Opportunities and Challenges Around a Tool for Social and Public Web Activity Tracking

Amy X. Zhang
MIT CSAIL
Cambridge, MA
axz@mit.edu

Joshua Blum
MIT CSAIL
Cambridge, MA
joshblum@mit.edu

David R. Karger
MIT CSAIL
Cambridge, MA
karger@mit.edu

ABSTRACT
While the web contains many social websites, people are generally left in the dark about the activities of other people traversing the web as a whole. In this paper, we explore the potential benefits and privacy considerations around generating a real-time, publicly accessible stream of web activity where users can publish chosen parts of their web browsing data. Taking inspiration from social media systems, we describe individual benefits that can be unlocked by such sharing and that may incentivize users to publish aspects of their browsing. We ask whether and how these benefits outweigh potential costs in lost privacy. We conduct our study of public web activity sharing through scenario-based interviews and a field deployment of a tool for web activity sharing.

Author Keywords
web browsing; privacy; web tracking; sharing motivations; activity traces; self-presentation; web analytics; social media

ACM Classification Keywords
H.5.3. Group and Organization Interfaces: Web-based interaction; Computer-supported cooperative work

INTRODUCTION
The global record of people's browsing activity on the web is a treasure trove of valuable data. To see this, we need only consider the tremendous amount of effort and money that corporations put into tracking, collecting, analyzing, and selling this data, often without users' consent or even awareness [1, 29]. Arguably, this is one of the primary drivers of the Internet market. It is notable, therefore, how little of this valuable data is accessible to the end users who generate it.

In contrast to this involuntarily-shared and generally inaccessible data, social media tools like Facebook and Twitter help users voluntarily share details about their whereabouts, interests, emotions, and other aspects of their day-to-day life, for purposes such as building a public persona and interacting with friends. This does not imply, however, that users do not care about their privacy; rather, users are often very aware and can tailor their sharing depending on the context [20].

If users could also publish chosen parts of their web activity traces this way, then instead of only benefiting organizations, shared data could directly benefit its sharers. The information could provide a wealth of new social opportunities that users could leverage, much as with social media. For instance, extroverts may welcome the opportunity to "nanoblog," sharing even more details about their daily browsing without needing to construct tweets. Others may consume this information to discover interesting new content, or may value encountering and communicating with friends or strangers.

In this paper, we explore a variety of ways that sharing web activity information to friends or to the public can offer direct benefits to the individuals involved. We describe a prototype system, Eyebrowse, that uses a real-time global web log to let people "bump into" their friends on the web, discuss what they're seeing in real time, see where crowds are gathering and follow them to find new places, blaze trails though the web, and leave marks of their presence for others to see.

While appealing, these potential benefits also potentially incur a cost in lost privacy. Therefore, we devote a substantial portion of this paper to exploring people's privacy preferences. While there is widespread revulsion for the *involuntary surveillance* imposed by corporations and governments for their own benefit, we focus here on exploring what kinds of *voluntary sharing* people may wish to engage in to benefit themselves, their friends, or society. We do so first through a series of scenario-based inteviews that describe hypothetical sharing features that people might find appealing on the web. We then present results of a preliminary field deployment of the Eyebrowse system.

While there has been substantial work studying attitudes towards, methods for, and protections from the current adversarial practices of involuntary web tracking, we have found little work considering cost-benefit tradeoffs around letting end-users leverage the global web trace for their own benefit. We believe this work offers new insights into this opportunity. Furthermore, a model where users explicitly choose to share their data pushes back against the current assumption that the only way to collect browsing data is to take it surreptitiously.

RELATED WORK
Our work is informed by three lines of research: the first on the issues of privacy that have arisen due to online surveil-

CSCW '16, February 27–March 02, 2016, San Francisco, CA, USA
Copyright © 2016 ACM. ISBN 978-1-4503-3592-8/16/02 $15.00
DOI: http://dx.doi.org/10.1145/2818048.2819949

lance and ways to mitigate them, the second on the interplay of privacy and publicity in social media, and the third on the opportunities that lie in harnessing web activity.

Privacy, Surveillance, and the Web

In the domain of web browsing, researchers have shown that many Internet users have little idea about the extent to which and the nature of how they are tracked [29, 30], and when questioned, reject the status quo of online tracking [17]. Some researchers posit that it is not so much the restriction of access to personal information that users care about, but that information moves appropriately according to the context [25]. Studies have also shown that different aspects of web activity data have different privacy levels depending on the user and the content [1, 8]. This leads us to consider whether there are areas on the web that users consider public and are willing or even eager to share with others.

There is a great deal of research on designing for privacy in technical systems [14], such as ubiquitous computing systems [10], mobile apps [26], location-based services [28], and more. Work related to privacy and web browsing primarily focus on restricting tracking or alerting users [7], or on how to share browsing data, but only to researchers [21]. Our research expands on this work by considering how to design for sharing browsing activity in the cases where users may want to share, while still respecting users' privacy preferences.

Balance between Privacy and Publicity with Social Media

With the advent of social media, individuals now share large amounts of information about themselves online and also gather information about their peers, in a form of social surveillance [19]. This type of sharing allows users to gain social capital [4] and also relies on them to self-monitor their actions and balance their levels of publicity and privacy [20]. There are a host of privacy issues related to social media, including that users make decisions about sharing they regret [32], that interfaces aren't as nuanced in their controls as necessary [20], and that a power imbalance exists between users and companies that collect this data [2]. Yet a voluntary model of self-disclosure can counter the hierarchical models of power that come with traditional surveillance and potentially put more control in the hands of users.

Many social media users care more about how and what information gets shared with peers than with corporations [20], and thus privacy controls must be developed with this in mind. From studying social media sites like Facebook, it is clear that the design of privacy controls cannot be so cumbersome that users ignore them entirely nor so one-dimensional as to prevent sufficient amount of personalization [11]. Instead privacy interventions succeed when they take into account the inherent design conflicts [34] and existing social dynamics of how and why users share in the first place [32].

Using and Sharing Web Activity

Previous attempts to release large repositories of web activity all involved anonymized logs. One infamous attempt was the AOL Query Log Dataset, which showed that anonymity is not enough to hide the identities of users. Other projects collected web activity for the purpose of contributing to research. Challenges there involved convincing users to contribute if they were volunteering [9], inability to scale if they were being paid [21], and complications around preserving anonymity [31]. Our research aims to tackle motivation by introducing social incentives that would benefit users if their data were released for peer and public consumption.

When it comes to the social benefits of sharing activity traces, researchers have demonstrated the advantages of having a transparent activity stream within communities such as Wikipedia [13] and Github [3], or for social recommendations [24]. Also, a long line of research has studied how to make online spaces mimic those of real-life ones in order to promote qualities like serendipity [12]. We use this prior work to consider what social benefits to study and build into our web activity sharing tool. It also informs our choice of participants and the use cases we envision.

Finally, researchers have studied and built tools that involve sharing web activity traces for more niche purposes, such as visualizing actions within a page [15], adding citations within a technical post [6], or generating content recommendations [16]. Some web browsing applications that are more social in nature also exist, such as Flock, Torch, or Beamrise, though none have been studied academically. However, most of these browsers focus on interfacing with different social media systems as opposed to harnessing browsing information, and none have considered public sharing. Our research builds upon this prior work by considering a general-purpose use case for publicly sharing browsing data.

METHOD

We conducted a qualitative study with scenario-based interviews followed by a field study of a web activity sharing tool. The main goal of the interviews was to understand users' perceptions about sharing aspects of their browsing activity and to assist in the design of said tool. The field study then allowed users to share aspects of their activity publicly and participate socially while browsing.

Participants

Because we were interested in the social aspects of web browsing, we chose to recruit small groups of friends or colleagues to participate in both the interviews and field study. The friend groups represent general Internet users that might be interested in sharing primarily with other friends. We also recruited professionals in the areas of news and technology, where users may additionally be interested in the benefits of sharing with the public or with work colleagues. Our choice of these fields draws from popular systems such as Twitter or Github, where many users publicly share the articles they read or the work they conduct, respectively. These kinds of users may also be critical early adopters that draw in more everyday users interested in their activity.

- FRIENDS is a group of 5 close friends and recent graduates of a mid-sized U.S. university that currently live in different areas of the world. They keep in touch via email to share and discuss news and issues. Their occupations range from non-profit manager, to consultant, to business analyst.

	Interviewed	Average Age	Ratio of Females
FRIENDS	5	22.6	2/5
NEWS	7	35.4	4/7
TECH	4	28	0/4

Table 1. Number of people participating in the interviews and basic demographics.

They were chosen in order to understand how friend groups and people that use the Internet mostly for personal entertainment might feel about sharing browsing information.

- FRIENDS-2 is a separate friend group of 11 undergraduate and graduate exchange students at a university in Israel. This group was not interviewed but participated in the field deployment. They were chosen as yet another friend group to study but that was co-located.
- NEWS is a group of 7 mid-career acquaintances related to the news industry, primarily journalists. They are an international group, with members coming from 5 different countries. During the study, they were all co-located in the U.S. and enrolled in a seminar together. This group was chosen as a set of people that are heavier news readers, interact with the news on a professional basis, and maintain active public personas.
- TECH is a group of 4 graduate students working together in a computer science lab at an Israeli university. This group was chosen as a set of technologically savvy individuals that have an understanding of online tracking, contribute to open source, and conduct most of their work online.

Interviews

The interviews were conducted with participants in 3 out of the 4 groups during April and May of 2015. Demographic information about each group is listed in Table 1. Each interview took between 20 and 80 minutes, and interviewees were compensated $10 for their time. In the interviews, we delved into how users felt about a hypothetical tool that could share aspects of their web activity with friends and with the public. We were interested in the utility, interest, and privacy concerns that a specific feature would generate for a user, both as a consumer and a producer. The format of the interviews took inspiration from prior research [18] that used scenarios and hypotheticals to identify users' privacy boundaries. The specific scenarios will be described in more detail below. We chose to use interviews as opposed to surveys as many of the hypotheticals necessitated further explanation by the interviewee, and we were interested in users' reasoning. The interviews then concluded with a set of more open-ended questions about interviewees' opinions on surveillance on the web and sharing web activity with third parties and peers.

The interviews were conducted in person by the 1st and 2nd authors. They were recorded then transcribed and coded by the 1st and 2nd authors using standard qualitative coding techniques [22], which led to 74 codes after multiple iterations. During the course of the interviews and coding process, the codes as well as emerging themes were discussed among members of the research team. Because of the small number of interview participants, our interview findings should be regarded as indicative.

	Participants	Average Age	Ratio of Females
FRIENDS	4	22.5	2/4
FRIENDS-2	11	22.6	5/11
NEWS	5	35.2	3/5
TECH	4	28	0/4

Table 2. Number of people participating in the field study and basic demographics.

Field Study

Besides the scenario-based interviews, we also wanted to observe what users would do when it came to real-life situations. For this reason, we designed and built Eyebrowse, an open source social media tool for viewing people's shared web activity and a companion Chrome extension that allows users to share certain aspects of their browsing and that provides some social experiences. We describe the tool in more detail further in the paper. We built the tool over the course of many months and allowed people on the web to freely use it as we iterated. Though we did not release it widely, we publicized it at different times to various groups, including to a number of the participants mentioned earlier. Thus several participants were trying Eyebrowse voluntarily for different extended periods of time. While these users helped provide information about what they were willing to share, the lack of peers using the system meant that social features would be less useful or interesting. Thus, at the conclusion of the interviews, we also conducted a formal week-long field study in order to ensure greater adoption within peer groups.

We invited all interview subjects to participate in the field study and compensated the ones who agreed to participate $10. In Table 2, we report their demographics. We asked all participants to register for the Eyebrowse website and download the extension. Participants were also asked to visit the website at least once a day to see what other people shared. They were not required to share anything, nor were they required to participate in any social activities. They were told that all web activity they chose to share would be released on the website and viewable publicly. For the purposes of this study, we separately obtained consent to our collecting additional statistics about these users' Eyebrowse activities, such as links they clicked on from the Eyebrowse website. None of this additional information was visible to any of the participants. At the conclusion of the field study, participants were given an exit survey where they were asked about their experience with the tool. While we report numbers on the usage of the system and some raw survey results, these do not hold statistical significance due to the small sample size. Further discussion of the limitations of our findings can be found in the Limitations section.

PRIVACY AND WEB ACTIVITY SHARING

We first turn towards the potential privacy concerns around sharing browsing activity and how to design a system so as to alleviate these concerns. Following this section, we consider potential social features that could incentivize users to share aspects of their browsing data. Finally, we report the results of the field study, where we analyze what people actually chose to share and their reactions to using the tool.

Web Surveillance Today: Perception versus Reality

We begin by exploring the perceptions that interviewees had about sharing their data with third parties. We were interested in understanding how this might motivate a new method of data collection as well as inform our design of a web activity sharing tool. We found that interviewees had varying levels of understanding of how they were tracked while browsing, as well as perceptions that did not always align with reality. As this topic is well covered in prior research [17, 29], we only briefly mention the main themes that went on to be relevant in our consideration and design of Eyebrowse.

Ownership of Browsing Data. First, most interviewees (11/16) expressed a strong sense of ownership over their browsing data, including a right to know who was collecting their data, as well as when, where, and for what purpose:

> *Companies that...track what you do and then sell it...[is] analogous to theft because they are taking away from you information that you produced yourself...without you knowing it to make money...that's where the problem is, the not knowing. You feel a bit like a fool you know?* –NEWS

Ownership also suggests power to give the data away. Some interviewees considered how if they had the ability to share their data with others, they could then decide who they wanted to benefit from their data:

> *If I'm getting the most relevant results from Google because they have the most information about me? OK. If Bing had the same amount of information about my behavior? Could they actually be doing a better job and might they offer me more controls? Being able to transfer that from one party to another may be kind of cool.* –NEWS

Differing Levels of Privacy. All interviewees expressed that browsing was often a private activity. Interviewees believed that some aspects of browsing were more sensitive and some aspects less, with many of these considerations personal in nature, echoing prior work [1, 8]. When asked, interviewees cited examples of areas on the web where they preferred no one knew they visited as well as areas where they would be unhappy with strangers knowing, including medical websites, dating, shopping, politics, banking, and more.

Expectation of a Trade. Most interviewees (14/16) had an expectation that when releasing their browsing data to companies, they would receive something in return. These interviewees treated the choice to use a service as implicit consent to be tracked while on that service [25]. It was because of this expectation of a trade that many interviewees were affronted by companies that tracked them without providing services in return. Speaking about Facebook, one interviewee said:

> *Why should they be able to know what sites I'm going on if I didn't go there through them? It's not through their merit that I ended up on this site about blueberry pie. They have no claim to that information, as far as I'm concerned.* –FRIENDS

In several cases, interviewees saw the trade as beneficial to them only so long as the data was used to improve their experience. Unfortunately, some interviewees expressed unhappiness with some of the services their data was used to create, such as personalized advertising or incorrect or limiting recommendations, with little recourse to rectify the situation.

Summary. From the interviews, there was a stark difference between users' perceptions and desires and the status quo of web surveillance, which is arguably unethical [25, 29]. Though browsing data provides immense value to corporations, the people providing the data are not receiving much of the benefit. They cannot choose to give their data to services they wish to support nor do they have a say in how the services behave or what the services can and cannot track. It is also apparent that far more data is being gathered than would be if users were in charge.

Privacy Considerations when Sharing Web Activity

Based on these findings, we consider how a system for web activity sharing with peers and the public would need to act in order to be successful. In many ways, sharing information to the public, including to friends or peers, is more fraught than sharing to corporations or governments [20]:

> *I don't really care as much about Big Brother or whatever knowing, because I figure the internet companies already know everything. It's more that I don't want to have an embarrassing instance with friends or family.* –FRIENDS

Many of the concerns users had were similar to ones voiced earlier about sharing with organizations, but were now exacerbated due to the possibility that friends or acquaintances would also know. We mention the most salient ones below.

Misrepresentation. Several interviewees (4/16) were concerned with being misrepresented by others based on aspects of their browsing behavior. Echoing prior work on context collapse in social media [20], some interviewees mentioned innocuous data that might be misinterpreted by loved ones or friends when not understood in context.

Inadvertently Revealing Information. Other interviewees (6/16) were concerned about revealing information about themselves or about others. For instance, three journalists were concerned about how they might leak information about their work that they wanted to keep secret or about trusted sources. Other people mentioned scenarios such as keeping information away from bosses that were nosey or having a secret love affair. Some people discussed how they were worried they or others might interpret the settings incorrectly and accidentally share things they meant to keep private.

Concern About Judgement. Some interviewees (4/16) were worried about being criticized for the pages that they visited, such as a journalist concerned her audience might disapprove of her research methods. One TECH interviewee mentioned his worry that people might judge him for looking up answers to simple questions on StackOverflow. Other interviewees did not want to broadcast things such as how long they spend on Facebook or how they read articles such as gossip columns.

Summary. From the interviewees' overall responses, it is clear a tool for web activity sharing with friends and with the public needs effective and transparent privacy controls, even if users are already giving away their data wholesale to unknown corporations. Given users' expectation of a trade for data, there should also be obvious benefits to users in return for the use of their data.

Ways to Control What to Share

Given the concerns voiced by interviewees, we turn to considering how users might want to control what they share.

Domain-Level Control. As we found earlier, interviewees had differing acceptable levels of privacy for different websites that were often content-based, echoing prior work [1][8]. This led us to consider whether an opt-in whitelisting approach at the domain level might work well. Many interviewees (9/16) were receptive to the idea of sharing on a domain-by-domain basis. Several interviewees felt that domains related to news would be for the most part uncontroversial and perhaps also interesting to share. In addition, TECH users felt it would be fine to share programming-related websites. One interviewee explained why he would whitelist news websites and share with his friends or even the public:

> *...I don't think there any domains for which I would be sometimes okay and sometimes not okay. If it were sometimes not okay, I would just not whitelist the domain...I can't think of a situation in which I would read an article and not want my friends to know. And then, as a separate question, not want the public to know.* –FRIENDS

Time-Based Toggling. Some interviewees wanted the ability to be able to hide the time or to turn sharing on and off during certain parts of the day, such as during work hours or while they were doing certain tasks:

> *No, I hate to lie but maybe to hide the time would be nice because you did not answer the phone and you don't want them to know that you were awake.* –TECH

Another interviewee, a journalist, wanted to be able to turn all sharing on during parts of the day when working instead of following a more cumbersome domain whitelisting approach.

Sharing Anonymously or in Aggregate. Most interviewees (14/16) felt uncomfortable about sharing aspects of their data publicly. Some interviewees thought they would feel more comfortable if it were somehow anonymous or only appeared in aggregate, though this would make some social benefits unattainable. Given the possibility of de-anonymization, as happened with the AOL Query Log Dataset, a anonymization feature might be unwise to offer, even if users want it. There were other ways that interviewees wanted to be able to control their sharing. Some interviewees felt that they would be more willing to whitelist certain domains if they could obfuscate certain details, such as the specific page they were on within the domain, the amount of time spent on a page, or the time of day the visit occurred. However, these controls would necessarily add complexity to the decision to share and would need to be balanced with clear design.

SOCIAL BENEFITS OF WEB ACTIVITY SHARING

As expressed earlier and in prior work [26], people wanted incentives for sharing data. In this section, we turn to existing social media features to consider what potential benefits could be unlocked with shared web activity data. With each feature, we gauged impressions using scenarios posed to interviewees. A scenario began with something that the interviewee was familiar with, and we provided screenshots of existing systems as visual cues. Then for each scenario, we designed a set of hypotheticals to present to interviewees that varied different aspects of the system and asked for their reaction. One set of hypotheticals broadened the existing system in question so that it would exist first on other pages on the web and then on every page on the web. Another set of hypotheticals asked interviewees to consider sharing information first only with friends and then to the public. Some hypotheticals included screenshots of aspects of the Eyebrowse tool we built to give users visual examples.

Visibility on a Webpage. We asked interviewees to consider websites that showed who else was online at the same time, such as on Facebook or Google Hangouts. When asked to then consider having the same visibility feature on specific webpages on the Internet or on any webpage on the Internet, many interviewees (8/16) responded positively:

> *Maybe I'll realize that this random guy in my entryway is always on Epicurious looking up recipes just like me. That's cool. Then, if you're on specific pages and it's showing you who is on that actual page...then you could know that you can have a conversation with someone about it later.* –FRIENDS

Most interviewees wanted the visibility feature to only appear on certain websites, much as they preferred to share their information publicly only about certain domains. Some interviewees preferred to only be visible to friends instead of the public. Other interviewees didn't mind being visible publicly but were only interested in learning about people they knew. Interviewees were also concerned about getting distracted by others and wanted the ability to go incognito when doing work or focusing on a private task.

Chatting and Commenting on a Webpage. Many interviewees (9/16) saw benefit in being able to chat in real-time with other people on the same webpage or leave comments on pages anywhere on the web. One interviewee mentioned trying to learn more about data hosted on a page online, while another saw utility in real-time chat on StackOverflow:

> *The comments that were available on that page were from other people asking for updated data... They were relatively recent. I knew that there was a lot of interest in it. It would have been really great to be able to see who else was actively on the page...and maybe have a, "Hey are you looking at this too? Do you know anywhere else?" Being able to use them as a piece of support for the Internet.* –NEWS

> *StackOverflow...people will post their comments, but it will be much easier to chat with other people that are currently solving the same problem, and to discuss things.* –TECH

Interviewees also preferred having the ability to block people or to leave chats when they were done or were busy elsewhere. More members of NEWS and TECH were interested in conversing with strangers and had a greater history of participating in public comment spaces or on public social media. Members of FRIENDS were more interested in chatting if it were limited to a small group of friends. Some were not against the idea of chatting with strangers but saw little benefit. Many interviewees expressed a desire for the conversations to be moderated or organized in some fashion, citing the chaotic nature of many public chat rooms and forums.

Ambient Awareness. For this scenario, interviewees were asked to think of more auto-shared and ambient activity streams such as sometimes present on the Facebook home page or on Spotify. They were then asked to consider such a stream but comprised of real-time visits to webpages that their friends or colleagues were currently on. Many interviewees (6/16) were interested in the serendipitous benefits that such a stream might provide:

> *I feel like our web patterns...are less likely to be only useful to us... If the person's attending an arbitrary Facebook event, odds are I'm probably not interested and/or invited. But if they're visiting a certain webpage, I think the odds that it might be interesting to me, especially if I read the title and judge whether I want to click, are higher.* –FRIENDS

Interviewees mostly wanted to keep such a stream limited to information from their friends. While some could see it as potentially interesting if it were integrated into their browser, others were concerned that it would be too distracting and wanted the ability to turn it on or off or not have it showing when they were doing work or composing an email. Others wanted the ability to filter the information or have the system only show the most interesting visits and not mundane ones.

Reflection and Self-Improvement. Several interviewees (7/16) were interested in the idea of monitoring their web browsing. While personal applications for online monitoring such as RescueTime exist, a few interviewees were interested in the social aspect of monitoring, such as sharing summaries or streams of their browsing habits as a way to be more accountable, not unlike many of the fitness and food dieting social applications people use. Similarly to those systems [5], interviewees were concerned about coming off as a braggart and about annoying or overwhelming their friends. Several interviewees mentioned that they felt that sharing would cause them to be more mindful of how they browse the web:

> *...I would make more of an effort to, say, go to diverse news sources... I respect my friends for doing that so I would want them to feel like I do that, too... Yeah, I think the potential for...approval from friends is potentially a good incentive to change behavior...* –FRIENDS

Other interviewees felt that sharing what articles they had read with others would cause them to read more carefully. Some interviewees that were interested in monitoring their browsing felt that there were parts of the web that they wanted to monitor but didn't want to share in detail, such as all the pages they visit on Facebook. Here, a strategy for sharing the domain only and not specific pages might be more fitting.

Self-Presentation. One important reason people participate on social media is self-presentation [23]. The act of self-presentation could also extend to the web browsing realm. Several interviewees (7/16) thought this feature was interesting. One interviewee expressed how with some sites she even wanted people to know that she read them:

> *I think that there are some websites that are very normal, and almost good that you read them, right?...I like people to know that I read the news or that I try to keep up with politics and energy stuff...* –FRIENDS

Another benefit that was more of a draw for members of NEWS was self-presentation for a public audience. One interviewee mentioned a goal of being more transparent about her work process while reporting on the news:

> *It scares me, but I'm interested. I think that's what appeals to me about...how can I make my recording process more transparent and invite people to give me criticism about how I research... It probably will provide more transparency for how something becomes news which I think would be great.* –NEWS

Several TECH interviewees were also interested in being able to broadcast their web activity streams for instance when working on an open-source software project. Prior research has also shown that showing one's browsing behavior can cause audiences to trust the advice of the sharer more [24].

Content Recommendation. Finally, we ask interviewees to consider how browsing information collected from friends or the public could recommend interesting content for them to read, similarly to many aggregators. Almost all interviewees (13/16) found this feature interesting. Some interviewees thought this information would be interesting in different ways than recommendations from other social media because there would be less curation:

> *I think there are certain filters to what people post on Facebook... I often read about pregnancy or depression ...not things that I want to be throwing over Facebook... I think it would be interesting to see what people are reading in addition to what people are comfortable posting.* –FRIENDS

Some interviewees preferred recommendations from friends, while others were also interested in recommendations from the public:

> *My friends are smarter than the vast majority of New York Times readers...My friends tend to be thoughtful people and interested in...the same issues I'm interested in.* –NEWS

> *I would really love it, definitely in sort of gawking way...What do people know about? What do people care about? From a curious news-producing perspective that would be incredibly interesting.* –NEWS

Interviewees were also interested in how the analytics provided by a web activity dataset could be useful:

> *I have a sense that the New York Times and the Wall Street Journal are really legit. If I'm on a random website, I don't have a good sense of its legit-ness. So it would be nice if you could...see how many readers this site gets in a day...[and] sense whether it's an important news source.* –FRIENDS

Many interviewees brought up the need for more sophisticated filtering capabilities to be able to remove the types of content that they didn't find interesting. Some interviewees also felt overwhelmed with the amount of news and other content they already read and felt that something that wasn't carefully curated by people would exacerbate that issue.

FIELD STUDY RESULTS

We now turn to our field study, which allowed us to see how people would behave in a more real-world setting while browsing as usual, as opposed to hypothetically. We first describe the web activity sharing tool that we built.

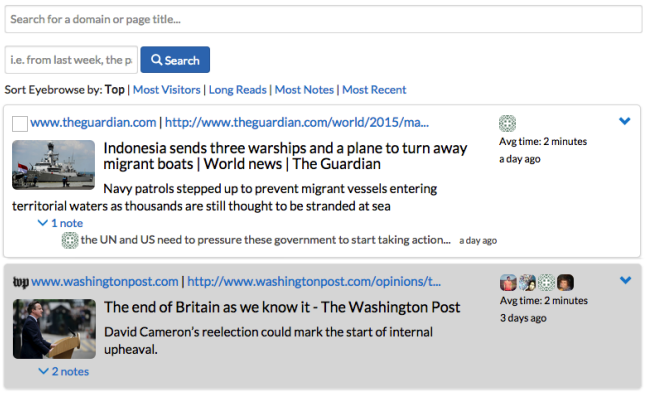

Figure 1. Screenshot of the live stream of data presented on one of the the Eyebrowse feeds.

Figure 2. Prompt showing recent visitors and comments on the webpage the user is viewing.

The Eyebrowse System

The Eyebrowse system[1] consists of a web server coupled with a Chrome extension that allows a user to build up a whitelist of domains from which they permit their activity to be automatically shared. At the start, a user's whitelist is empty and no browsing activity is tracked by default. When visiting a new domain, the user is occasionally asked via a small popup on the corner of their window if they wish to whitelist that domain. If they click no, the domain is stored within a separate "blacklist," or list of domains the user does not want to share, and they no longer receive prompts about that domain. If they click yes, the domain is whitelisted, and information about each visit to a page within that domain, such as the page title and overall time spent, is automatically sent to the Eyebrowse server. Besides whitelisting domains, Eyebrowse also lets users publish one-off "check-ins" to pages on domains they do not wish to whitelist. Users can temporarily disable all automatic sharing via a toggle link in the extension.

At the server, shared browsing data is collected into several real-time and aggregate streams. Each user has a public profile page with their shared webpage visits. On their home page, they can see a feed of the pages visited by people they follow called "Following", as well as a "Firehose" feed of all web activity sent to Eyebrowse. An example of such a feed is in Figure 1, where already visited pages are grayed out, and comments made on a page can be viewed. We use a simple algorithm to aggregate and rank page visits based on factors including time decay, number of people that visited a page, average time spent on a page, and number of comments on a page. Users can mute specific key phrases or subdomains from their Following or Firehose feeds and delete items from their own visits retroactively (though someone may see the item before its deletion). A real-time version of this feed is also available that sorts in reverse-chronological order.

[1] http://eyebrowse.csail.mit.edu

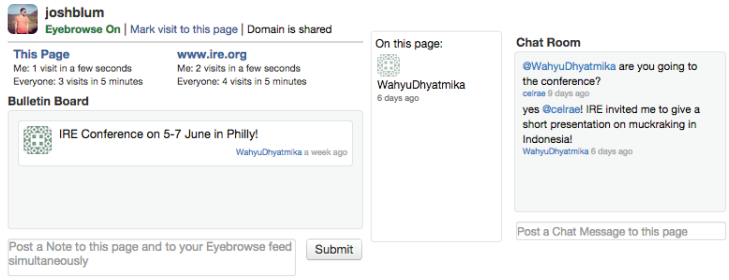

Figure 3. Chrome extension popup, showing an example of a comment and a chat conversation on a page.

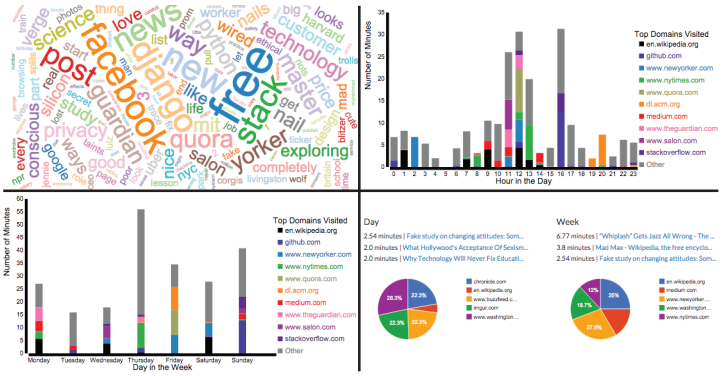

Figure 4. Personal informatics visualizations available about a user.

We enable social interactions while browsing the web using the Chrome extension. As shown in Figure 2, when visiting a webpage that another Eyebrowse followee recently visited, a small prompt appears in the corner of the page showing the followees that have been to that page as well as the most recent comment left on the page. Followees currently on the page are highlighted, allowing users to "bump into" followees while browsing. Clicking on the extension icon opens a small pop-up window, as shown in Figure 3, where users can see all comments made on the page as well as a dedicated public chat board for that page. From this window, users can post comments that then also appear on the Eyebrowse website feeds, or participate in chat, which is only accessible on that page. While posting, users can tag followees who will then get an email notification that they were tagged. We also provide some visualizations for users to see how they browse, including a word cloud of terms appearing in page titles, time-of-day and day-of-week breakdowns, and top webpages and domains visited, as shown in Figure 4.

Activity on Eyebrowse

We show the use of the Eyebrowse tool by participants during the field study in Table 3. We noticed in the usage logs that overall, sharing of page visits was quite variable, even though users' whitelists did not change very often. From asking users, we found this was due to issues like using other browsers, computers, or mobile devices, or not browsing the web on certain days. Since on these occasions, users were not actually using Eyebrowse, we normalize several measures by days of *active usage* of Eyebrowse, or days in which at least one item was shared.

	FRIENDS	FRIENDS-2	NEWS	TECH
avg # followees†	3.8	5.5	4.0	4.5
avg # followers†	4.0	5.2	3.0	3.2
avg # domains in whitelist †	19.2	10.2	38.8	12.8
avg # times pop-up opened*	37.6	6.1	2.7	2.6
avg # items shared*	11.1	27.8	13.3	11.9
avg # visits to web feed*	5.9	19.7	7.5	4.3
avg # clicks from web feed*	0.2	0.6	0.4	0.1

Table 3. Average counts of user actions from field study. †Average per user. *Average per user per day of active usage.

When they used the system, users were overall engaged, including accessing feeds on the Eyebrowse website and clicking on the extension 7.5 and 7.6 times per day on average, respectively. Users visited their own profile the most often (45% of all visits to Eyebrowse), and visited other people's profiles much less often (17%). They also chose to view the Firehose feed (26% of all visits) more often than they viewed their Following feed (12%). We also measured how often users clicked on a link from the Following or Firehose feed on the Eyebrowse website. Overall, we found most users chose to click on news articles. Out of 55 total clicks, 27% were news related articles or websites. Some use of Eyebrowse was playful when it came to sharing information; for instance, one TECH participant shared Google Search result pages of silly queries about another user that he knew would see the visit on Eyebrowse. This demonstrates the performative aspects of sharing information.

While trying out Eyebrowse, many users chose to share, either automatically through whitelisting or actively "checking in," their page visits on the web. We can see in Figure 5 that people generally shared anywhere from 0 to 30 items a day. There were a few outliers in FRIENDS-2 that shared upwards of 60 to 80 items during the time they tried out Eyebrowse, though other members of FRIENDS-2 shared much less. Within the other groups as well, one can see the spread of number of items shared, demonstrating the differences in how much people are willing to share. Figure 6 builds upon this, showing how the FRIENDS group was willing to whitelist many domains right away, while members of NEWS were more cautious and slowly built their whitelist over time while browsing the web.

In Figure 7, we show whitelist and blacklist decisions broken down by category among all participants using Eyebrowse. As users were prompted about whitelisting a domain only occasionally while browsing, these decisions do not account for all domains visited during the field study. Categories were extracted from an open API provided by alexa.com. Many domains return no category information from the API, including 150 whitelisted domains and 100 blacklisted domains and are excluded from the graph. As reflected in the interviews, many participants chose to whitelist news websites, reference sites such as Wikipedia, and technology sites. It is interesting to note that many categories can be considered both public and private. For example, both the "Society" and "World" categories contained a mix of whitelisted and blacklisted items. Conversely, some domains were considered public by the majority of users, such as "News" or "Regional." Overall, there

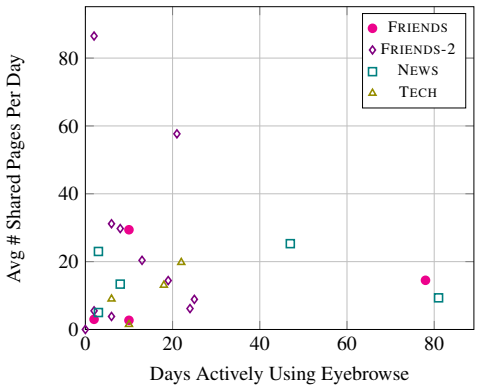

Figure 5. Scatterplot of the average number of pages shared by a user over the course of actively using the Eyebrowse system.

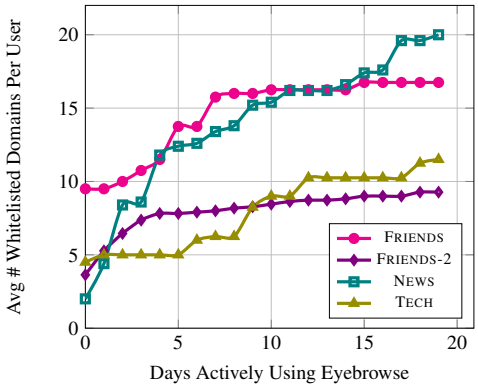

Figure 6. Timeseries of the average number of whitelisted domains per user within each group over first 20 days of actively using the Eyebrowse system.

was little overlap in domains whitelisted, with 312 domains whitelisted only by 1 person and 14 domains whitelisted by two or more people. This speaks to the diversity of domains online and the very different browsing patterns that people have, suggesting that finding out about new domains could be an interesting feature in addition to new articles or webpages.

When it came to usage of social features, the NEWS participants were the heaviest users of the chat and comment functions, posting 11 out of 12 total chat messages and 19 out of 30 total comments over the course of the 1-week field study. This may be because they all had public Twitter accounts and were accustomed to maintaining public personas. Many participants used the commenting capabilities to alert their friends of pages, to post a short reaction to the public, or to leave a note to themselves, somewhat similarly to Twitter. There were several cases of serendipitously coming across information about other members. For instance, one user in NEWS who was from San Francisco noticed that another user had been visiting the webpages of restaurants in San Francisco and asked on chat if she would be making a trip there. The other user replied to say yes and to ask for recommendations. As another example, two people discovered they'd be attending the same conference due to visit traces left on the conference website.

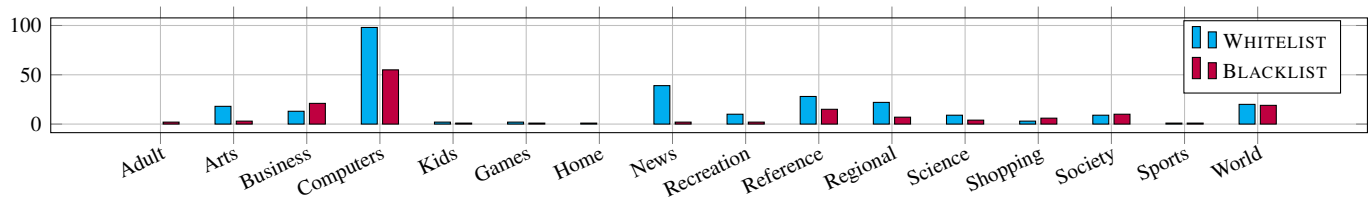

Figure 7. The number of items whitelisted and blacklisted by all participants within each domain category.

Post-Study Survey

We sent out a survey of mostly open-ended free-response questions after the completion of the field study to all participants of Eyebrowse. 23 (60% male, average age of 28.4) out of 24 total participants filled out the survey.

When we asked about people's experiences with whitelisting domains and deciding what to share versus not, some respondents expressed that they were surprised by what they actually felt or did in the moment. One user, who in the interview had been confident about what he was comfortable sharing, realized that there were some domains where he was okay sharing the domain but not the specific pages he was on. He later went back and deleted visits to those domains from his feed. Another user realized that Wikipedia was actually something that he did not want to share: *"I realized I don't really want to be revealing my ignorance plus it can be embarrassing to share the random stuff you browse on a Wikipedia dive."* Some people's concerns were more related to self-presentation to their peers instead of about privacy. One respondent expressed trying to curate their feed similarly to how they curated their Twitter feed: *"I'm politically waayyyyyy left and decided I didn't want everybody to be able to see some of the things that would indicate that."*

When it came to features that were useful or interesting versus not, many respondents enjoyed being able to see their own content (16/23) as well as other people's content (16/23). Ten respondents stated that they found interesting content on their feeds, including things about their friends that they didn't realize before, for instance, that a friend had a Github profile or a personal blog. When it came to their own content, 12 respondents indicated aspects about their own browsing that they found interesting or surprising. Several mentioned habits that they were interested in changing as a result, with one person stating they did a lot of multitasking and other users realizing they spent a lot of time on Facebook or Twitter. One NEWS user said: *"I spent a lot of time on my own newsroom website, and on mainstream news sites. I realize I should expand my horizon and start looking at smaller and more diverse sites."* When it came to the social features, users were split. Eight people stated social features were the most useful features: *"I liked the social bits way more than I expected. I was wishing I had some specific project to work on where I could ask others for their input."* On the other hand, five people stated they were the least useful feature: *"I didn't have a lot of interaction with other users."*

People also had different things to say about their overall experience. Some respondents saw the potential in Eyebrowse and were interested in continuing to participate voluntarily:

"I can chat about stuff I found online with people I knew have been reading the same stuff...If Eyebrowse's goal is creating a public space online, I believe the app has met its objective. Well done!"

"I thought using Eyebrowse is really great, I think the more people to use it the more it can offer."

"I think it's a good idea that can be successful, because it's implementing the social component with the statistical component in a fun way."

Other respondents had criticisms of the experience:

"Maybe if more of my friends would use it than it will be interesting but currently the content which is presented to me is just not that interesting."

"It was pretty boring, most of the people I was following only marked pages in StackOverflow or Github which wasn't very interesting."

One user expressed greater concern about privacy after trying the system: *"I got more aware about what pages I visit on the web and how my browsing history...might be perceived by others. I am not really comfortable with sharing my browsing history, not even for a few selected domains."* From the interviews and field study results, it was clear that there were some people for whom sharing any web activity data was out of the question. In the end, 18 of the 23 respondents indicated that they would continue to use Eyebrowse.

DISCUSSION

From our field study results, it appears that many users enjoyed or were willing to share aspects of their browsing data in order to learn more about their friends and themselves and to participate socially. Most users readily understood and participated in using whitelisting to segment their private and public spaces on the web. Many users also expressed that they became more mindful of their browsing behavior while using Eyebrowse and more aware of the information their browsing data communicated.

Providing Social Incentives

As several users commented, it may be that benefits for this system may not be fully realized until many more people are using the tool. With more peer users, social interactions such as "bumping into" someone while browsing the web would happen more often, and better web-scale analytics and personalized recommendations could be provided. Perhaps benefits may not be clear until a user has followees in the tens or hundreds as is often the case on Facebook or Twitter [27].

Issues such as impression management to different social circles might become more salient at those scales as well.

While a larger field study might provide more insight into the use of Eyebrowse's social features, we chose to start with a smaller field study combined with interviews in order to get a more in-depth understanding of users' reactions to sharing browsing data. Our findings from this study serve to indicate how we might better design personal and social incentives and privacy controls, areas that Eyebrowse needs to get right in order to have a successful wider rollout or public launch.

Even if features are designed well, bootstrapping a social system can still be difficult as well as risky. It may be beneficial to focus on slowly growing the user base by improving the individual benefits of Eyebrowse, such as self-monitoring and reflection, before social ones in order to attract initial contributors. It also might be useful to develop features that piggyback onto existing systems, such as allowing easier integration with Twitter or Facebook or providing widgets to integrate into websites.

Tensions Between Different Goals

As we saw in both interview and field study results, there was a tension between sharing information publicly and accruing some of the social benefits that we introduced. For instance, users interested in monitoring their own information diet and browsing habits might want a more comprehensive set of aggregate statistics than what they are willing to share at the webpage level. Interviewees also mentioned wanting to be able to conduct some social activities only with a small group of friends or work colleagues instead of with the public. Another tension that arose was between the desire for self-presentation or fear of judgement and the interest in providing more data for transparency purposes, participation in discussion, or other reasons.

One approach to alleviate some of these issues would be to allow users to designate some visits as private to themselves or private to a set of users. A concern with simply allowing users to form private groups is that users might then default to not sharing anything publicly. Here we can also look to social media systems that still have considerable public activity even with the ability to make some posts or profiles private. For instance, a system might support or emphasize the more public-facing aspects in order to encourage public sharing. Another issue with this approach would be that the system would now be in possession of private and potentially sensitive data, making security more of an issue.

A different approach would be to allow users to designate certain visits as anonymous. This should be implemented with care since web activity can sometimes be de-anonymized, such as happened with the AOL Query Log Dataset. One way might be to associate each anonymous visit to a separate identifier, so that no web traces were stored; however, this reduces the usefulness of the data collected. Other ways that reduce the likelihood of de-anonymization but still provide insight into browsing paths would be to only group page visits under a single identifier within a certain time frame, within a domain, or that arise from clicking links only.

Opportunities for Public Benefit

As stated earlier, some of the social benefits we mentioned could accrue even in a social non-public model, where friends and acquaintances share privately with each other. Indeed, some interviewees stated that they would prefer this. However, some benefits require a public dataset, such as seeing what a general population is doing or building a public persona. An additional benefit that would come with publicly shared data is the enabling of more research and development of new tools in the public domain. As one interviewee stated:

> *If you ask me, I would rather prefer a thousand, a billion times that researchers for an academic and humanistic and knowledge purpose are the ones that...can map that information and store it for humanity, than a private company.* – News

Currently, few publicly accessible repositories of web activity data exist due to concerns related to the de-anonymization of anonymous logs. But if users chose to share their web activity publicly, research and development could flourish. For our tool, we chose to explore the public sharing of web activity due to these benefits and because it was the most extreme in terms of getting away from the inaccessibility of current browsing data.

Along these lines, we included with the Eyebrowse system a publicly accessible API to grab, filter, and query different parts of the existing data. One question given this approach then, is whether the data shared is of any usefulness. The data provided by this form of tracking will always be inferior to that collected by indiscriminate tracking because it is not comprehensive for an individual and also is not used by everyone. The benefit of data gathered in this way is that it is collected and released with the knowledge and consent of the user. Concerns over data privacy or identity leakage are therefore reduced. However, we still need to take care with data that is public, even if it is shared voluntarily by users, as it might still not be ethical to analyze or release more widely publicly accessible data [2]. This is due to the context of when the user chose to share the data, a context that may change drastically once the audience changes. This issue may be of particular concern when it comes to web browsing data, even data that users have consented to share publicly. It may be exacerbated if users are unaware of what hidden attributes of themselves can be gleaned from their publicly-released browsing data or the data in combination with data from other sources. Our model of consent does not account for this, and the solution may not be a purely technical one.

At the least, this requires that people who may analyze browsing data show consideration when publishing aspects of that data to a wider audience. It may be possible to consider limiting API access to the data to people that specify how they will use the data, or indicate that they will abide by rules such as using the data only in ways that benefit the producers of that data. Therefore, while the data would still be publicly available, the system emphasizes accountability and appropriate use [33], similarly to the Creative Commons system. Additionally, while this model of consent helps with some concerns we discussed in the paper, it does not offer a complete solution to others, for instance the issue of context col-

lapse and fear of misrepresentation. Here too, additional privacy features such as the ability to share in private or to share anonymously might diminish the issues, though they may not completely erase them.

LIMITATIONS AND FUTURE WORK

Our study involved only a small number of people, both in the interviews and in the field study. Thus, the findings we present are not meant to be conclusive but instead provide insight into how users might think about and use the system as well as suggestions for ways to improve the system. Because we partially targeted people from areas such as news and technology, we may have introduced additional biases in our results. As mentioned earlier, it was clear from our results that Eyebrowse in its current form may be more interesting or useful to some people and not others. In order to understand if web activity sharing is feasible as a strategy for collecting a large repository of data that is useful, a deployment needs to be made to larger and more diverse audiences. It may be that new concerns or benefits may arise with a different set of people or larger social groups that we did not account for. Also, as mentioned earlier, several potential social benefits may not be fully realized until greater adoption or greater peer network penetration occurs.

In the future, we aim to conduct field studies with much larger peer groups, such as at the community or campus level, as well as conduct in-the-wild studies where we publicize Eyebrowse extensively. However, this study has brought up key design considerations that should be addressed before a wider rollout. For instance, the algorithms for determining feeds could be improved to recommend more interesting content. More flexibility regarding sharing preferences could be introduced, such as obfuscating or anonymizing certain aspects. In terms of new features and directions, a strategy will need to be developed to incorporate mobile browsing data, as mobile browsers currently do not permit extensions. Certain additional data, such as referrer information or within-page scrolling data, could be very interesting to collect and share, though they introduce new privacy challenges. We also did not consider other aspects that might be useful or entertaining to users, such as allowing users to rate webpages or gamification, where users can be "mayors" of a site, similarly to Foursquare.

Once a larger web activity data repository is available, it would also be interesting to recruit members of the community to build upon or study the data. Introducing the tool to other user populations or inviting additional research or development may provide other interesting use cases for this data that we may not have envisioned. For example, though recommending web content was a natural feature to incorporate, many other kinds of recommenders could be built given the richness of browsing data. Indeed, releasing this data to outside developers could open up an ecosystem for the development of more tools that benefit the creators of this data as opposed to tools and analysis that stays within corporations. This would create a feedback loop that further incentivizes people to contribute their data.

CONCLUSION

Organizations that track users' browsing activity, often without the knowledge or consent of users, gain massive benefits from this privately held data while users themselves benefit little. At the same time, social media has demonstrated how many users will freely share their data and self-monitor their levels of privacy and publicity when social incentives are built in. In this paper, we investigated the social opportunities and privacy challenges surrounding a tool for public web activity sharing. We outlined several social benefits that could be unlocked with sharing web activity, examined privacy concerns with sharing browsing data, and considered ways for allowing users to control what they share. Our results from conducting both scenario-based hypothetical interviews and a preliminary field deployment of a public web activity sharing tool suggest that a system for sharing some aspects of web activity publicly may be feasible and interesting to users given the proper privacy controls and social incentives.

REFERENCES

1. Mark S. Ackerman, Lorrie F. Cranor, and Joseph Reagle. 1999. Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences. In *Proceedings of the ACM Conference on Electronic Commerce (EC '99)*. ACM, New York, NY, USA, 1–8. DOI: http://dx.doi.org/10.1145/336992.336995

2. danah boyd and Kate Crawford. 2012. Critical questions for big data. *Information, Communication & Society* 15, 5 (2012), 662–679. DOI: http://dx.doi.org/10.1080/1369118X.2012.678878

3. Laura Dabbish, Colleen Stuart, Jason Tsay, and Jim Herbsleb. 2012. Social Coding in GitHub: Transparency and Collaboration in an Open Software Repository. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '12)*. ACM, New York, NY, USA, 1277–1286. DOI:http://dx.doi.org/10.1145/2145204.2145396

4. Nicole B. Ellison, Charles Steinfield, and Cliff Lampe. 2007. The benefits of Facebook "friends:" social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication* 12, 4 (2007), 1143–1168. DOI: http://dx.doi.org/10.1111/j.1083-6101.2007.00367.x

5. Daniel A. Epstein, Bradley H. Jacobson, Elizabeth Bales, David W. McDonald, and Sean A. Munson. 2015. From "Nobody Cares" to "Way to Go!": A Design Framework for Social Sharing in Personal Informatics. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM, New York, NY, USA, 1622–1636. DOI:http://dx.doi.org/10.1145/2675133.2675135

6. Adam Fourney and Meredith R. Morris. 2013. Enhancing Technical Q&A Forums with CiteHistory. In *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM '13)*. AAAI, Palo Alto, CA, USA.

7. Batya Friedman, Daniel C. Howe, and Edward Felten. 2002. Informed Consent in the Mozilla Browser: Implementing Value-Sensitive Design. In *Proceedings of the IEEE Hawaii International Conference on System Sciences (HICSS '02)*. IEEE, New York, NY, USA. DOI: http://dx.doi.org/10.1109/HICSS.2002.994366
8. Kirstie Hawkey and Kori M. Inkpen. 2006. Examining the Content and Privacy of Web Browsing Incidental Information. In *Proceedings of the International Conference on World Wide Web (WWW '06)*. ACM, New York, NY, USA, 123–132. DOI: http://dx.doi.org/10.1145/1135777.1135801
9. E. Herder, R. Kawase, and G. Papadakis. 2011. Experiences in Building the Public Web History Repository. In *Proceedings of DataTEL Workshop of STELLARnet Alpine Rendez-Vous (ARV '11)*.
10. Jason I. Hong and James A. Landay. 2004. An Architecture for Privacy-sensitive Ubiquitous Computing. In *Proceedings of the International Conference on Mobile Systems, Applications, and Services (MobiSys '04)*. ACM, New York, NY, USA, 177–189. DOI: http://dx.doi.org/10.1145/990064.990087
11. Gordon Hull, Heather Richter Lipford, and Celine Latulipe. 2011. Contextual gaps: Privacy issues on Facebook. *Ethics and Information Technology* 13, 4 (2011), 289–302. DOI: http://dx.doi.org/10.1007/s10676-010-9224-8
12. Ellen A. Isaacs, John C. Tang, and Trevor Morris. 1996. Piazza: A Desktop Environment Supporting Impromptu and Planned Interactions. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW '96)*. ACM, New York, NY, USA, 315–324. DOI: http://dx.doi.org/10.1145/240080.240316
13. Aniket Kittur, Bongwon Suh, and Ed H. Chi. 2008. Can You Ever Trust a Wiki?: Impacting Perceived Trustworthiness in Wikipedia. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW '08)*. ACM, New York, NY, USA, 477–480. DOI: http://dx.doi.org/10.1145/1460563.1460639
14. Scott Lederer, Jason I. Hong, Anind K. Dey, and James A. Landay. 2004. Personal privacy through understanding and action: five pitfalls for designers. *Personal and Ubiquitous Computing* 8, 6 (Nov. 2004), 440–454. DOI: http://dx.doi.org/10.1007/s00779-004-0304-9
15. Ian Li, Jeffrey Nichols, Tessa Lau, Clemens Drews, and Allen Cypher. 2010. Here's What I Did: Sharing and Reusing Web Activity with ActionShot. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 723–732. DOI: http://dx.doi.org/10.1145/1753326.1753432
16. Alessandra Alaniz Macedo, Khai N. Truong, José Antonio Camacho-Guerrero, and Maria da Graça Pimentel. 2003. Automatically Sharing Web Experiences Through a Hyperdocument Recommender System. In *Proceedings of the ACM Conference on Hypertext and Hypermedia (HT '03)*. ACM, New York, NY, USA, 48–56. DOI: http://dx.doi.org/10.1145/900051.900061
17. Mary Madden and Lee Rainie. 2015. Americans' Attitudes About Privacy, Security and Surveillance. Pew Research Center. (20 May 2015). Retrieved May 21, 2015 from http://www.pewinternet.org/2015/05/20/americans-attitudes-about-privacy-security-and-surveillance/.
18. Catherine C. Marshall and Frank M. Shipman. 2015. Exploring the Ownership and Persistent Value of Facebook Content. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM, New York, NY, USA, 712–723. DOI: http://dx.doi.org/10.1145/2675133.2675203
19. Alice E. Marwick. 2012. The public domain: Surveillance in everyday life. *Surveillance & Society* 9, 4 (2012), 378–393.
20. Alice E. Marwick and danah boyd. 2011. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society* 13, 1 (2011), 114–133. DOI: http://dx.doi.org/10.1177/1461444810365313
21. Ericka Menchen-Trevino and Chris Karr. 2012. Researching real-world web use with Roxy: Collecting observational web data with informed consent. *Journal of Information Technology & Politics* 9, 3 (2012), 254–268. DOI: http://dx.doi.org/10.1080/19331681.2012.664966
22. Matthew B. Miles and Michael Huberman. 1994. *Qualitative Data Analysis: An Expanded Sourcebook*. Sage, Thousand Oaks, CA, USA.
23. Ashwini Nadkarni and Stefan G. Hofmann. 2012. Why do people use Facebook? *Personality and Individual Differences* 52, 3 (2012), 243–249. DOI: http://dx.doi.org/10.1016/j.paid.2011.11.007
24. Duyen T. Nguyen, Laura A. Dabbish, and Sara Kiesler. 2015. The Perverse Effects of Social Transparency on Online Advice Taking. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM, New York, NY, USA, 207–217. DOI: http://dx.doi.org/10.1145/2675133.2675253
25. Helen Nissenbaum. 2009. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press, Palo Alto, CA, USA.
26. Fuming Shih, Ilaria Liccardi, and Daniel Weitzner. 2015. Privacy Tipping Points in Smartphones Privacy Preferences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 807–816. DOI: http://dx.doi.org/10.1145/2702123.2702404

27. Aaron Smith. 2014. 6 new facts about Facebook. Pew Research Center. (3 February 2014). Retrieved July 27, 2015 from **http://www.pewresearch.org/fact-tank/2014/02/03/6-new-facts-about-facebook/**.

28. Karen P. Tang, Pedram Keyani, James Fogarty, and Jason I. Hong. 2006. Putting People in Their Place: An Anonymous and Privacy-sensitive Approach to Collecting Sensed Data in Location-based Applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 93–102. DOI: **http://dx.doi.org/10.1145/1124772.1124788**

29. Joseph Turow. 2003. *Americans & Online Privacy: The System is Broken*. Annenberg Public Policy Center, Philadelphia, PA, USA.

30. Joseph Turow, Jennifer King, Chris J. Hoofnagle, Amy Bleakley, and Michael Hennessy. 2009. Americans reject tailored advertising and three activities that enable it. *Social Science Research Network* (2009). DOI: **http://dx.doi.org/10.2139/ssrn.1478214**

31. Christian von der Weth and Manfred Hauswirth. DOBBS: Towards a Comprehensive Dataset to Study the Browsing Behavior of Online Users. In *Proceedings of International Joint Conferences on Web Intelligence & Intelligent Agent Technologies (WI/IAT '13)*.

32. Yang Wang, Pedro Giovanni Leon, Alessandro Acquisti, Lorrie F. Cranor, Alain Forget, and Norman Sadeh. 2014. A Field Trial of Privacy Nudges for Facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2367–2376. DOI: **http://dx.doi.org/10.1145/2556288.2557413**

33. Daniel J. Weitzner, Harold Abelson, Tim Berners-Lee, Joan Feigenbaum, James Hendler, and Gerald J. Sussman. 2008. Information Accountability. *Commun. ACM* 51, 6 (June 2008), 82–87. DOI: **http://dx.doi.org/10.1145/1349026.1349043**

34. Chi Zhang, Jinyuan Sun, Xiaoyan Zhu, and Yuguang Fang. 2010. Privacy and security for online social networks: Challenges and opportunities. *IEEE Network* 24, 4 (2010), 13–18. DOI: **http://dx.doi.org/10.1109/MNET.2010.5510913**