior as well as the subset of unshared content participants would have potentially shared if they could have specifically targeted desired audiences. We asked participants to report all content they thought about sharing but decided not to share on Facebook and interviewed participants about why they made sharing decisions and with whom they would have liked to have shared or not shared. Participants reported that they would have shared approximately half the unshared content if they had been able to exactly target their desired audiences.

## Author Keywords

Social networking sites; usability; privacy; self-censorship

## ACM Classification Keywords

H.1.2 User/Machine Systems: Human factors

## INTRODUCTION

Social Networking Site (SNS) users make decisions about what content to share and with whom. Sharing inappropriately can result in consequences ranging from regret to job loss [20]. SNSs provide tools that allow users to share content with some people and block other people from viewing content. However, sometimes instead of targeting a particular audience, users will self-censor or choose not to share.

Lampinen et al. describe self-censorship as one of the techniques SNS users rely on to manage the co-existence of different social groups on SNSs [13]. Self censorship is an important ability; SNS users choose not to post content for a

variety of reasons, including to protect their own and others' privacy and to prevent regret [12, 13, 20, 22]. In this paper we explore users' self-censorship decisions on Facebook, as well as the types of content they choose to self-censor.

While self-censorship can be a desirable behavior both on- and offline, users sometimes choose to self-censor on SNSs because available access-control tools don't meet their needs. For a subset of self-censored content, users choose not to share because they would like only specific audiences to see the content, and those audiences are difficult, or impossible, to target given current interface design. We focus on understanding this subset of self-censored content and the potential impact of optimizing selective sharing tools to allow users to share this content with their preferred audiences.

Selective sharing [9] occurs when users can share with only their desired audiences, by selecting people to share with or block. We look specifically at sharing that could potentially have occurred if participants had been able to target exactly their desired audiences (optimal selective sharing). Our intention is to explore the potential ability of tools to allow users to share a subset of currently unshared content.

This paper has two primary contributions. Self-censorship has been established as a means for preserving SNS privacy but has not been thoroughly examined. We seek to expand understanding of types of, and reasons for, self-censorship on SNSs by examining self-censorship on Facebook. Second, we provide insight into the subset of self-censored content users could potentially share given improved SNS selective sharing mechanisms, as well as the types of tools that would be necessary to allow users to share this content. Previous work tended to focus on shared content; by focusing on unshared content, we provide additional insight for creating selective-sharing tools.

To address these issues we examined the types of Facebook content that users were not sharing, and why. Specifically, we looked at the following research questions:

- **Q1:** What types of content are users currently not sharing?
- **Q2:** Why do users choose not to share different types of content?
- **Q3:** What subset of content that users currently don't share (unshared content) could potentially be shared if they could

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CSCW'13, February 23–27, 2013, San Antonio, Texas, USA.

Copyright 2012 ACM 978-1-4503-1209-7/13/02...\$15.00.

exactly target their intended audiences (i.e., given optimal selective sharing mechanisms)?

- **Q4:** What attributes typify the groups with whom users would like to selectively share currently unshared content?

We ran a weeklong, 18-participant, diary study during which we asked participants to send us Facebook posts they thought about posting but decided not to share. We used an in-lab interview to gather additional information about the content. We found that participants chose not to share a variety of types of content, especially entertainment and personal content. Participants would have shared approximately half of the unshared content if they were able to share with or block some combination of specific individuals, groups of individuals, and more ambiguous, attribute-defined groups.

## RELATED WORK

Prior work on SNS user behavior, online identity management, and grouping for online privacy informed our study. We expand on prior work by examining unshared content and by looking at a subset of self-censored content to understand the interplay between sharing decisions and grouping for audience management.

### User behavior on social networks

Users behave in risky ways when using SNSs, often due to audience mismanagement. Reynolds et al. found that Facebook users' behaviors did not coincide with their privacy concerns [17]. Stutzman and Kramer-Duffield compared intended audiences to actual audiences on friend networks, in a privacy context. They found that friends-only privacy settings could include "weak tie" relationships that presented privacy challenges [19].

Users can also regret SNS posts. Wang et al. found that one of the reasons Facebook users regretted posts was visibility to unintended audiences. Self-censorship was among the responses that they saw [20]. We complement this work on sharing regrettable content by exploring the subset of content users self-censor.

### Identity management

Farnham and Churchill showed that users often manage multiple identities when sharing online. Users behave differently based on the social groups with which they interact, many of which are affiliated with different social roles [5]. These findings suggest that SNSs would benefit from tools that enable users to create appropriate boundaries for sharing.

When users make sharing decisions on SNSs, they are often trying to control how they present themselves. On Twitter, Marwick and boyd observed the tendency of users to collapse multiple audiences into a single context. The imagined audience with which the users interacted affected their sharing behaviors, including levels of self-censorship [14]. Strater and Lipford found that this perceived audience tended to narrow over time, often falling out of line with users' actual privacy settings [18]. Lampinen et al. found that Facebook users identified with multiple groups that co-existed on Facebook and used proactive strategies, including self-censorship, to manage this "group co-presence" [13].

SNS users also try to manage interpersonal boundaries on SNSs. On Facebook, they use mechanisms not formally built into the interface. These strategies include using self-censorship to avoid disclosing inappropriate information about others [12]. Wisniewski et al. also found that users employed "coping mechanisms," such as posting less content, to manage social boundaries on SNSs [22].

## Grouping

One way to manage audiences is to create groups of individuals to share with or block from viewing content. Karr-Wisniewski et al. included group creation as a boundary regulation mechanism that allows users to define "relationship boundaries" on several SNSs [10].

Olson et al. found that access preferences for different types of content depended on the types of information they contained and the people who would potentially view them, thus making sharing conducive to grouping [16]. Kelley et al. asked users to create groups of their Facebook friends and found that common groups included "general friends," "college," "other education," "family," and "work." Static groups created ahead of time tended not to perform well when faced with real-time sharing decisions, suggesting a need for dynamic grouping [11]. Wiese et al. also found that "self-reported closeness," co-location, "frequency of communication," and common contacts were predictive features of sharing behavior [21]. Looking at Google+ circle names, Kairam et al. found that the majority of groups could be sorted into some combination of "life facet" (e.g., "work" and "school") and "tie strength" ("strong tie" and "weak tie") groups [9].

Mechanisms have been proposed to automatically support users' abilities to create groups on SNSs. Jones and O'Neil compared an automated SCAN clustering algorithm to groups created using card sorting and found that human-created groups outperformed algorithm-created groups in best and worst cases [8]. Fang and LeFevre explored the possibility of an interactive tool for specifying privacy preferences by asking users to make sharing decisions about a subset of their friends [4]. ReGroup is a more interactive machine-learning tool that looks for patterns in who users choose to share with through continually updated groups; users preferred it to an equivalent Facebook-style interface [1].

The underlying goal of these systems and studies was to understand sharing decisions and the groups with whom users wanted to share. In this work, we also examine unshared content. We explicitly explore reasons users do not share and the groups of people with whom a subset of this unshared content could potentially be shared.

## METHODOLOGY

We wanted to determine what users were not sharing, and why (Q1,2), as well as the subset of unshared content that could potentially be shared using optimal selective sharing (Q3). We also wanted to explore attributes of the groups with whom our participants would have wanted to selectively share or block from viewing this subset of unshared content (Q4).

The study had two phases and took place in April and May of 2012. First, participants took part in a weeklong diary study during which they used SMS messaging to report all instances of unshared content on Facebook (i.e., content intentionally self-censored). Participants also filled out nightly surveys to further describe unshared content and any shared content that they decided to post on Facebook. Next, qualified participants took part in in-lab interviews. The interview provided more details about reported, unshared content and a better understanding of participants' decisions on when to share. We asked about participants' reasons for deciding against sharing, as well as the people, if any, participants hoped would see or wanted to block from viewing their content.

We iteratively coded each piece of unshared and shared content that we were able to ask participants about in the final interviews (122 piece of unshared and 83 pieces of shared content) for types of content, the types of groups the participant wanted to share with or block from viewing the content, and the participant's reasons for not sharing.

### Recruitment and Demographics

We recruited 18 participants from a campus participant pool website, Craigslist, flyers, and a targeted Facebook ad. They were screened online for high English proficiency, a minimum age of 18, at least 6 months of Facebook use, frequent Facebook use (more than once per week), texting regularly (at least once per week), and having frequently held back content on Facebook (at least 3 pieces of unshared content over the past week). Thirty potential, qualified participants were sent online instructions for participating in the diary study, including 8 students. Nineteen of the 30 recruited participants completed at least one nightly survey, and 16 out of 30 completed the full study including the final interview. Two additional qualified student participants received the link to the instructions from friends and participated in the full study, resulting in 18 participants.

Participants ranged in age from 20 to 51. Ten were female, and seven were students. Table 1 summarizes participant demographics. Participants were compensated \$20 for the final interview and \$2 per nightly survey completed, up to a total of \$34. We also reimbursed \$6 for parking.

### Diary Study

The diary study lasted seven days. Participants had continuous access to a set of online instructions. Participants sent SMS text messages whenever they thought "of things that they would like to post on Facebook but decide[d] not to post." They were asked to describe the potential post and include the type of post it would have been (e.g., wall post, photo, link, etc). This SMS-based approach was based on the technique used by Brandt et al. [2].

Every night, each participant was also sent a link to an online survey, which contained questions for each piece of unshared content. Participants could provide more detailed descriptions of unshared content and reasons for not posting. Participants were also prompted with questions about the people with whom they would have liked to share or would have liked to block from viewing each item. These questions were

<i>Code</i>	<i>Age</i>	<i>Gender</i>	<i>Occupation</i>	<i>Unshared items</i>	<i>Shared items</i>
P01	20	F	engineering student	7	4
P02	26	M	engineering student	8	4
P03	20	M	bus admin student	1	3
P04	33	F	social science student	24	0
P05	30	M	dental student	4	3
P06	26	M	unemployed	2	2
P07	23	F	non-profit	13	5
P08	29	F	art/writing/journalism	3	11
P09	25	F	non-profit	1	8
P10	28	M	human resources	10	4
P11	26	M	unemployed	6	7
P12	25	F	art/writing/journalism	4	5
P13	51	F	bus/mgt/fin	9	7
P14	24	F	lab mgr	8	1
P15	24	M	art/writing/journalism	2	7
P16	32	F	unemployed	12	8
P17	22	M	architecture student	4	2
P18	21	F	engineering student	4	2

Table 1. Participant demographics

open ended, allowing participants to either name specific individuals or define their own notions of the people that would have constituted a "group" for sharing. The interface allowed participants to add additional unshared content, so they were not bound by the SMS messaging system. The survey also asked participants to describe content they had shared that day (shared content). If a participant had not shared any content, they were asked to fill out an auxiliary question about why they had not shared. We hoped to ensure a baseline level of effort and minimize incentives not to report.

We used this diary study and survey system because users can think of unshared content throughout the day, and we wanted to capture this as it occurred. This technique also allowed a participant to provide a quick "digest" of unshared content through the SMS system, and, if they were busy, return to the survey at a more convenient time to provide details.

### Semi-Structured Interview

Participants who completed at least four surveys qualified for a final, in-lab interview (18 participants). We chose a semi-structured approach, which allowed us to capture similar types of data across all the interviews while maintaining the flexibility to explore the varied content reported. The interviews each lasted approximately one hour and occurred in a lab. One researcher served as the primary interviewer and interacted with the participant. A second researcher served primarily as a note taker. All interviews were audio recorded.

We used participants' shared and unshared content to explore our four research questions. We went through each piece of unshared content, and the participant's nightly survey responses, and probed for details on the content, reasons for wanting to share and not sharing, and, when relevant, details about the groups the participant would potentially have wanted to share with or block. For example, we asked the participant to describe the unshared content in more detail, to further explain why they decided not to post it, and to expand on their relationships with or common characteristics of the people they would have wanted to share content with or block. We also asked participants about their willingness to share each item, given selective sharing mechanisms, as

well as for additional details on shared content and a series of questions on SNS usage and privacy habits.

Prior to study launch, we refined our methodology by piloting with 10 additional participants who are not included in analyses.

### Data coding and analysis

To analyze the data, we looked at each piece of nightly survey content that we were able to discuss with participants during the final interviews. We removed any content that we were unable to discuss in the interviews, either because the participant did not complete enough nightly surveys to qualify for the interview or because the participant submitted too many items to allow for discussion of all content (only P04, who submitted 52 pieces of unshared and 32 pieces of shared content). Table 1 lists the number of shared and unshared items included in the analyses for each participant. We coded each item for the type of content, the participant's reason for not sharing, and the types of people with whom the participant would have wanted to share and/or block (where relevant). Our coding process was based on that used by Kairam et al. to code content shared on Google+ [9] as well as the technique used by Naaman et al. to code Twitter data [15].

To create codes, two researchers each independently coded a random selection of 50 items, using data from the nightly surveys and notes from the interviews. Based on those codes, the researchers collaboratively created a set of high level codes and independently coded the majority of the remaining data. The two researchers then iteratively coded all the data with updated codes two additional times. Between each iteration, the researchers updated the coding scheme based on shortcomings from the previous round. Using the final codes, the researchers went through their independent codings and discussed and agreed on any codes that differed. This process produced the set of codes used in the analyses.

Analyses presented in this paper are intended to be entirely qualitative. Numbers are intended to illustrate results from the sample but are not meant to indicate statistical significance or quantitative generalizability. Examples are only intended to illustrate trends seen during the study.

## RESULTS

Participants self-censored a variety of types of content, especially those related to external material (content unrelated to the participant), like entertainment. They most commonly chose not to share because they were trying to control how they presented themselves, and they would have shared about half of unshared content, given optimal selective sharing. The groups participants wanted to use for this optimal selective sharing included specific individuals, specific groups of individuals, as well as more dynamic groups that depended on context.

In this section we outline the types of unshared content, the reasons participants chose not to share, how much of that content would have been amenable to selective sharing, and the characteristics of the groups that would have been necessary

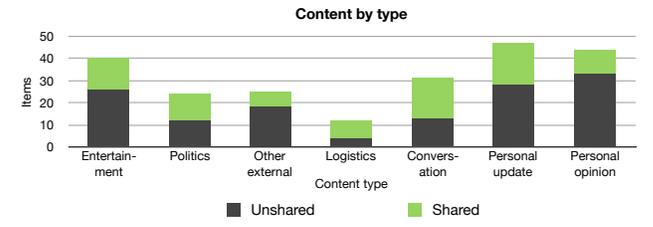


Figure 1. Shared and unshared content by type

to allow participants to have selectively shared that subset of unshared content.

### Q1: Types of unshared content

We coded shared and unshared items into one or more of seven categories. We split items based on whether they were external content (e.g., entertainment, politics, or other external content), personal information, or related to planning or conversation. This led to three external categories: **entertainment**, **politics**, and **other**; two personal content categories: **personal update** and **personal opinion**; and categories for **conversational** content and **logistics**. Figure 1 shows the number of unshared and shared items in each category. These categories were roughly based on those Kairam et al. used for reasons to share Google+ content [9]. Items could be coded in multiple categories.

#### External content

External content included references unrelated to the participant. It could be intended to entertain, inform others, or allow the individual to express an opinion about the outside world. There were three subcategories: **entertainment**, **politics**, and **other**. We also noted when it included an opinion.

**Entertainment:** Examples included references to or articles about movies, television, sports, or music. This category contained 21% of unshared and 17% of shared content (26 and 14 items respectively). Unshared entertainment content tended to contain more material that could potentially offend. Several items contained explicit language or drug references. P17, for example, considered sharing a “weird” video that included drug-related content, but decided not to because her “family in Austin is really religious... they would’ve called [her] about it.”

Half of unshared entertainment items included opinions, as opposed to three shared items. In contrast, shared entertainment content tended to advertise without a stated opinion. For example, several participants posted to advertise concerts.

**Politics:** Content that referenced politics, current events, or activism was coded as politics, which included 10% of unshared and 15% of shared content (12 items each). The majority of the unshared political content was considered potentially controversial. P04, for example, decided not to post a “Link to article about young black republicans” to try to avoid controversy. On the other hand, shared content trended more toward current events. P12, for example, “posted a link to an article about the slow recovery from the BP Oil Spill in Louisiana.” She explained “it was one of the few instances

when there was something kind of political and I put it up anyway,” because it “was the true story and what’s seen.”

**Other external content:** This category included items that referenced content not related to the participant, entertainment, or politics. It included facts, quotes, pictures, and jokes, and included 15% of unshared and 9% of shared content (18 and 7 items respectively). Many sharing decisions depended on context. For example, P18 considered sharing “a recipe for a cake I saw posted by a friend from high school” but decided not to share because “I haven’t spoken to her in a while and it would be awkward.”

#### Personal content

Personal content related to a participant’s life or general opinions and included **personal updates** and **personal opinions**.

**Personal updates:** These were items that described something that happened in a participant’s life. Examples included content about the participant’s day or about events the participant took part in, including photos. Personal updates made up 23% of the unshared and shared content (28 and 19 items respectively). Participants often decided not to post personal updates because they were too “frivolous” or not “creative” enough. For example, P10 thought about posting “Kicking ass and taking names!!! Happy Monday!!!” but decided not to because it was “very vague very generic, didn’t think it was very creative.” Participants also didn’t post because they felt their personal updates were too negative or sounded like they were “whining.” P16, for example, thought about posting about a fight between her and her boyfriend but decided not to because it was “grumpy.” Shared personal updates tended to be relatively positive or straightforward.

**Personal opinions:** Opinions unrelated to external content were coded as personal opinions. These included how the participant generally felt about life, such as “having a stressful day,” or more general opinions such as “We are way too old to be celebrating 420 day.” Personal opinions included 27% of unshared and 13% of shared items (33 and 11, respectively). As with updates, many unshared personal opinions tended to be negative. Participants also worried that some might offend or start an argument. For example, P05 considered posting about how she disapproved of the Pokemon tattoo her brother-in-law was considering but decided against the post “because he wouldn’t have liked it and it really wouldn’t have made a difference anyway.”

#### Conversation and planning

**Conversational:** This category included conversational niceties without additional content, such as birthday wishes or replies to posts that did not include additional content. This category included 11% of unshared and 22% of shared content (13 and 18 items respectively). Participants tended to not post conversational content based on potential social awkwardness. For example, P07 thought about wishing a friend happy birthday but decided not to because she hadn’t “talked to him in a long time.”

**Logistics:** Logistics included posts related to making plans. More were shared (8 items, 10%) rather than unshared (4 items, 3%). When participants didn’t share, it tended to be

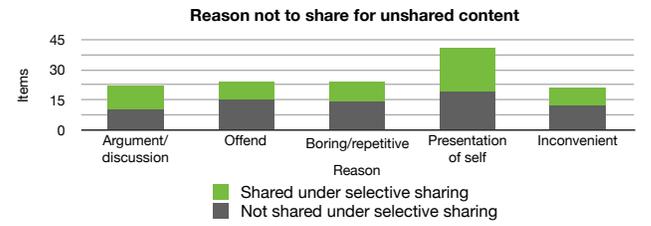


Figure 2. Reasons not to share unshared content

due to offline, social reasons. For example, P05 decided not to discuss lunch plans because he didn’t know one of the people involved in the conversation well enough.

#### Q2: Reasons for not sharing

We were also interested in reasons for self-censorship. We asked in the surveys and the final interview, for each unshared item, why participants decided not to share. Responses tended to fall into one or more of five categories:

- **Argument/discussion:** Didn’t want to start or participate in an argument or discussion.
- **Offend:** Didn’t want to offend or hurt someone.
- **Boring/repetitive:** Felt the content was redundant, boring, or not interesting enough.
- **Presentation of self:** Felt the content went against the way the participant wanted to present him/herself (e.g., “seemed silly” or “don’t like to post that kind of thing”).
- **Inconvenient:** Prevented from posting due to time or technology (e.g., location made it difficult to post).

Figure 2 summarizes the number of items in each category. Presentation-of-self issues were most common (34%, 41 items) by a small margin; however, the remaining reasons each applied to approximately 20% of items. Percentages add up to over 100% because some items were not shared for multiple reasons.

Several reasons emerged more frequently for different types of content. Approximately half of entertainment and personal updates weren’t shared because of presentation-of-self concerns, and slightly over half of political items weren’t shared because participants didn’t want to start or participate in arguments or discussions. Almost half of personal updates also weren’t shared because participants were worried that the items would be boring or repetitive.

#### Q3: Potential for selective sharing

Participants would potentially have shared a subset of the unshared content if they could have exactly targeted particular audiences under optimal selective sharing. To isolate this subset, we used two Likert scale questions to judge participants’ willingness to share given an optimal ability to selectively share with desired audiences. For each item of unshared content for which a participant provided a potential group that they would have liked to have shared with or blocked, we asked the participant to imagine that they either “could have shared this content only with” the people they wanted to share

it with or could have “prevent[ed]” the people they didn’t want to see it from viewing the content. To increase generalizability, we did not specify the interface that would be used to share the content, only that it would exactly target desired audiences. Responses were on a five-point Likert scale where a one was “very unlikely” and a five was “very likely.” We consider a participant who indicated above a three for either question to have been potentially willing to share given optimal selective-sharing mechanisms. If a participant answered above a three for sharing and/or blocking selected people, we analyzed the people with whom the participant indicated they wanted to share and/or block. Overall, 60 out of 122 unshared items (49%) would have potentially been shared given optimal selective sharing. Of those, 57 would have been shared if the participants could have shared with only a desired set of people, and 25 would have been shared if the participants could have blocked people from viewing content. Figure 3 shows each participant’s potential willingness to share given optimal selective-sharing tools.

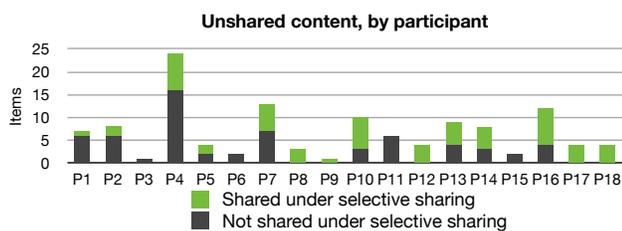


Figure 3. Unshared content participants were willing or unwilling to share, given optimal selective sharing

Although our small sample limits the generalizability of these results, this indicates that participants could have potentially shared a relatively large subset of their self-censored content if they could have exactly targeted desired audiences.

#### Types of content for selective sharing

Participants would potentially have selectively shared approximately half of each type of content. External content tended to be amenable to selective sharing because participants wanted to share items with people who would have been interested and block people who might have been offended. For example, P08 considered posting “a lot of angry status updates” during a hockey game she was watching but decided not to because others were already doing so. She would have posted if only her hockey friends had been able to see the posts because they would have been interested. This was common for unshared entertainment content, which participants often felt only a subset of people would be interested in and/or other people might be offended by. A similar dynamic occurred with political content, with more of an emphasis on avoiding debate. This might have been expected because participants often decided not to share such political content to avoid argument or discussion.

Approximately half (15 out of 28 items) of personal updates would have potentially been shared with optimal selective sharing. Participants often wanted to share with people who were “close friends” or who they saw regularly who would understand or appreciate the posts. For example, P16, who

considered posting about a fight with her boyfriend, only wanted to share with a small group of friends because “they can relate, because they know more about me and we talk about more personal things with each other.” Participants would have potentially shared 42% of personal opinions (14 items) given optimal selective sharing. In several cases, participants wanted to share opinions with people who would understand the context. P07, for example, wanted to post “don’t have pets if you’re not prepared to take care of them!!!” after a bad experience cat sitting but only wanted mutual friends of the person who the post was directed at to see it.

Items for which participants wanted to control presentation of self or didn’t want to start an argument or discussion were most amenable to selective sharing (slightly over half). Participants tended to want to share items with presentation-of-self issues with close friends or people who would be interested in or understand the content. For example, P07 thought about posting about her frustrations at her babysitting job to get advice, but chose to self-censor because she didn’t think babysitting was “cool.” She would have preferred to share only with particular people who also babysat. Participants who chose not to share because they didn’t want to get involved in an argument or discussion tended to want to share with people who agreed or thought the same way about potentially controversial content. P04, for example, considered posting a link to an article about “cohabitation and divorce.” She decided against posting because she had a lot of Facebook friends who were religious Christians who disapproved of cohabitation, and she wanted to avoid a long discussion.

#### Q4: Types of groups

For participants to selectively share the desired subsets of content, they would need to be able to specify, using the interface, the individuals or groups with whom they wanted to share. We asked participants to specify who they did and did not want to view each unshared item, so we could understand the kinds of groups participants would need to create to express their optimal selective-sharing preferences. We looked at the number of people in, and characteristics of, the groups.

##### Number of people in group

We coded the people with whom participants did or did not want to share each item into one or more of the following: a **specific person** (e.g., “my sister,” “Tim”); **specific people** defined as a countable set of people (e.g., a group of ten close friends); or an **ambiguous group** defined by one or more attributes or relationships (e.g., “hockey friends”). Percentages add up to over 100% because participants sometimes specified multiple sets of people they wanted to share with or block (e.g., a specific person and an ambiguous group).

Participants specified individuals or groups with whom they wanted to share for 92 out of the 122 unshared items (75%). For the remaining items they were willing to share with everyone or weren’t willing to share with anyone. Of the groups associated with the 92 pieces of content, we looked at those with which participants would have shared given optimal selective sharing (53/92 items). Of the groups that would have been useful for selective sharing, 47% (25) were ambiguously defined, 30% (16) were groups of specific people, and 33%

(17) were specific individuals. Participants specified individuals or groups to block from viewing content for 57 items (47%). Of these, participants said that blocking 23 of the groups would allow them to share the content items under optimal selective sharing. Of these groups that would have been useful for selective sharing, 74% (17) were ambiguously defined, 13% (3) were specific groups of people, and 26% (6) were specific individuals.

These results partially imply that our participants were not using Facebook's current custom privacy settings. Participants indicated that they wanted to share with single individuals or specific people, which could be done on Facebook. Participants' reasons for not doing so are further addressed in the Discussion. More ambiguous groups also accounted for a relatively large percent of potentially useful groups. They tended to be attribute-based and consisted of both concrete groups (e.g., classmates) and more context-specific groups (e.g., people who would disagree with a post). Such groups would require more extensive user effort or new tools. For example, a user could set up a school-based group ahead of time but might have more difficulty creating a group defined by people's feelings toward a topic.

#### Group characteristics

We also looked at characteristics associated with the individuals and groups with whom the participants would have liked to have selectively shared. We coded each individual or group into one or more of the following categories:

- **Work/school:** Work or school at any stage of the participant's life (e.g., coworkers, classmates, high school).
- **Demographics:** Age, gender, geography, race (e.g., younger relatives, male/female).
- **Family:** Relatives (e.g., mother, extended family).
- **Close friends:** Close relationships (e.g., close friends, people seen on a regular basis, boyfriend/girlfriend).
- **Not close friends:** Lacking close relationships (e.g., "not close to," someone never met, "frenemies").
- **Relationship to post:** Interested in the post, felt a certain way about the post, personally relevant to the post (e.g., "feel the same way as me," person the post was directed at, interested in the content).

A summary of the characteristics of the groups associated with the items participants would have been willing to share if they could have targeted or blocked specified people is in Tables 2 and 3. These categories are similar to those that emerged in other work on grouping [9, 11, 13, 21].

Based on the 53 items that participants said that they would have been willing to share if they could have shared with selected individuals, the most frequent attribute was the person or group's relationship to the post (62%, 33 items). It was slightly more likely to occur for a specific person or ambiguous groups. Participants tended to want to share only with people at whom the content was directed or people who would be interested in an item. For example, P08 "had tickets

SHARED	Total	Specific person	Specific group	Ambiguous group
	Work/school	17	0	6
Demographics	10	2	4	6
Family	6	3	2	3
Close	9	3	6	0
Not close	2	1	1	1
Relationship to post	33	11	11	16
<b>Total items</b>	<b>53</b>	<b>17</b>	<b>16</b>	<b>25</b>

**Table 2. Characteristics of groups participants wanted to share with for optimal selective sharing, by type of group**

BLOCKED	Total	Specific person	Specific group	Ambiguous group
	Work/school	8	1	2
Demographics	4	0	0	4
Family	2	0	2	1
Close	1	1	0	0
Not close	7	2	0	6
Relationship to post	13	5	1	10
<b>Total items</b>	<b>23</b>	<b>6</b>	<b>3</b>	<b>17</b>

**Table 3. Characteristics of groups participants wanted to block for optimal selective sharing, by type of group**

to an advanced screening of The Avengers and almost posted about how excited [she] was to see it using a bunch of profanity." She wanted to share it with her friends who liked comic books and video games and was "sure I would have posted it if it was just like the people I know like it would've seen it." But, as she pointed out, "I don't have a group for comic book friends, mostly because I don't know who would like it, there are people who like things I don't know about." Determining these more complex, ambiguous relationships to posts that rely on time-of-post decisions would be relatively difficult.

The second most common attribute was work/school (32%, 17 items), which only occurred for groups of specific people and ambiguous groups. For specific groups of people, this attribute tended to be associated with a close group of friends that included people from school. Such specific groups would be relatively easy to define using a selective-sharing mechanism, because they are at least partially defined by a concrete common attribute. When participants defined more ambiguous groups using work/school, they tended to be people who would be interested in the content and who either currently went to school or worked with the participant or went to school with the participant in the past. P08, for example, wanted to share content about a hockey game with "hockey friends," who also tended to be college friends. Defining these more ambiguous groups would be more difficult with current tools and might not be encompassed by the work/school attribute.

Relationship to post also occurred most frequently for the 23 items that participants would have been willing to share if they could have blocked a specific group of individuals (56%, 13 items). Again, it was more likely for specific people and ambiguous groups. For specific people, participants tended to want to block the person who originally posted the content they were planning to comment on or people who might be offended. For example, P12 considered posting "some links

to articles I read on NPR and WeArePowerShift.org - very political stuff." She didn't mind the general public seeing the content, but wanted to block her boyfriend's dad and other conservative friends from viewing it.

Work/school was also the second most common attribute for people participants wanted to block. However, for blocking selected people, but not for sharing with selected people, "not close friends" emerged as the third most common attribute. This attribute characterized specific individuals and ambiguous groups. Participants tended not to want to share more personal content with people who didn't know them as well. For example, P14 considered posting about a stressful day but didn't want to share it with people she wasn't as close to. As she put it "if they're better friends with you then they don't necessarily care if you're venting or complaining." Such groups would be relatively difficult to capture using current tools because they tended to be context-specific. They ranged from friends-of-friends to the "frenemies" P13 considered too "weird" to know about her evening plans.

### LIMITATIONS

This study had several primary limitations. First, it was qualitative, limited to a small sample, and did not consider unshared content in a cross-cultural context. Conclusions, therefore, lack broad generalizability. Our sample also skewed young. This age skew partly reflects SNSs; in 2010, approximately three-quarters of SNS users were 35 or younger [7]. However, future work examining differences in self-censorship across age levels would also be interesting.

Using a diary study also introduced bias. Participants were aware of the purpose of the study; as part of a "study on Facebook usage" they were asked to report "everything you think about sharing on Facebook but decide not to post." Texting in content and filling out surveys likely primed them to think about Facebook, unshared content, and audiences. When asked, participants did not feel they had changed their behavior due to the study. However, about a third mentioned being more aware of what they posted and unshared content. This may have pushed them to think more about self-censorship.

The study structure also relied on self-reported data based on hypothetical scenarios. Actual behavior does not always match what participants say they will, or mean, to do. These issues could be partially addressed in future studies by designing studies to focus on actual behavior. One possibility would be to examine the differences in types and levels of sharing that occur under different interface designs or when a user is instructed to share in different manners (e.g., posting only for oneself, for close friends, etc). Focusing on behavior might reduce the limitations of self-reported hypothetical data and could allow for less priming.

Finally, this study was only able to capture a subset of self-censored content. There is a spectrum of how likely a user would be to post an unshared item, which ranges from content they are almost prepared to post (e.g., at the keyboard and have fully composed) to vague ideas that they decide they probably shouldn't post. Responses to this study mostly included more fully-thought-out ideas, although there were

some vague thoughts. It likely missed more of what people self-censor before ideas are fully developed.

Participants may also have been less likely to report sensitive or embarrassing content. To reduce participants' sensitivity, we avoided face-to-face interaction until after the diary study. We believe this was at least partially successful; participants reported some potentially sensitive items that included profanity, political opinions, and drug references. There is also likely content that is so sensitive that it is self-censored in an ingrained way and was not captured. Future work might accompany an approach like this by using a survey to try to probe more ingrained self-censorship by asking participants if they would consider posting content on a variety of more extreme topics (e.g., sexual content, violence, etc.).

### DISCUSSION AND CONCLUSIONS

Participants self-censored content, often because they wanted to manage how they presented themselves to various audiences or to avoid argument or discussion. They indicated that they would have potentially shared about half of this self-censored content, across content types, given the ability to optimally target audiences. The people participants wanted to share with, or block, ranged from those captured by current Facebook privacy controls to ambiguous, context-specific groups that would require more sophisticated mechanisms. We discuss why participants seemed not to use Facebook's custom privacy settings, participants' uses of alternatives to self-censorship, and some high-level design suggestions for capturing selective sharing preferences.

#### Reasons for not using Facebook custom privacy settings

In general, participants didn't use Facebook's custom privacy settings to control who could see content. At the time of the study (April-May 2012), Facebook offered the ability to set the visibility of a post to the general public, friends-only, lists of friends defined either automatically by Facebook or manually by the user, or a post-specific list of people. Users could set a default to public, friends, or a custom list. Most participants used a friends-only default setting. Some, like P03, felt that friends should be able to "see everything," while others, like P15, assumed that anything posted on Facebook was available for a general audience. A few participants had set up friend lists at some point but tended to have used them once or set them up and then stopped maintaining them. P18, for example, had used the friend lists feature when it first appeared but hadn't continued to actively use them. This behavior is consistent with the literature. Kelley et al. found that users tended not to want to use previously created groups for sharing [11]. Strater and Lipford found that users both had trouble understanding Facebook privacy settings and tended not to revisit them once set up [18], and Karr-Wisniewski et al. found that users did not use provided grouping tools [10].

Several participants found Facebook's grouping and privacy features too confusing or difficult to use. P01, for example, hadn't recently adjusted her settings, even though she realized Facebook had changed their privacy settings, and said that "it kind of worries me that I haven't messed with it." She both found the settings confusing, admitting that "I'm not really

sure how lists work” and felt that Facebook was something she quickly logged on and checked, not something she sat down and used long enough to bother with the settings. P08 pointed out that she frequently posted status updates from the Facebook mobile app on her phone, where “it’s easier to just not post than to go in and mess with the settings.”

Other participants didn’t trust Facebook. Some didn’t trust Facebook to maintain their privacy settings. P16 put it “Like maybe one day they’ll just take off all the permissions, like just for fun. . . so I never know if that’s going to happen, since Facebook seems to have a negative track record in most people’s minds, I just try to censor myself.” Other participants didn’t trust themselves to configure the privacy settings and understand how they would propagate.

This suggests that for preferences that could have been captured by Facebook’s current tools, users might require better, built-in, education about Facebook’s privacy controls, a better interface, or an overall increased level of trust in Facebook’s data privacy. One potential direction might be to increase the visibility of available tools and their impact on sharing. Many participants seemed confused about available custom privacy settings and friend lists and how they could be used; increasing transparency could increase their abilities to use such tools and might potentially increase trust in Facebook.

#### **Alternatives to self-censorship**

Lampinen et al. describe strategies SNS users rely on to mitigate the co-presence of multiple social groups on SNSs. One of these strategies is self-censorship. However, users also rely on other strategies, including choosing “channels of communication” and dividing up who can see what content [13] both of which our participants described.

Consistent with “dividing the platform,” Facebook includes “group pages” that allow users to post content to particular groups. Unlike the other privacy features, participants used the group pages to post content for particular, self-selected groups. Several participants used groups affiliated with interests, school, or work to post and read content, participate in discussions, and advertise events to interested people.

Stutzman and Hartzog also describe how SNS users can choose to use different social network services to maintain “privacy, identity, utility, and propriety” [6]. Participants mentioned using different channels of communication as ad hoc privacy controls to varying degrees. Some used chat for more private communications. Others used locked and unlocked Twitter accounts to post personal content they felt was unfit for Facebook. Several felt they could better limit who was following their content on Twitter. P08 for example, was friends with her young sisters on Facebook, and said “I have been kind of watching things I post [on Facebook] because they’re on it a lot, so I’m trying not to swear as much or post a whole lot of crazy things.” Instead, she would “post it somewhere else like on my Twitter or on my blog or something” where she felt her sisters couldn’t find the content as easily. Such behavior was in line with Stutzman and Hartzog’s observation that SNS users relied on a strategy of “practical obscurity” to make it difficult to find certain accounts and main-

tain privacy [6]. Other participants felt they could better track who viewed Twitter content, even with public accounts. This might indicate a desire for a simpler sharing interface. Participants tended to be wary of how their content would be shared through friends of their friends on Facebook.

#### **Potential improvements to selective sharing**

Participant interest in selectively sharing currently self-censored content (approximately half of currently unshared items) suggests a desire for selective sharing tools that would allow them to share with the groups encompassed by their desired audiences. To allow users to share such content would require interface tools that captured the more ambiguous groups participants wanted to target for selective sharing. They would require context-specific information or information often unknown to the user. As outlined in Related Work, machine learning solutions are being developed to help users dynamically create groups. Facebook provides a rich dataset for machine learning, including posts, group pages, likes and a user’s own and friends’ profile data.

Participants often wanted to share with or block audiences that were relevant to posts. This might require tools that could target groups related to topics people are interested in or might disagree with. Defining these traits could require discovering traits, like “geeky comic book friends” or friends with liberal political views that users might not know themselves but might find useful for sharing. One potential method would be to rely on self-identification. For example, a user could indicate that she wanted to share with “comic book lovers” and wait for people to indicate an interest. Alternatively, a user or algorithm could try to identify characteristics that typified a trait. Such a system might be similar to the Hummingbird system for Twitter. Hummingbird uses Twitter hashtags to allow users to indicate the topics of their tweets and then request and approve others’ requests for access by topic [3]. Participants also sometimes wanted to just target individuals involved in a conversation around a post; future tools could make it easier for users to limit a post’s audience to people who had previously been involved in a conversational thread. Some combination of these tools could help facilitate users’ abilities to target content to people relevant to posts.

#### **Conclusions**

Participants chose to self-censor a variety of different types of content. We found that participants most commonly self-censored external content, especially items related to entertainment, closely followed by personal content, including personal opinions and updates. Participants most often tended to self-censor to better control presentation of self. Of the content participants chose to self-censor, they indicated that they would have shared slightly over half if they had been able to exactly target their desired audiences. However, to allow participants to exactly target these audiences would require tools that could specify such desired audiences. We found that participants wanted to target specific individuals and groups of specific people that could potentially be captured using current interface tools but also wanted to share with or block more ambiguously-defined, often context-dependent

groups that would be more difficult to target. Across categories these groups were most frequently characterized by their relationship to the content, including relevance to the post or participation in a conversation around the item.

Based on the content participants indicated that they would share if given the opportunity for optimal selective sharing, there may be opportunities to create improved selective sharing tools both by improving the usability of tools to target individuals and specific groups and by creating selective sharing tools for dynamic group creation.

#### ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grants No. 0946825 and DGE-0903659, as well as the ARCS Foundation. We thank Patrick Gage Kelley, Michelle Mazurek, and Blase Ur for their assistance.

#### REFERENCES

1. S. Amershi, J. Fogarty, and D. Weld. Regroup: interactive machine learning for on-demand group creation in social networks. CHI '12, pages 21–30, New York, NY, USA, 2012. ACM.
2. J. Brandt, N. Weiss, and S. R. Klemmer. txt 4 l8r: lowering the burden for diary studies under mobile conditions. CHI EA '07, pages 2303–2308, New York, NY, USA, 2007. ACM.
3. E. D. Cristofaro, C. Soriente, G. Tsudik, and A. Williams. Hummingbird: privacy at the time of Twitter. Oakland '12. IEEE, 2012.
4. L. Fang and K. LeFevre. Privacy wizards for social networking sites. WWW '10, pages 351–360, New York, NY, USA, 2010. ACM.
5. S. D. Farnham and E. F. Churchill. Faceted identity, faceted lives: social and technical issues with being yourself online. CSCW '11, pages 359–368, New York, NY, USA, 2011. ACM.
6. S. Frederic and H. Woodrow. Boundary regulation in social media. CSCW '12, pages 769–778, New York, NY, USA, 2012. ACM.
7. K. N. Hampton, L. S. Goulet, L. Rainie, and K. Purcell. Social networking sites and our lives. *Pew Internet*, June 2011.
8. S. Jones and E. O'Neill. Feasibility of structural network clustering for group-based privacy control in social networks. SOUPS '10, pages 9:1–9:13, New York, NY, USA, 2010. ACM.
9. S. Kairam, M. Brzozowski, D. Huffaker, and E. Chi. Talking in circles: selective sharing in Google+. CHI '12, pages 1065–1074, New York, NY, USA, 2012. ACM.
10. P. Karr-Wisniewski, D. Wilson, and H. Richter-Lipford. A new social order: mechanisms for social network site boundary regulation. AMCIS 2011, 2011.
11. P. Kelley, R. Brewer, Y. Mayer, L. Cranor, and N. Sadeh. An investigation into Facebook friend grouping. Interact 2011, pages 216–233. Springer Berlin / Heidelberg, 2011.
12. A. Lampinen, V. Lehtinen, A. Lehmuskallio, and S. Tamminen. We're in it together: interpersonal management of disclosure in social network services. CHI '11, pages 3217–3226, New York, NY, USA, 2011. ACM.
13. A. Lampinen, S. Tamminen, and A. Oulasvirta. "All my people right here, right now": management of group co-presence on a social networking site. GROUP '09, pages 281–290, New York, NY, USA, 2009. ACM.
14. A. E. Marwick and D. Boyd. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1):114–133, 2010.
15. M. Naaman, J. Boase, and C.-H. Lai. Is it really about me?: message content in social awareness streams. CSCW '10, pages 189–192, New York, NY, USA, 2010. ACM.
16. J. S. Olson, J. Grudin, and E. Horvitz. A study of preferences for sharing and privacy. CHI EA '05, pages 1985–1988, New York, NY, USA, 2005. ACM.
17. B. Reynolds, J. Venkatanathan, J. Gonçalves, and V. Kostakos. Sharing ephemeral information in online social networks: privacy perceptions and behaviours. Interact 2011, pages 204–215. Springer, 2011.
18. K. Strater and H. R. Lipford. Strategies and struggles with privacy in an online social networking community. BCS-HCI '08, pages 111–119, Swinton, UK, UK, 2008. British Computer Society.
19. F. Stutzman and J. Kramer-Duffield. Friends only: examining a privacy-enhancing behavior in Facebook. CHI '10, pages 1553–1562, New York, NY, USA, 2010. ACM.
20. Y. Wang, G. Norcie, S. Komanduri, A. Acquisti, P. G. Leon, and L. F. Cranor. "I regretted the minute I pressed share": a qualitative study of regrets on Facebook. SOUPS '11, pages 10:1–10:16, New York, NY, USA, 2011. ACM.
21. J. Wiese, P. G. Kelley, L. F. Cranor, L. Dabbish, J. I. Hong, and J. Zimmerman. Are you close with me? are you nearby?: investigating social groups, closeness, and willingness to share. UbiComp '11, pages 197–206, New York, NY, USA, 2011. ACM.
22. P. Wisniewski, H. Lipford, and D. Wilson. Fighting for my space: coping mechanisms for SNS boundary regulation. CHI '12, pages 609–618, New York, NY, USA, 2012. ACM.