Silent Listeners:
The Evolution of Privacy and
Disclosure on Facebook

Fred Stutzman*, Ralph Gross†, Alessandro Acquisti‡

Abstract.   Over the past decade, social network sites have experienced dramatic growth in popularity, reaching most demographics and providing new opportunities for interaction and socialization. Through this growth, users have been challenged to manage novel privacy concerns and balance nuanced trade-offs between disclosing and withholding personal information. To date, however, no study has documented how privacy and disclosure evolved on social network sites over an extended period of time. In this manuscript we use profile data from a longitudinal panel of 5,076 Facebook users to understand how their privacy and disclosure behavior changed between 2005—the early days of the network—and 2011. Our analysis highlights three contrasting trends. First, over time Facebook users in our dataset exhibited increasingly privacy-seeking behavior, progressively decreasing the amount of personal data shared publicly with unconnected profiles in the same network. However, and second, changes implemented by Facebook near the end of the period of time under our observation arrested or in some cases inverted that trend. Third, the amount and scope of personal information that Facebook users revealed privately to other connected profiles actually increased over time—and because of that, so did disclosures to "silent listeners" on the network: Facebook itself, third-party apps, and (indirectly) advertisers. These findings highlight the tension between privacy choices as expressions of individual subjective preferences, and the role of the environment in shaping those choices.

Keywords: social network sites, Facebook, disclosure, privacy

1   Introduction

In recent years, the landscape of social network sites has been characterized by growth and evolution. Virtually all US teenagers use social network sites, as well as almost half of all US online adults [44]—an approximate five-fold increase since 2005. At the same time, social network site platforms have evolved. Facebook and LinkedIn (and, more recently, Google+) have experienced tremendous growth, while early players such as Myspace and Friendster have contracted. The rise of micro-blogging services such as Twitter and Tumblr has threatened more traditional networking sites, contributing to rapid development of new products and evolving norms of privacy. It is against this backdrop of evolution and change that we present an analysis of privacy and disclosure behavior in a longitudinal panel of 5,076 members of the Carnegie Mellon University

*Heinz College, Carnegie Mellon University, Pittsburgh, PA, `mailto:fstutzman@cmu.edu`
†Heinz College, Carnegie Mellon University, Pittsburgh, PA, `mailto:rgross@cs.cmu.edu`
‡Heinz College, Carnegie Mellon University, Pittsburgh, PA, `mailto:acquisti@andrew.cmu.edu`

Facebook network, whose public disclosures we collected between 2005 and 2011.

Much of the current research on privacy in social network sites has been cross-sectional, using surveys or data mining to explore social network sites at one or two points in time. However, in the last decade, the use of social network sites has rapidly evolved, as new participants joined various platforms and individuals iteratively changed disclosure and privacy settings. In this manuscript, we attempt to tell the story of this evolution through the lens of a set of Facebook members, combining quantitative analysis of their public disclosures and qualitative analysis of changes to the Facebook network over time. The longitudinal dataset, created with Institutional Review Board (IRB) permission and de-identified, contains only publicly visible content shared by members of the Carnegie Mellon University (CMU) Facebook network with the rest of the network. Within the dataset, we construct and examine the disclosure behaviors of a specific subset of users. In particular, we follow a panel of early joiners of the network, and we explore what was being shared on their Facebook profiles, with whom, and how this sharing changed over time. One limitation of this data is that it cannot reflect a random sample of *current* Facebook users; hence, extrapolations to the general Facebook population should be considered with caution. On the other hand, its longitudinal nature offers an unprecedented view of the long term evolution of privacy and disclosure behavior on a social network site.

Taking a long-term longitudinal perspective allows us to connect three contrasting trends, and uncover field evidence consistent with the results of lab experiments focused on understanding privacy and disclosure behavior. When considered together, those trends bear witness to the role of, and tensions between, both endogenous (user-driven) and exogenous (network-driven) factors in influencing privacy choices and the dynamics of online disclosure.

First, we find quantitative evidence that, over time, Facebook users in our dataset exhibited increasingly privacy-seeking behavior: they became more protective of their personal information by progressively limiting data publicly shared with "strangers" (profiles in the same Facebook network but unconnected to the user).[1] This pattern is consistent across all profile fields (or data types) we investigated. This first result may be interpreted as an example of consumers, as autonomous agents, expressing and acting upon their subjective preferences while negotiating with a service—the resulting outcome being that some consumers increasingly took privacy choices.

However, we also observe a second result: policy and interface changes implemented by Facebook near the end of the period of time under our observation seemingly altered that outcome and countered such privacy-seeking behavior by arresting and in some cases inverting the trend we just described.[2] Specifically, between 2009 and 2010, we observe a significant *increase* in the public sharing of various types of personal informa-

---

[1]The scope and meaning of "public" sharing changed over time on Facebook. In early 2005, users could share publicly with the rest of their .edu network, but not with members of other networks. By 2011, it was possible to share publicly with any other user on Facebook. For consistency, and since our analysis begins in 2005 and is conducted from the point of view of an unconnected profile on the CMU Facebook network, here and in the rest of the manuscript by the term "public" we refer to sharing with the rest of the network.

tion. We causally link this trend reversal to specific changes in Facebook's site interface and default settings, and find that by the time our data collection ended (May 2011), disclosures in the majority of fields had *not* gone back down to the levels reached before those changes. This second result, therefore, highlights the power of the environment in affecting individual choices: the entity that controls the structure (in this case, Facebook), ultimately remains able to affect how actors make choices in that environment. Such finding, based on field data, is consistent with previous experimental evidence highlighting the role of asymmetric information [1] and default settings [5] in affecting privacy choices.

Third, and finally, we discuss qualitative evidence that, over time, the amount and scope of personal information that Facebook users revealed to their Facebook "friends" (that is, to other connected profiles) actually increased. In doing so, however—and in parallel to their *reducing* disclosures to stranger profiles—users ended up increasing their personal disclosures to other entities on the network as well: third-party apps, (indirectly) advertisers, and Facebook itself. Sometimes, this occurred without users' explicit consent or even awareness [7, 34, 60].[3] Hence, the "silent listeners." When linked to the two previous findings, this latter trend appears consistent with recent experimental evidence in the field of privacy decision making. Access to increasingly granular settings (which help individuals determine which profile data other Facebook users get to peruse) may increase members' feeling of control and selectively direct their attention towards the sharing taking place with other members of the network covered under those settings; in turn, perceptions of control over personal data [11] and misdirection of users' attention [4] have been linked in the literature to increases in disclosures of sensitive information to strangers. Social network sites remain in part "imagined" communities [2], where intended audiences do not necessarily map to actual audiences.

Considered together, these findings illustrate the challenges users of social network sites face when trying to manage online privacy, the power of providers of social media services to affect their disclosure and privacy behavior, and the potential limits of notice and consent mechanisms in addressing consumers' online privacy concerns [4]. More broadly, the results highlight the tension between privacy choices as expressions of individual subjective preferences, and the role of the environment—as well as factors such as asymmetric information, bounded rationality, or cognitive biases—in affecting those choices. Like a modern Sisyphus, some consumers strive to reach their chosen "privacy spot"—their desired balance between revealing and protecting—only to be taken aback by the next privacy challenge.

---

[2]Our references to policy and interface changes and reversals of previous trends merely describe a process and its consequences; in this manuscript, we do not investigate the motivations or goals that inspired said changes.

[3]See, also, the Federal Trade Commission's 2011 complaint *In the Matter of FACEBOOK, INC.*, at `http://ftc.gov/os/caselist/0923184/111129facebookcmpt.pdf`, last accessed on February 25, 2013.

2   Background

Between 2004 (the year Thefacebook was launched) and 2011, social network sites experienced remarkable growth in active users, as well as shifts in popularity. In particular, even as more niche and diversified players started appearing, the market consolidated, with Facebook achieving dominance.

Facebook's transition from a university-focused social network site for students to a global social network site, however, was not seamless. In particular, changes to Facebook network structure and a series of unpopular moves raised users' privacy concerns. When Facebook debuted in 2004, the site was segmented by university, so that university network membership created a meaningful privacy boundary. Although university networks could be quite large (e.g., [2, 54]), this boundary generally separated students from family, employers, and municipal law enforcement. Starting in 2006, Facebook gradually liberalized its policies for site membership, and began changing (and eventually discounting) the value of "networks" within the service. Facebook was rewarded for these moves, with adoption climbing through a billion users as of late 2012.[4] As for its users, Facebook's growth in popularity proved both an opportunity and challenge. Increasingly, users could articulate a greater portion of their "social graph" in the service, and take advantage of the benefits reaped from the establishment and maintenance of large weak-tie networks [20]. With the presence of a larger portion of users' personal networks, however, came new challenges, such as the presence of multiple contextual networks in the site—a phenomenon highlighted by Acquisti and Gross by referring to early campus-based Facebook networks as "imagined" communities [2]. Individuals faced challenges as they attempted to share information in the presence of coworkers, family, and distant friends in a single social network site (e.g., [17, 21, 53]).

The hurdles of managing disclosure across multiple social contexts in a social network site have been referred to as "context collapse" [50]. According to Lampinen and colleagues [40, 41], individuals employ a range of strategies to manage multiple contexts in social network sites, including self-censorship and withdrawal of content, creating more inclusive group identities, and sharing different types of content in different spaces. In addition to these behavioral and mental strategies for context and privacy management, individuals also turn towards the application of privacy settings within the site. Numerous studies documented both an increase in privacy awareness and privacy-seeking behaviors within Facebook by students [9, 44, 46, 59] and the contextual application of privacy settings in relation to perceived harms [15, 55].

Research on social network sites often explores the disclosure behavior of social network site users. Because social network sites thrive on peer-produced content, disclosure is often concomitant with site use. The definition of social network sites, posed by boyd and Ellison, illustrates the importance of disclosure. The authors state that social network sites allow the creation of "public or semi-public profiles within a bounded system" and foster the articulation of lists of personal connections within the system [8]. That is, a social network site is characterized by what you say about yourself, who you choose

---

[4]See "Newsroom - Key Facts," Facebook: `http://newsroom.fb.com/content/default.aspx?NewsAreaId=22`, last accessed on February 25, 2013.

to publicly associate with, and increasingly, what your connections say about you.

The earliest studies of social network site use provided empirical evidence of the remarkable disclosure practices within the site. Early work by Gross and Acquisti [28], and Acquisti and Gross [2] found that students in the Carnegie Mellon University Facebook network extensively shared sensitive information such as political views and sexual orientation on Facebook, and that information shared on Facebook was generally self-reported as valid. Other studies conducted at the time in different university networks, including Stutzman [54] and Lampe et al. [37], further evidenced the high degree of personal disclosure within social network sites. Large-scale studies such as Thelwall [57] and Caverlee and Webb [14] provided evidence of similar disclosure phenomena in Myspace, once the leading social network site. These findings were corroborated by a national probability study conducted by Lenhart and Madden [43].

Researchers have considered a range of motivations for disclosure in social network sites. Drawing on the work of Erving Goffman [25], Donath and boyd [18] and boyd and Heer [10] described the use of a social network site as a performance of identity. Users of social network sites are therefore challenged to strategically present themselves, through the constructed profile, to their increasingly diverse network of social ties. As noted by Lampe et al. [37], the motivations for use, and disclosure within, a social network site are a function of offline outcomes such as relational formation and deepening. Work by Bumgarner [12] and Joinson [33] illustrated the social motive of social network site use: use of social network sites was driven by participant desire to connect and learn about one another; without significant personal sharing in these sites, these motives of use would not be addressed.

In addition to the shifting nature of networks in Facebook, changes to the interface and site policies have produced public backlash that has increased awareness of privacy in social networks (e.g., [27, 30, 31]). To encourage disclosure on the platform, Facebook has consistently changed the nature of sharing certain items in the platform, and the default sharing settings for new accounts.[5] These site-directed changes may affect disclosure by altering the level of trust individuals have in Facebook, which was initially described as a more trusted social network site when compared with Myspace [19, 22].

While Facebook has changed substantially over the years, to our knowledge no large-scale, long-term longitudinal analysis has yet described the effects of these changes. For example, the large-scale analysis employed by Gjoka et al. [24, 23] focuses on a short time period. Recent work by Dey, Jelveh, and Ross [16] employs a longitudinal perspective, but is limited to a two-year window (2010 and 2011). Similar works by Lewis [45] and colleagues [46] draw on the Time, Taste, and Ties Harvard network Facebook dataset.

---

[5]See, for instance, "Facebook's Eroding Privacy Policy: A Timeline," Electronic Frontier Foundation: https://www.eff.org/deeplinks/2010/04/facebook-timeline, last accessed on February 25, 2013. An insightful graphical representation of how default visibility settings for Facebook profiles changed over the period 2005–2010 was provided by Matt McKeon at http://mattmckeon.com/facebook-privacy/, last accessed on February 25, 2013.

3 The Carnegie Mellon Facebook Dataset

Between 2005 and 2011, we collected a longitudinal database of information publicly disclosed on Facebook profiles, documenting the online disclosure behavior of Facebook users at Carnegie Mellon University (the *Carnegie Mellon Facebook Dataset*, or CMFD). The information captured in the dataset reflects public disclosures within the Carnegie Mellon University (CMU) Facebook network. The CMFD is comprised of multiple snapshots of the CMU Facebook network over a seven year period. This "omnibus" dataset starts with just over 3,000 profiles in the summer of 2005, and grows to over 20,000 profiles by 2011, as more students enrolled at Carnegie Mellon University and joined its Facebook network. To collect these snapshots, we navigated the CMU Facebook network, capturing information disclosed publicly to other members of the network. The CMFD was collected with the approval of the Carnegie Mellon University Institutional Review Board, and then de-identified.[6] To produce this dataset, we post-processed the snapshots to create archival and analytic datasets. In the period of time covered by the database, Facebook grew substantially, shifting from being a college-centric social network site (initially open only to individuals with a Harvard University email address, then extended to other educational institutions) to the industry leader it is today (eventually, open to everyone). While a detailed history of Facebook's features is beyond the scope of this manuscript, our data does capture how Facebook's growth deeply impacted its members' disclosures.

To develop the initial seeds for profile data collection, we used Facebook's predictable user ID scheme to identify Carnegie Mellon University Facebook profiles. In early iterations of the service, Facebook generated network-specific user ID's, which auto-incremented when new users joined the network. For example, the first user of the Carnegie Mellon University network has a Facebook user ID of 4800001, the second has a user ID of 4800002, and so on. Due to this predictable user ID scheme, we were able to effectively take an early census of virtually all Facebook users in the Carnegie Mellon network.

The data for the Carnegie Mellon Facebook panel was collected from the perspective of an unconnected member of the Carnegie Mellon Facebook network. Practically, this means that we employed a profile that was situated in the Carnegie Mellon Facebook network but had no friendship connections. It is important to utilize an unconnected profile both for privacy considerations and to control for confounds in disclosure behavior that may be attributable to friendship connections, or friends-of-friends status.

Over the course of the seven years of data collection, Facebook's interface changed substantively and several times, requiring the refinement of existing techniques and the development of new data collection techniques. For example, as Facebook became more popular, individuals joined the CMU Facebook network as secondary networks (where the previous network was the primary high school, undergraduate, or work network). To capture these members of the CMU Facebook network that did not have a predictable

---

[6]By de-identification we refer to the process of removing unique identifiers (such as names or Facebook IDs) from the dataset. Naturally, extant research on statistical re-identification—see, for instance, [56]—has demonstrated that de-identified data can often be re-identified by third-parties.

Table 1: Collection Dates, Carnegie Mellon Yearly Snapshot Dataset

Time	Year	Collection Date	Total Obs.	Panel Obs.
t_0	2005	Nov 20	6380	5076
t_1	2006	Nov 29	10254	5076
t_2	2007	Nov 02	15041	5076
t_3	2008	Mar 01	15324	5076
t_4	2009	Oct 04	15024	5076
t_5	2010	Nov 12	15731	5076
t_6	2011	May 5	22124	5076

user ID (as their user IDs were not created as part of the CMU Facebook network), we analyzed seed members friends lists to identify new entrants. Through inspection of the network and observation of the new account ID assignment pattern we were able to effectively capture existing Facebook members as they joined the CMU network. Although we attempted to exhaustively capture the Carnegie Mellon Facebook network at each snapshot, it is likely that earlier snapshots, such as 2005 and 2006, are more exhaustive given the linear nature of Facebook ID assignment.

3.1   The Carnegie Mellon Yearly Snapshot Dataset

Within the omnibus Carnegie Mellon Facebook dataset, we constructed a yearly panel dataset comprised of early-adopting individuals for whom we have profile data for the years 2005–2011. This dataset, the *Carnegie Mellon Yearly Snapshot Dataset*, is the unique longitudinal collection of Facebook profile dynamics which we analyze in this paper. The general structure of the dataset is presented in Table 1. Over the seven snapshots $t_0 \ldots t_6$ we were able to construct a balanced panel with 5,076 profile observations, collected at one year intervals in the fall (with the exception of 2008, which was collected in March).

To ensure that panel mortality (the decrease in the size of the panel from one snapshot to the next) did not bias our estimates, we explored why observations dropped from the panel over time. At t_0, our crawler recorded 6,380 possible Facebook profiles using an incremental crawl of the CMU Facebook user ID space. For a number of reasons, not every user ID corresponded to an active profile,[7] which reduced our valid t_0 observation set to 6,142 observations, producing an effective panel mortality rate of 1066 members. We explored the patterns of mortality of the 1066 members and found that 72% of mortality was attributable to a single missed observation between $t_1 \ldots t_6$. These missed observations are likely the result of download errors, Facebook website troubles, or other random error, which leads us to believe that mortality patterns in our panel are random and non-systematic.

---

[7] For example, an individual may have created, and then deleted, their account, which effectively
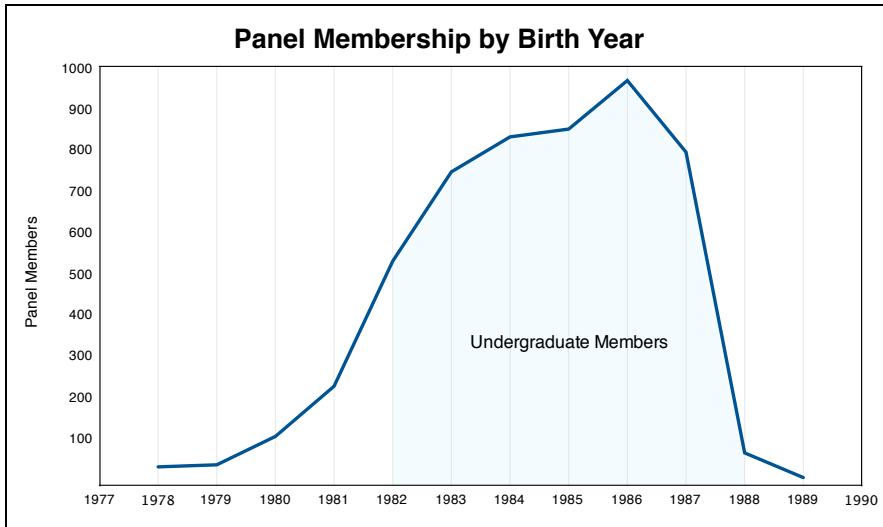
Panel Membership by Birth Year

Figure 1: Raw distribution of reported birth years (n=5,076) for the Carnegie Mellon Yearly Snapshot Dataset. The shaded area represents likely undergraduate members of the panel. An additional 284 members of the panel reported birth dates that were not within the 1978–1989 range.

To understand the panel composition, we analyzed member self-disclosures at t_0, the date of panel entry for all members. At the time of panel entry, approximately 87% (n=4,415) of panel members shared their date of birth. Of those members that shared their birth, approximately 95% indicated a birthdate between 1978 and 1989, and approximately 88% indicated a birthdate between 1982 and 1988, the likely birth years for Carnegie Mellon undergraduates in 2005 (Figure 1). Practically, this means that the majority of the panel is composed of undergraduate students, which matches expectations (Facebook was primarily an undergraduate phenomenon in 2005). We found that 60.93% (n=3,093) of panel members are male, and 37.57% (n=1,907) of panel members are female. Seventy-five panel members did not self-report gender. According to official Carnegie Mellon statistics,[8] the school maintains a 3:2 Male/Female ratio. Therefore, our self-reported statistics are in line with the expected CMU population, though we do observe significant overrepresentation of males in the panel ($p = 0.0018$, one sample t-test) as compared to the CMU population. Finally, we found that approximately 16% of our panel explicitly reported that their hometown was located in Pennsylvania (as is CMU).

"locked" the user ID to a non-existent profile. To examine this, we visited a selection of "null" t_0 profiles and did not find any content (i.e., 404 error), which means these IDs are not appropriate for inclusion in the panel.

[8] "Carnegie Mellon at a Glance," http://www.cmu.edu/enrollment/pre-college/about-facts.html, last accessed on February 25, 2013.

4 Data Trends in the Carnegie Mellon Yearly Snapshot Dataset

In this section we describe the dynamics of disclosure in the Carnegie Mellon Yearly Snapshot Dataset over the last seven years. We focus on "public" disclosures to the rest of the CMU network, and on a subset of profile elements that were common across all of the years.

As previously noted, Facebook initially employed a network model of privacy segmentation. Individuals that created an account in a network, or joined a network, would have their privacy settings defaulted to open for that network, a functionality that continues today. Therefore, public disclosures to the Carnegie Mellon network are significant for two reasons. First, a public disclosure to a network is a disclosure to a deterministically large public of Carnegie Mellon students, faculty, staff, and alumni. As this network contains individuals that may hire or write letters of recommendation for an individual, disclosures to this large public are not without risk (or potential reward). Second, we are able to infer generally that individuals that no longer have content available to the network have likely either changed their privacy settings, or stopped sharing content. While we are not able to completely differentiate between the two events, both are noteworthy from the perspective of our research question as they indicate a shift in individual disclosure behavior in the social network site.

Facebook changed substantially between 2005 and 2011, with the introduction of technologies such as the news feed, the "Like" button, and photo tagging. Furthermore, the degree of control afforded to users over the visibility of their personal information changed as well, becoming more granular over time (from coarse settings commanding the visibility of an entire profile in 2005, to minute oversight over the visibility of distinct pieces of disclosed information in 2011). Even with these changes, the core of the Facebook profile—in which individuals share information about themselves—has remained fairly consistent. Just as in 2005, today an individual can share her birthdate, hometown, contact information, and favorite books and movies on their Facebook profile. In this manuscript, we focus on a subset of Facebook profile elements, or "fields", which remained consistent over the panel's life. We do not claim that this analysis is necessarily generalizable to all types of sharing on the profile, as innovations like the status updates and photo tagging may account for a large portion of an individual's current Facebook disclosures. However, we do argue that this analysis provides an internally consistent, longitudinal description of the changes in public sharing to a large campus network over time, and from this analysis we can more generally understand Facebook privacy attitudes and action.

In this analysis, we concentrate on twelve profile elements, described in Table 2. These elements were selected because they are commonly shared, and generally applicable to all subjects in the dataset. Other fields, such as graduate school, or employment history, may not be applicable to all subjects, and create a confound when studying disclosure behavior. We feel the twelve profile elements are diverse, and provide general insight into the dynamics of profile disclosure over the panel.

Table 2: Facebook Profile Elements

Element	Disclosure Category	Note
Birthdate	Personal	FT (mm-dd-yyyy)
High School	Personal	FT
Hometown	Personal	FT
Political Affiliation	Personal	Initially DD, later FT
Instant Messenger	Contact	FT, any IM (AIM, Skype, Y!, etc)
Phone	Contact	FT, any phone (mobile or landline)
Address	Contact	FT
Looking For	Contact	DD
Interests	Interests	Initially FT, later L
Favorite Music	Interests	Initially FT, later L
Favorite Books	Interests	Initially FT, later L
Favorite Movies	Interests	Initially FT, later L

Disclosure category refers to the type of disclosure, as categorized by the authors. Under Note, FT: free text input, DD: drop down list, L: "Like" button. For a number of items, the way the items are added to the profile changes over time, which is noted.

In constructing the list of twelve profile elements, we categorized the disclosures either as personal information, contact information, or interests.[9] Personal information is profile sharing that reveals significant personal and background information about an individual. Contact information, on the other hand, is information that either enables communication or signals opportunity for communication. For these reasons, we included "Looking for" (i.e., looking for a type of relationship) as a communication opportunity. Finally, "Interests" information refers to declared personal interest, such as favorite books or movies. As many of the variables were unstructured, we created a number of derivative variables that are included in this analysis. These include a variable that measures if the subject shares the profile element, and a count of items shared within the profile element (for full-text interests).

4.1 What Are People Sharing?

Figure 2 is a heat map that displays the proportion of the population sharing with the rest of the network the twelve profile elements our manuscript focuses on, at each time interval $t_0 \ldots t_6$. In the visualization, higher proportions of disclosure are indicated by red colors, and lower proportions are indicated by blue colors. The raw means corresponding to the heat map are reported in Table 3. They should be interpreted as the proportion of the population sharing the particular profile element.

In the early days of Facebook, when profiles were only available to be viewed from

---

[9]The categorization was inspired by [38]. It does not affect the results of the analysis we present in the rest of this section, which is at level of individual profile elements.
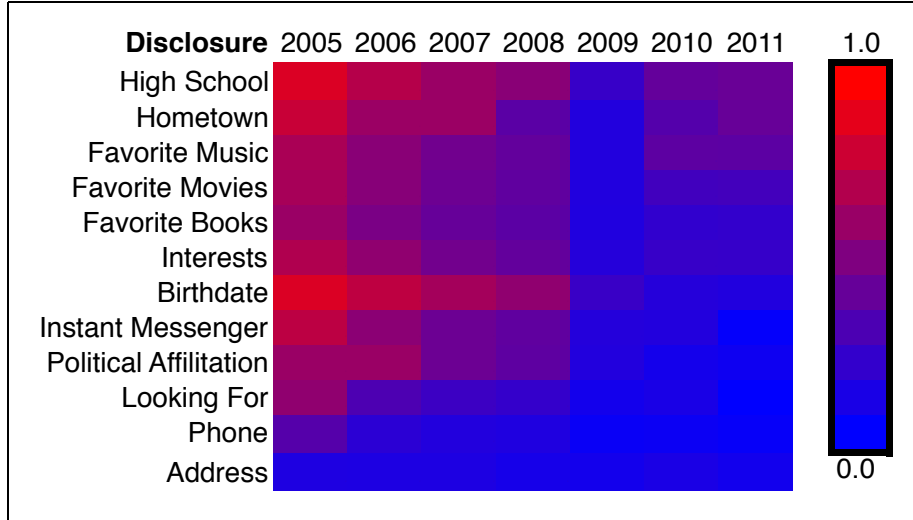
Figure 2: Heatmap visualization of disclosure patterns in the Carnegie Mellon Yearly Snapshot Dataset. As indicated by the legend, red colors indicate higher proportions of disclosure, and blues indicate lower proportions of disclosure. All variables are effects-coded at the unit level to capture if the participant shares the specific element.

within a Facebook network, users tended to share much information publicly [28]. For example, over 70% of our panel shared their birthdate, high school, hometown, and instant messenger information. However, there were some types of information less commonly shared—the prime examples being the individual's residential address (12%) and their phone number (33%). This finding is consistent with previous research, and it hints to Facebook's use as a vehicle for identity formation and expression as a primary early motivator of use (e.g., [39]).

Examining panel sharing over time, we see a consistent reduction in the proportion of the population sharing all profile elements, as the chart moves from reds and purples in 2005 to dark blues in 2009. As Facebook grew in popularity, and new features were added, the individuals in our panel became less public with the information they shared. In 2010, we begin to notice a reversal in the trend of certain types of information sharing. That is, in previous years we have observed declining public sharing for almost every variable we collected; in 2010, we see this trend of declining public sharing "reverse" in that for certain variables, public disclosure starts increasing. While this reversal does not mitigate all of the previous reduced disclosure (e.g., it does not restore disclosure levels to their 2004 maximums), it is interesting to observe. For this reason, we explore some of the potential causes for this reversal in depth throughout the paper.

Table 3: Public Disclosure of Information, 2005–2011

Element	t_0	t_1	t_2	t_3	t_4	t_5	t_6
Birthdate	.862	.747	.645	.568	.224	.146	.132
High School	.860	.713	.606	.538	.219	.392	.413
Hometown	.785	m	.611	.355	.136	.333	.408
Political Affiliation	.604	.498	.423	.369	.136	.082	.058
IM	.739	.547	.426	.378	.144	.135	.019
Phone	.336	.170	.134	.122	.037	.046	.026
Address	.120	m	.114	.086	.077	.104	.074
Looking For	.567	.305	.237	.204	.077	.101	m
Interests	.691	.564	.447	.390	.149	.218	.215
Favorite Music	.669	.537	.445	.389	.135	.368	.362
Favorite Books	.607	.481	.401	.357	.128	.194	.204
Favorite Movies	.656	.529	.431	.380	.131	.257	.264

Note: Raw means represent proportions of the population sharing the particular profile element. Missing observations coded as "m."

4.2   Public Disclosure Trends: Personal Information

We now present an in-depth analysis of profile disclosure for each of the three disclosure categories: personal information, contact information, and interests. By personal information disclosure we refer to profile elements such as birthdate, high school, hometown, and political affiliation. Lampe et al. [38] note that the sharing of personal information may serve as a signal that effectively "verifies" a profile as an accurate representation of the offline identity. As Facebook profiles exist between online and offline context, these types of disclosures establish the individual in the eye of observers by providing important, if somewhat mundane details about the individual. Of course, sharing even mundane information may have significant implications. For instance, as Acquisti and Gross [3] identified, the combination of hometown and birthdate information could be effectively used to predict an individual's social security number with reasonable accuracy. Because personal information disclosures are important in relational formation, we expect high levels of disclosure early in the panel. Over time, the need to share personal information publicly may decrease as this information tends to be static (generally, one's birthdate or hometown does not change).

As demonstrated in Figure 3, individuals at waves t_0 (2005) and t_1 (2006) overwhelmingly share personal information to the campus network. We observe a robust and quasi-linear downward slope across the years in the sharing of personal information for all four profile elements. While the obvious trend in the data is less public disclosure over time, the data appears to show a trend reversal taking place between 2009 and 2010 for two personal disclosure elements: high school ($p=0.000$, group-means t-test) and hometown ($p=0.000$, group-means t-test). To ensure this reversal was not a data
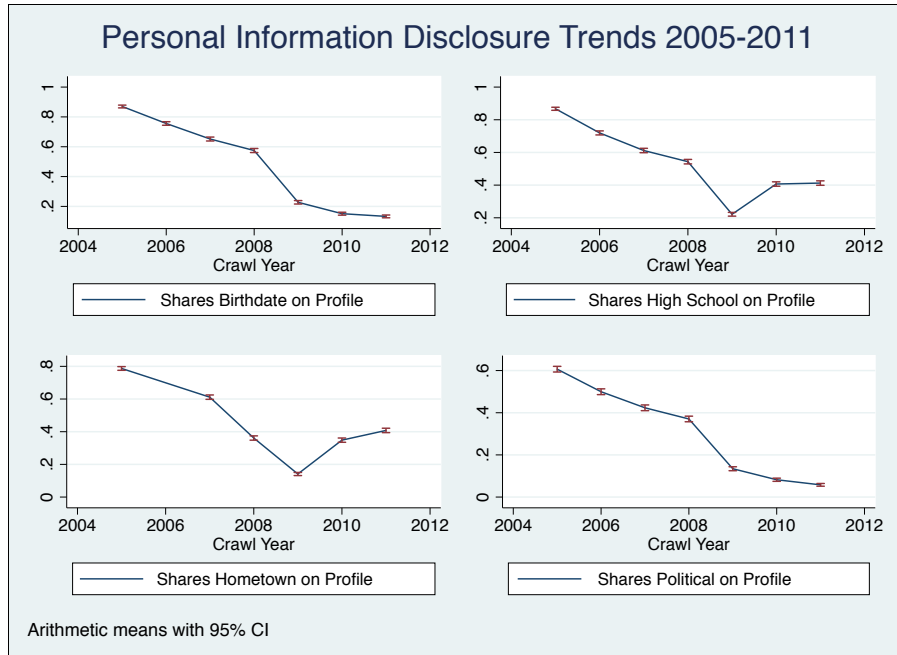
Figure 3: Personal information disclosure trends, 2005–2011. Note: trend lines are scaled.

collection error, we went back to raw data and verified that this trend reversal was legitimate. As we note below, the reversal is robust across other profile elements (such as contacts and interests). We believe this trend reversal to be the direct result of policy and interface changes on behalf of Facebook, and we discuss those further in Section 4.5.

4.3 Public Disclosure Trends: Contact Information

We define contact information disclosure to be represented by the profile elements instant messenger, phone, address, and *Looking for*. The first three elements (IM, phone, and residential address) represent explicit vectors for communication, as in information that can be used to gain access to the person. The fourth element (*Looking for*) is a Facebook convention that signals a degree of willingness for communication. *Looking for* is shorthand for "what type of a relationship are you looking for?" and responses can range from simple friendship to a serious relationship. Whereas the first three elements are vectors for contact information, *Looking for* can be understood as a call-to-action for contact.

Figure 4 displays a similar set of trends in public disclosure of contact information
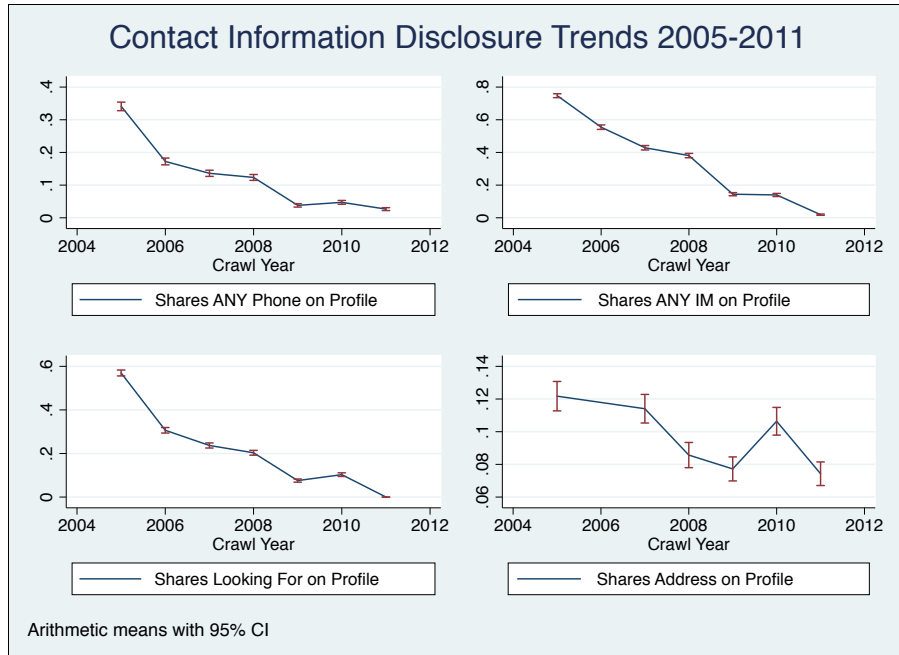
Figure 4: Contact information disclosure trends, 2005–2011. Note: trend lines are scaled.

to the one we observed in public disclosure of personal information. Raw means are reported in Table 3, and should be interpreted as the proportion of the population sharing that particular profile element. Between waves t_0 and t_4, we observe less contact information being shared with the network. As with personal information, we attribute the general declining trend to the turn towards private profiles. At t_5, we observe the trend either pausing (e.g., phone, IM, Looking for) or in fact reversing (e.g., address information; $p=0.000$, group-means t-test).[10] We believe this general reversal to be the direct result of policy and interface changes on behalf of Facebook (see Section 4.5).

4.4   Public Disclosure Trends: Interest Information

We conclude the analysis by exploring disclosure of interest information, which we define to consist of an individual's stated interests, and his/her stated favorite movies, books, and music. The sharing of interest information is an explicit performance of identity, from which outside individuals can draw inference about an individual's favorite pastimes, their personality, and even their social status (e.g., [47]). Lampe and colleagues

---

[10]For address information, the decreasing trend resumes in t_6. Note, however, that trend lines are scaled in Figure 4, and the y-axis scale for Address information is different—and reflects much smaller absolute changes—than the y-axes in the other quadrants in the figure.
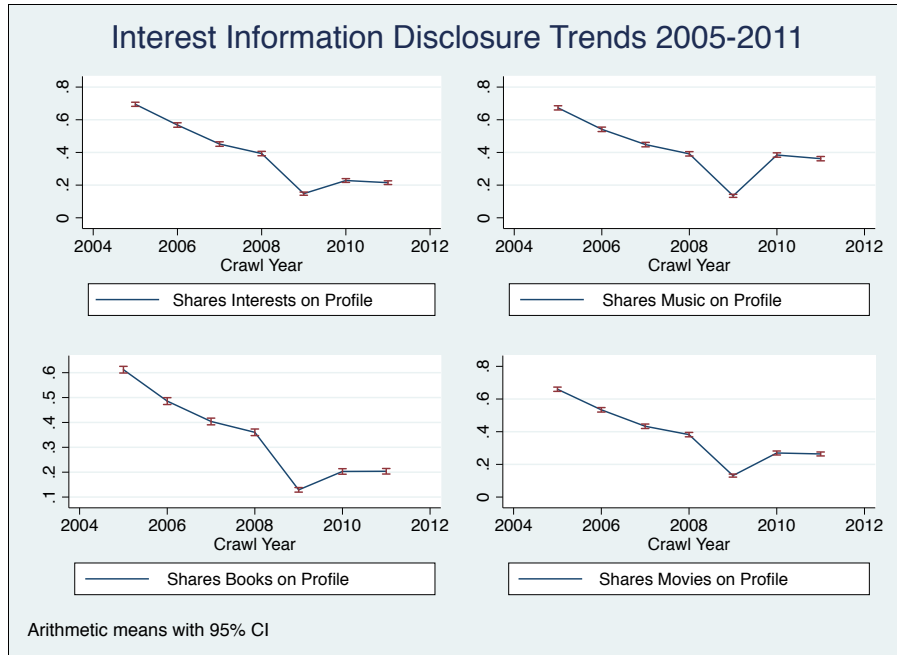
Figure 5: Interest information disclosure trends, 2005–2011. Note: trend lines are scaled.

[38] argue that interest information is important because it creates opportunities for finding common ground between individuals. This process of finding common ground—such as a shared favorite book or movie—may ease the process of relational formation or deepening. As college students are frequently building new relationships and deepening existing relationships, we expect a high degree of interest information disclosure. Panel trends are reported in Figure 5.

Again, we observe the consistent trend of decreasing disclosure of interest information between waves t_0 and t_4, and a sharp reversal of the trend between waves t_4 and t_5 ($p=0.000$, group-means t-test). Raw means are reported in Table 3, and should be interpreted as the proportion of the population sharing the particular interest element.

4.5   Trend Reversal in Public Disclosures

In the previous sections, we have highlighted a robust decreasing trend and a smaller but significant trend reversal. The general trend consisted of a continuous decline in public disclosure of Facebook elements between the years 2005 and 2009. Over time and as Facebook expanded, individuals in the panel either locked down their accounts with privacy settings or employed other disclosure reduction techniques, such as the removal

of content from the profile. Indeed, and broadly speaking, our panel individuals shared less information publicly in 2011 than they did when they first joined Facebook. This could be due to a variety of factors, including: evolution in subjective privacy preferences; access to more granular privacy settings; growing expertise with settings and controls; increasing awareness of, or attention devoted to, privacy risks in online social networks. This finding is consistent with other recent research, which has highlighted growing privacy awareness among Facebook users [27, 30, 31] (compared with early studies of the network [2]) and an uptick in privacy-seeking behaviors [15, 55, 16]). However, to our knowledge, the Carnegie Mellon Yearly Snapshot Dataset offers the first longitudinal field evidence of how the trend started in the early days of the network, and progressed over several years of Facebook history. Our data is also consistent with self-report surveys indicating that, over time, an increasing number of Facebook users chose more protective privacy settings [42, 48], which leads us to believe that the reduction in disclosure in our dataset was likely due to the adoption of privacy settings, rather than merely information removal.[11]

For a certain number of elements, however, we identified a reversal of this trend: between the years 2009 and 2010, individuals started sharing their high school, hometown, address, interests, and their favorite movies, books, and music in significantly higher amounts. While the general decreasing trend appears consistent with extant research, the combination of that trend and the smaller trend reversal is surprising. Having observed this trend to be robust across a number of profile information elements, we turn to an analysis of its possible explanations, focusing on changes in Facebook's site interface and default settings in late 2009 and early 2010.

Note, first, that the trend reversal was observed in our dataset between t_4 and t_5; that is, between the snapshot of the CMU Facebook network taken on October 4, 2009, and the snapshot taken on November 12, 2010. In analyzing such trend reversal, we were first struck by its selective nature of the reversal; in particular, we were interested in why the reversal only affected certain types of profile elements. After inspecting the data, and ruling out data corruption as a cause, we explored policy and design changes instituted by Facebook in the period between our 2009 and 2010 data collections, and ended up concluding that the reversal was, with high probability, the result of policy and interface changes by Facebook—in particular, the redesign of Facebook's privacy interface announced on December 9, 2009[12] and the introduction of Facebook community pages and connected profiles announced on April 19, 2010[13]—which took place in the period within our t_4 and t_5 snapshots.

The introduction of a new privacy interface to Facebook was described as follows: "We're happy to be offering you simpler tools to control your experience on Facebook.

---

[11]We can exclude the possibility that the results are driven by students graduating and leaving Carnegie Mellon University: over 98% of panel participants remained members of the CMU Facebook network as of 2011. Besides, leaving the network would not invalidate the basic premise of the analysis, which focuses on what of a Facebook profile remains visible to someone else in the CMU network.

[12]See "New Tools to Control Your Experience," Facebook Blog: `https://blog.facebook.com/blog.php?post=196629387130`, last accessed on February 25, 2013.

[13]See "Connecting to Everything You Care About," Facebook Blog: `https://blog.facebook.com/blog.php?post=382978412130`, last accessed on February 25, 2013.

We encourage you to take the time to explore them and consider what settings are right for you."[14] Through the addition of highly granular privacy controls, Facebook argued that individuals would be better able to share information with audiences of their choice. However, Facebook's new privacy interface proved to be confusing to users, resulting in public retractions and updates by the company.[15] In fact, the Facebook privacy changes became one of the major issues in a 2011 FTC complaint against Facebook that led to the settlement barring "further deceptive privacy changes."[16] Among other parts of the FTC's claim, the following is alleged:

> The Privacy Wizard did not adequately disclose that users no longer could restrict access to their newly-designated PAI via their Profile Privacy Settings, Friends App Settings, or Search Privacy Settings, or that their existing choices to restrict access to such information via these settings would be overridden. For example, the Wizard did not disclose that a user's existing choice to share her Friend List with Only Friends would be overridden, and that this information would be made accessible to the public.

Choosing Facebook privacy settings to correctly match one's preferences can be difficult. Although our data does suggest that, between 2005 and 2009, Facebook users may have become increasingly willing—and able—to protect information on their profiles, disclosure errors remain possible.[17] Following Facebook's December 2009 changes, users' cognitive burden—and therefore opportunities for errors—arguably increased. New information was deemed "Publicly Accessible Information" (PAI), which was not directly identified in the Facebook Privacy Wizard. Hence, when individuals stepped through the wizard, in agreeing to the terms, they unknowingly grandfathered in previous private information. Thus, now it was public (an example of how asymmetric information between users and providers of a service [1], and default settings [5], may affect privacy outcomes).

The new interface may have caused users to make information public that was not previously public, therefore contributing to the trend reversal. However, this does not address the question about why the trend reversal was selective, affecting only certain elements of the profile. For this question, we turned to the rollout of Facebook's community pages and connected profiles. These changes are described on the Facebook blog as follows:

> What if you could take this one step further, by linking your profile to

---

[14]See "New Tools to Control Your Experience," Facebook Blog: `https://blog.facebook.com/blog.php?post=196629387130`, last accessed on February 25, 2013.

[15]See "Updates on Your New Privacy Tools," Facebook Blog: `https://blog.facebook.com/blog.php?post=197943902130`, last accessed on February 25, 2013.

[16]See "Facebook Settles FTC Charges That It Deceived Consumers By Failing To Keep Privacy Promises," Federal Trade Commission: `http://ftc.gov/opa/2011/11/privacysettlement.shtm`, last accessed on February 25, 2013.

[17]Even in 2006—when privacy settings were fewer and simpler than in 2009—Acquisti and Gross found dichotomies between how visible some Facebook members believed their profiles to be, and how visible their profiles actually were [2]; similar results were also found by [49].

Pages about your interests, affiliations and favorite activities? Today, we're adding two features that do just that.

Community Pages

Community Pages are a new type of Facebook Page dedicated to a topic or experience that is owned collectively by the community connected to it. Just like official Pages for businesses, organizations and public figures, Community Pages let you connect with others who share similar interests and experiences.

More Connected Profiles

Some of you added information about yourself, such as your likes and interests, favorite books, music and movies, when you first joined Facebook. But we've noticed that more than three times as many of you have connected to Facebook Pages, such as those for bands, non-profits, universities or anything else you care about, as a way to express yourself. So to make it even easier to display your affiliations, we've improved the profile.

Now, certain parts of your profile, including your current city, hometown, education and work, and likes and interests, will contain "connections." Instead of just boring text, these connections are actually Pages, so your profile will become immediately more connected to the places, things and experiences that matter to you.[18]

During 2009 and 2010, Facebook moved a large number of profile elements to the new status of a Page. Pages are Facebook's official representation of an entity—such as a band, restaurant, or TV show. Using the *Like* button users can link the page to one's profile. However, there was a catch:

> Keep in mind that Facebook Pages you connect to are public. You can control which friends are able to see connections listed on your profile, but you may still show up on Pages you're connected to. If you don't want to show up on those Pages, simply disconnect from them by clicking the "Unlike" link in the bottom left column of the Page. You always decide what connections to make.[19]

By default, connections to pages were public. In 2010, when individuals stepped through this new Facebook privacy interface (Fig. 6), items that were previously private became public—consistent with experimental evidence of default settings shaping privacy outcomes [5] (the interface contains options such as "Link All" and "Ask Me Later," but no "Do Not Link," option; the option to "Choose Pages individually" is provided, but in a smaller font). As noted by Facebook, in addition to favorite bands,

---

[18]See "Connecting to Everything You Care About," Facebook Blog: `https://blog.facebook.com/blog.php?post=382978412130`, last accessed on February 25, 2013.

[19]See "Connecting to Everything You Care About," Facebook Blog: `https://blog.facebook.com/blog.php?post=382978412130`, last accessed on February 25, 2013.
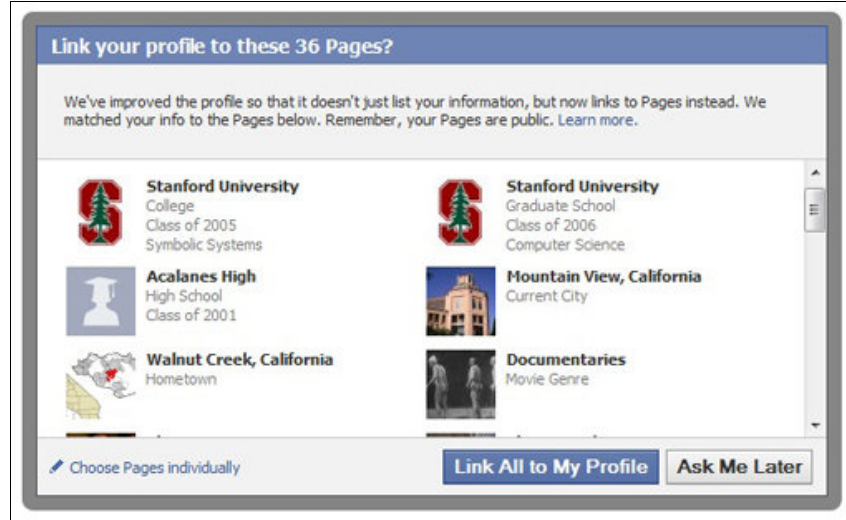
Figure 6: Facebook Page Opt-in Wizard

books, and movies, the profile elements "current city, hometown, education and work, and likes and interests" now were connected as pages. Reflecting back on the items that comprise our reversal—high school, hometown, address, interests, and their favorite movies, books, and music—we see a complete matchup between this information and the information converted to Pages by Facebook in 2010. For this reason, we conclude that the trend reversal that occurred in our dataset between October 4, 2009 and November 12, 2010 was not indicative of a popular trend to share more information publicly, but more likely the result of the December 9, 2009 and April 19, 2010 policy and interface changes on a largely unsuspecting public. By the time our data collection ended (May 5, 2011), with few exceptions (such as address information) disclosures of most types of personal information had *not* gone back down to the levels they had reached before Facebook's changes.

5   The Silent Listeners

So far, we have observed that Facebook users in our dataset became increasingly protective of their personal information over time, limiting the access to it they provided to unconnected profiles. We also observed that policy and interface changes implemented by Facebook near the end of the period of time under our observation arrested or even inverted that trend for certain profile elements. In this section, we discuss how *private* disclosures of Facebook users (that is, information selectively disclosed to subsets of users on the CMU network, such as friends or friends of friends) evolved during the same period of time.

Unlike the analysis we presented in Section 4, which was primarily quantitative, the analysis in this section is mainly qualitative, as it focuses on structural changes to the Facebook network during the period under our observation. However, below we show that the findings of the qualitative investigation are consistent with data points from a variety of sources. Furthermore, while the analysis in the previous sections was based on a novel dataset, here we rely on a combination of existing data and research; however, we are able to revisit that extant research under the lights of what we learned from the analysis of the CMU data. In doing so, we connect trends in public disclosures to the seemingly contrasting trends in private disclosures. Specifically, we observe that, while decreasing their *public* disclosures, Facebook users at the same time likely increased their *private* disclosures to connected Facebook friends, both in terms of scope and amount of personal data. By doing so, they increased their disclosures to other entities as well: third-party apps, advertisers, and Facebook itself. Whereas the first trend presented in Section 4 of this manuscript, considered in isolation, would paint a picture of Facebook users as increasingly willing and able to take the protection of privacy in their own hands, the combination of these diverse disclosure trends highlights the challenges associated with effectively managing one's online disclosures.

The trend towards decreasing disclosures we highlighted in the previous sections was based on the subset of Facebook profile fields which existed in 2005 (Table 2), and was limited to what members of the CMU Facebook network chose to make *publicly* available to all other members of the same network. However, during the period of time under our observation, two significant system changes occurred: first, Facebook profile fields and the user interface kept evolving; second, modifications to privacy controls allowed Facebook members to privately disclose information to subsets of other Facebook users (rather than to the entire network). There are at least four reasons to believe that, as a result, Facebook users increased their private disclosures over time.

First, the number of profile fields available to Facebook members to share information expanded over time. Compare Table 5, which shows the list of fields of a Facebook profile in 2005, to Tables 6, 7, and 8, which list the fields at the time of writing (2012). The number of fields in 2012 is over three times the number in 2005.

Second: certain fields started being used dynamically, for repeated sharing: for instance, status updates in 2006 and the Timeline in 2011. These fields enable users to share new information frequently (often multiple times a day) as opposed to fields for sharing static information that never changes (e.g., hometown), or fields for sharing infrequently changing information (e.g., favorite music). This trend reflects Facebook profiles' transition from static representations of a member's personal information (his/her gender) or personally identifying information (name), to 'habitats' through which new information is frequently created (Likes, status updates, comments and messages to other users, places visited, events attended, and so forth) by virtue of interacting with others (users, companies, sites) through the network. This trends also exemplifies the transformation in the type and quality of personal information explicitly or implicitly disclosed by the user about herself.[20]

---

[20]See, also, the "taxonomy" of online social networking data proposed in [52].

Third, more diverse user data started being generated by third-party apps (which were introduced in 2007). Often, an app "creates" additional data points about a user (see Table 4, which includes a selection of currently popular Facebook apps, and the information they elicit or generate). These data can be posted by the app on the user's profile and transmitted to developers. For instance, the Spotify app shows songs the user is listening to on the user's profile, and the TripAdvisor app shows information about the user's trips. As of October 2012, the usage base for the apps listed in the table ranged from 55 million (ChefVille) to 3.9 million (Four Square) Facebook users (based on estimates reported at `Appdata.com`).

Fourth, friends connected to a Facebook user started being able to add information about that user: for instance, by tagging individuals in photos in 2006, and by tagging their location in 2010. Schneier [52] refers to this as "incidental" data: "what other people post about you."

The availability of additional and more dynamic data fields illustrates that the scope for *potential* disclosures increased over time, but offers no proof that Facebook users actually started revealing more, or more diverse, personal information over time to their connected friends.[21] However, a variety of data points support the conclusion that Facebook private disclosures did increase over time—in scope and amount.

First, survey data reported by [29] suggests that, on an average day in 2010, 15% of US adult Facebook users updated their own status, 22% commented on another's post or status, 20% commented on another user's photos, and 26% "Liked" another user's content (in fact, 20% of women and 9% of men "Like" multiple times daily). Whereas in 2005 the overwhelming majority of fields available to Facebook users consisted of static or rarely updatable information (see Table 5), by 2010 a significant proportion of Facebook users would generate new information daily.

Second, qualitative inferences from profile data recently made available by a small set of Facebook users on the "Europe vs. Facebook" site are consistent with the above data points. While far from being statistically significant or generalizable, the documents point at profiles with much vaster amounts of disclosed information in 2011 than the profiles in the 2005 snapshot of the CMU network.[22]

Third, and perhaps more importantly: in 2011, Facebook disclosed that the amount of information shared by the average Facebook user was, in fact, doubling every year [58]—an exponential rate of growth.

Considering that Facebook users in our dataset moved towards lower *public* disclosures over time across *all* data categories we analyzed, and that—as noted in Section 4.5—similar trends have been reported in shorter field studies or survey studies for other data fields as well,[23] it is reasonable to infer that the intended recipients of the *increased*

---

[21]For instance, one year after the introduction of "frictionless sharing" (which allows sharing of personal information without direct user intervention), a Facebook manager was quoted as stating that users' reaction to this passive form of sharing had not been "strong" [26].

[22]See `http://europe-v-facebook.org/EN/Data_Pool/data_pool.html`, last accessed on February 25, 2013.

disclosures described in this section were not the general public, but other selected Facebook users (for instance, the user's friends, and friends of friends). In other words, the increased disclosures were likely and primarily intended as *private* disclosures. However, such increased private disclosures ended up reaching entities other than a user's friends.

The first such entity is, of course, Facebook itself. In addition to all the fields listed in Tables 6, 7, and 8, the provider of the network also maintains access to users' less visible information, such as logins or visited profiles.[24] That information may later be made available to entities Facebook chooses, or is compelled, to share data with—such as law enforcement [13].

The second entity, or set of entities, is represented by third-party apps. By default, at present Facebook apps get access to a user's "basic information"—that is, name, profile picture, cover photo, networks, username, user ID, gender information, as well as any other information the user has shared publicly on their profiles. In addition, an app installed by a Facebook user can access the fields of *another* Facebook member's profile connected to that user (for instance, Family and Relationships, Interested in, or Religious and Political Views). Furthermore, apps can request access to additional fields of data (the user cannot opt-out; they can simply decide not to use the app). Wang et al. recently analyzed the information actually accessed by over 9,000 popular Facebook apps [60]. By comparing each app's popularity with the number of data permission requests issued by the app, the authors concluded that "[Facebook] users shared their basic information more than 500 million times with apps." In addition, Facebook users revealed to third-party apps more sensitive information (including a profile element for which this manuscript uncovered a trend of decreasing public disclosure within the CMU dataset): birthday (138 million permissions requested), user location (55 million permissions requested), user photos (nearly 25 million permissions requested), or the location of the user's *friends* (8 million permissions requested). Research has consistently shown that such user information is often harvested by Facebook apps without individuals' precise knowledge or even awareness [7, 34, 60].

Third, information from private fields can also be used by Facebook advertisers. Although advertisers do not have direct access to the data, they can microtarget their ads based on the information provided in profile fields, including private ones.[25] For instance, [51] used targeted ads to display an invitation to participate in a scientific study to "women aged 18 to 45 years with an IP address in Italy whom [were] selected by applying 2 Italian keywords (in English: 'pregnancy' and 'delivery') to 'likes' and

---

[23]Both self-report surveys [42, 48] and recent field data gathered by [16] for the NYC Facebook network suggest that the trend towards lower public disclosure was not unique to the fields we analyzed in Table 2, but has extended to other fields (such as friends lists), or the complete profile.

[24]Schneier [52] refers to service data (the data provided to a social networking site in order to use it), behavioral data (data the site collects "about your habits by recording what you do and who you do it with"), and derived data (which can be inferred about a user based on the rest of the available data). For a sample of the data stored by Facebook about a user, see, again, "Europe vs. Facebook" site at `http://europe-v-facebook.org/EN/Data_Pool/data_pool.html`, last accessed on February 25, 2013.

'interests' listed in Facebook profiles."[26]

In short, two contrasting dynamics emerged from our analysis. Over time, Facebook members in our panel disclosed less information publicly. However, and in parallel, Facebook users seemed willing to disclose more and more diverse information privately, to friends. In doing so, however, they disclosed more to other entities as well (Facebook, third-party apps, and indirectly advertisers)—often without awareness or explicit consent [7, 34, 60].[27] In addition to misconceptions about the access third-party apps gain to profile data, the literature has also documented users' errors in estimating how many other Facebook members can have access to their profile data (see [2] and, more recently, [32] and [6]). In particular, [6] report that "social media users consistently underestimate their audience size for their posts, guessing that their audience is just 27% of its true size." Collectively, these results echo early depictions of social network sites as "imagined" communities [2] where actual audiences are not necessarily the intended, or expected, audiences.

The increased disclosures to these "silent listeners" may be linked to a different set of factors than those that helped us explain the spike in public disclosure in the CMU dataset between 2009 and 2010. Our previous analysis linked the latter spike to asymmetric information and default settings. The growth in disclosures to first (Facebook) and third-parties (apps, advertises) may reflect more subtle dynamics at play. While we do not attempt to provide conclusive causal claims, we observe that those dynamics are consistent with recent experimental evidence in the field of privacy decision making. Access to increasingly granular settings, to determine which profile data other Facebook users get to peruse, may have increased a member's feeling of control and direct her attention towards the sharing taking place with other active, non silent, members of the network; in turn, perceptions of control over personal data [11] and misdirection of users' attention [4] have been linked to increases in disclosures of sensitive information to strangers. Therefore, inasmuch as the results we have presented in Section 4 can be interpreted as suggesting that Facebook users became increasingly willing and able to protect their personal data over time, and chose to reveal it to fewer strangers, the evidence presented in this section raises questions over the degree of awareness and intentionality under which Facebook users provided increasing amounts of personal information to the silent listeners in the network.

---

[25]Facebook help center notes: "For example, if you are a children's toy store located in Springfield you may want to target people who live in your city who are parents." See "What are the benefits of choosing a more targeted audience versus a broader audience?," Facebook Help Center: http://www.facebook.com/help/121933141221852/, last accessed on February 25, 2013.

[26]In addition, data leakages are possible. Krishnamurthy and Wills found in 2009 that some third-party apps were leaking users' unique identifiers for sharing with third-party aggregators and advertising partners [36]. Korolova found in 2010 that the microtargeting capabilities of Facebook advertising system could be exploited for privacy violations [35].

[27]Also see, again, the Federal Trade Commission's 2011 complaint *In the Matter of FACEBOOK, INC.*, at http://ftc.gov/os/caselist/0923184/111129facebookcmpt.pdf, last accessed on February 25, 2013. See also EPIC's FTC Complaint in re Facebook (filed December 17, 2009), at http://www.epic.org/privacy/inrefacebook/EPIC-FacebookComplaint.pdf, last accessed on February 28.

Table 4: Popular Facebook apps and the information they generate about a user (Sources: appdata.com, Wall Street Journal, and individual websites for the listed apps)

AppName	Type	Data
ChefVille	Game	Food preferences
TripAdvisor	Travel	Trip recs./history/checkins
Yahoo! Social Bar	News/Social Reader	Yahoo activity reported
Instagram	Photo	Photo sharing/check-ins (FB)
Microsoft Live	Utilities/Communication	Social/search engine
Bing	Utilities	Social/search engine
Spotify	Music/Entertainment	Music choices
Scribd	Utilities	Interests for reading/publishing
SchoolFeed	Online Communications	Connects users to others
Between You and Me	Dating	Dating
MyPad for iPad	FB for iPad	Recreates FB for an iPad
Skype	Utilities	Communications/networking
FourSquare	Utilities	Aggregates Check-ins

6   Limitations and Conclusions

We have presented the results of a longitudinal analysis of 5,076 Facebook users who were members of the Carnegie Mellon Facebook network in 2005. Over the course of seven years, we captured profile content with a goal of understanding how disclosure practices change over time. Before summarizing our results, we point out a number of limitations of the current analysis.

First, as we have observed in the introduction, one limitation of this data is that it does not originate from a random sample of Facebook users—nor could it, as the bulk of Facebook users in 2013 did not have Facebook accounts in 2005. Our trends are based on a panel of Facebook users dominated by undergraduate students, and our analysis focuses on one specific Facebook network—the Carnegie Mellon network—and only on those who were members of that network in 2005. Hence, our results may not extrapolate to more diverse samples of users. However, both survey data [42, 48] and analyses of other Facebook networks [16] are consistent with one of the results presented here: that over time Facebook users have become less likely to share their personal information publicly. Our analysis extends that research by offering evidence that the privacy-seeking behavior started early in the life of the network, and then progressed over several years of Facebook usage—until it was partly obstructed by Facebook's policies and interface changes.

Second, in this manuscript we did not distinguish between non-disclosures due to the member actually not filling out a field, and those due to the member altering the privacy settings of that field (or of their entire profile) in order to limit public disclosures. While such distinction is of interest (and is the object of our ongoing research), it does not

invalidate the main conclusions of the current analysis, which focused on the trends in *public* disclosures of personal information over time. It does affect, however, the discussion of how much information remains available to third-parties (such as apps providers) and to Facebook itself.

Third, our quantitative analysis was restricted to the fields which existed on Facebook in 2005, and the analysis presented in Section 5 was mainly qualitative, and included only a preliminary investigation of additional fields. However, using a consistent set of fields, and a consistent set of users, allowed us to more precisely define and explain trends in disclosure and privacy behavior over the past seven years.

As our analysis revealed, a robust trend of declining public disclosure emerged over the years across a broad range of Facebook profile elements—including personal, contact, and interest information. We also observed a significant shift for many of these profile elements between the years 2009 and 2010, when public disclosure increased. We concluded that changes to privacy policy and interface settings by Facebook produced greater public disclosures. In other words, exogenous changes effected by Facebook near the end of the period of time under our observation arrested or inverted an endogenous, user-driven trend of members trying to protect their privacy by managing the public disclosure of their personal information.

On the other hand, we also observed that, over time, the amount and scope of personal information that Facebook users have revealed to friends' profiles seems to have markedly increased—and thus, so have disclosures to Facebook itself, third-party apps, and (indirectly) advertisers. Such findings highlight the challenges users of social network sites face when trying to manage online privacy, and the power of providers of social media services to affect individuals' disclosure and privacy behavior through interfaces and default settings.

Acknowledgments

Table 5: Facebook Fields in 2005 (Sources: snapshots of Facebook profiles; Everything2.com)

Gender
Birthday
Hometown
State
Zip
High School
School
Status at School
Year
Concentration
Secondary concentration
Residence
Room in dorm
Mailbox at school
School email
Preferred email
AIM screen name
Phone number
Other phone
Address
Website
Interested in (gender)
Interested in meeting
Relationship status
In relationship with
Political views
Interests
Clubs and jobs
Favorite music
Favorite book
Favorite movies
Favorite quotes
About you
List of courses
Profile picture
My pictures
Groups
Friends
Messages

Table 6: Facebook Fields in 2012, Part A (Sources: snapshots of Facebook profiles; Everything2.com)

Living: Current City
Living: Hometown
Relationships: Relationship Status
Relationships: With
Relationships: Anniversary
About You: Description
Basic Info: Gender
Basic Info: Birthday
Basic Info: Interested in
Basic Info: Languages
Basic Info: Religion
Basic Info: Rel. description
Basic Info: Political Views
Basic Info: Pol. Description
Contact Info: Preferred email
Contact Info: FB email
Contact Info: Mobile Phone
Contact Info: Other Phones
Contact Info: IM Screen Names
Contact Info: Address
Contact Info: City
Contact Info: Zip
Contact Info: Neighborhood
Contact Info: Website
Contact Info: Networks
Favorite Quotation: Quotation
Photos: Profile (historical)
Photos: Cover Photo
Photos: Mobile Uploads
Photos: Albums
Photos: Videos
Maps/places: Hometown
Maps/places: Places Lived
Maps/places: Business
Maps/places: Public spaces e.g. parks
Maps/places: Life Events
Maps/places: Photos

Table 7: Facebook Fields in 2012, Part B (Sources: snapshots of Facebook profiles; Everything2.com)

Likes: Music
Likes: Movies
Likes: Television
Likes: Athletes
Likes: Sports Teams
Likes: Events
Likes: Interests
Likes: Business
Likes: Promotions
Likes: Items/Merchandise
Likes: People
Likes: Establishments
Applications: Recent Use
Notes: Posts
Notes: Linked Blog
Subscriptions: People
Subscribers: People who subscribe to user
Friends: Aggregate
Friends: Lists
Friends: Requests
Timeline: Activity e.g. updates
Timeline: Comments
Timeline: Tags
Timeline: Likes
Timeline: Shares
Timeline: Places
Timeline: Friends' Posts
Timeline: Highlights
Timeline: All Stories (all activity)
Work and Education: Colleges: Time Period
Work and Education: Colleges: Graduated
Work and Education: Colleges: Description
Work and Education: Colleges: Concentrations
Work and Education: Colleges: College or Graduate School
Work and Education: Colleges: Classes
Work and Education: High School: Location
Work and Education: High School: Time Period
Work and Education: High School: Graduated
Work and Education: High School: Description
Work and Education: Work Places: Position
Work and Education: Work Places: Location
Work and Education: Work Places: Description
Work and Education: Work Places: Time Period

Table 8: Facebook Fields in 2012, Part C (Sources: snapshots of Facebook profiles; Everything2.com)

Life Events: Work & Education: New Job
Life Events: Work & Education: Published Book or Paper
Life Events: Work & Education: Retirement
Life Events: Work & Education: New School
Life Events: Work & Education: Study Abroad
Life Events: Work & Education: Volunteer Work
Life Events: Work & Education: Military Service
Life Events: Work & Education: Other Life Event
Life Events: Family & Relationships: First Met
Life Events: Family & Relationships: New Relationship
Life Events: Family & Relationships: Engagement
Life Events: Family & Relationships: Marriage
Life Events: Family & Relationships: Expecting a Baby
Life Events: Family & Relationships: New Child
Life Events: Family & Relationships: New Family Member
Life Events: Family & Relationships: New Pet
Life Events: Family & Relationships: End of Relationship
Life Events: Family & Relationships: Loss of a Loved One
Life Events: Family & Relationships: Other Life Event
Life Events: Home & Living: Moved
Life Events: Home & Living: Bought a Home
Life Events: Home & Living: Home Improvement
Life Events: Home & Living: New Roommate
Life Events: Home & Living: New Vehicle
Life Events: Home & Living: Other Life Event
Life Events: Health & Wellness: Organ Donor
Life Events: Health & Wellness: Overcame an Illness
Life Events: Health & Wellness: Quit a Habit
Life Events: Health & Wellness: New Eating Habits
Life Events: Health & Wellness: Weight Loss
Life Events: Health & Wellness: Glasses, contacts, Other
Life Events: Health & Wellness: Broken Bone
Life Events: Health & Wellness: Removed Braces
Life Events: Health & Wellness: Other Life Event
Life Events: Travel & Experiences: New Hobby
Life Events: Travel & Experiences: New Instrument
Life Events: Travel & Experiences: New Language
Life Events: Travel & Experiences: Tattoo or Piercing
Life Events: Travel & Experiences: New License
Life Events: Travel & Experiences: Travel
Life Events: Travel & Experiences: Achievement or Award
Life Events: Travel & Experiences: Changed Beliefs
Life Events: Travel & Experiences: First Word, Kiss, Other
Life Events: Travel & Experiences: New Sport
Life Events: Travel & Experiences: Registered to Vote
Life Events: Travel & Experiences: Other Life Event

References

[1] Acquisti, A. (2004). Privacy in electronic commerce and the economics of immediate gratification. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, 21–29.

[2] Acquisti, A. and Gross, R. (2006). Imagined communities: Awareness, information sharing, and privacy on The Facebook. In G. Danezis and P. Golle (eds.), 6th *Workshop on Privacy Enhancing Technologies (PET '06)*, vol. 4258 of *LNCS*. Cambridge, UK: Springer-Verlag. 36–56.

[3] — (2009). Predicting social security numbers from public data. *Proceedings of the National Academy of Sciences*, 106(27):10975–10980. `http://www.pnas.org/content/106/27/10975.abstract`.

[4] Adjerid, I., Acquisti, A., and Brandimarte, L. (2012). Sleight of privacy: Disclosure, framing, and the limits of transparency. In *Workshop of Information Systems Economics (WISE)*.

[5] Acquisti, A., John, L., and Loewenstein, G. (2009). What is privacy worth? In *Workshop on Information Systems and Economics (WISE)*.

[6] Bernstein, M. S., Bakshy, E., Burke, M., and Karrer, B. (2013). Quantifying the invisible audience in social networks. In *ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2013)*. To appear.

[7] Besmer, A. and Lipford, H. R. (2010). Users' (mis)conceptions of social applications. In *Proceedings of Graphics Interface 2010*. Canadian Information Processing Society. 63–70.

[8] boyd, d. m. and Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1). `http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html`.

[9] boyd, d. m. and Hargittai, E. (2010). Facebook privacy settings: Who cares? *First Monday*, 15(8).

[10] boyd, d. m. and Heer, J. (2006). Profiles as conversation: Networked identity performance on friendster. In *Proceedings of the Hawaii International Conference on Systems Sciences (HICSS '06)*. Kauai, HI: IEEE Computer Society.

[11] Brandimarte, L., Acquisti, A., and Loewenstein, G. (2012). Misplaced confidences: Privacy and the control paradox. *Social Psychological and Personality Science*.

[12] Bumgarner, B. A. (2007). You have been poked: Exploring the uses and gratifications of Facebook among emerging adults. *First Monday*, 12(11). `http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2026/1897`.

[13] Carioli, C. (2012). When the cops subpoena your Facebook information, here's what Facebook sends the cops. *Boston Phoenix—The Phlog*, April 6.

[14] Caverlee, J. and Webb, S. (2008). A large-scale study of Myspace: Observations and implications for online social networks. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM 2008)*, vol. 8. Seattle, WA.

[15] Debatin, B., Lovejoy, J., Horn, A., and Hughes, B. (2009). Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *Journal of Computer-Mediated Communication*, 15(1):83–108.

[16] Dey, R., Jelveh, Z., and Ross, K. (2012). Facebook users have become much more private: A large-scale study. In *Proceedings of the 31st Annual International Conference on Computer Communications (IEEE Infocom 2012)*. Orlando, FL.

[17] DiMicco, J. M. and Millen, D. R. (2007). Identity management: Multiple presentations of self in Facebook. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work (GROUP '07)*. New York: ACM. 383–386.

[18] Donath, J. S. and boyd, d. m. (2004). Public displays of connection. *BT Technology Journal*, 22(4):71–82.

[19] Dwyer, C., Hiltz, S. R., and Passerini, K. (2007). Trust and privacy concern within social networking sites: A comparison of Facebook and Myspace. In *Proceedings of the 13th Americas Conference on Information Systems (AMCIS '07)*. Keystone, Colorado: Association for Information Systems. 339–350.

[20] Ellison, N. B., Steinfield, C., and Lampe, C. (2007). The benefits of Facebook 'friends:' Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4). `http://jcmc.indiana.edu/vol12/issue4/ellison.html`.

[21] Farnham, S. D. and Churchill, E. F. (2011). Faceted identity, faceted lives: Social and technical issues with being yourself online. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work (CSCW '11)*. New York: ACM. 359–368. `http://doi.acm.org/10.1145/1958824.1958880`.

[22] Fogel, J. and Nehmad, E. (2009). Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in Human Behavior*, 25(1): 153–160. `http://www.sciencedirect.com/science/article/B6VDC-4TDYNXD-1/2/d7b3cd73dd996e428849e3e6752ed1e2`.

[23] Gjoka, M., Kurant, M., Butts, C. T., and Markopoulou, A. (2010). A walk in Facebook: Uniform sampling of users in online social networks. In *Proceedings of the 29th Conference on Computer Communications (IEEE INFOCOM 2010)*. San Diego, CA: IEEE Computer Society. 1–14.

[24] Gjoka, M., Sirivianos, M., Markopoulou, A., and Yang, X. (2008). Poking Facebook: Characterization of OSN applications. In *Proceedings of ACM SIGCOMM Workshop on Online Social Networks (WOSN '08)*. Seattle, WA.

[25] Goffman, E. (1959). *The Presentation of Self in Everyday Life*. New York: Anchor Books.

[26] Greenfield, R. (2012). Frictionless sharing hits the skids at Facebook. *The Atlantic Wire*, September 21.

[27] Grimmelmann, J. T. (2009). Facebook and the social dynamics of privacy. *Iowa Law Review*, 95(4):1–52.

[28] Gross, R. and Acquisti, A. (2005). Information revelation and privacy in online social networks. In *Proceedings of Workshop on Privacy in the Electronic Society (WPES 2005)*.

[29] Hampton, K., Goulet, L. S., Rainie, L., and Purcell, K. (2011). Social Networking Sites and Our Lives. Technical report, Pew Internet and American Life Project. `http://www.pewinternet.org/Reports/2011/Technology-and-social-networks.aspx`.

[30] Hoadley, C. M., Xu, H., Lee, J. J., and Rosson, M. B. (2010). Privacy as information access and illusory control: The case of the Facebook news feed privacy outcry. *Electronic Commerce Research and Applications*, 9(1):50–60. `http://www.sciencedirect.com/science/article/B6X4K-4W85MD0-1/2/b4c518bb554d998aa61320944e40ca94`.

[31] Hull, G., Lipford, H., and Latulipe, C. (2010). Contextual gaps: Privacy issues on Facebook. *Ethics and Information Technology*, 1–14. `http://dx.doi.org/10.1007/s10676-010-9224-8`.

[32] Johnson, M., Egelman, S., and Bellovin, S. M. (2012). Facebook and privacy: it's complicated. In *Proceedings of the 8th Symposium on Usable Privacy and Security*, 9. ACM.

[33] Joinson, A. N. (2008). Looking at, looking up or keeping up with people?: Motives and use of Facebook. In *Proceeding of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. New York: ACM. 1027–1036.

[34] King, J., Lampinen, A., and Smolen, A. (2011). Privacy: is there an app for that? In *Proceedings of the 7th Symposium on Usable Privacy and Security*. New York: ACM. 12:1–12:20.

[35] Korolova, A. (2011). Privacy violations using microtargeted ads: A case study. *Journal of Privacy and Confidentiality*, 3(1):27–49.

[36] Krishnamurthy, B. and Wills, C. E. (2009). On the leakage of personally identifiable information via online social networks. In *Proceedings of the 2nd ACM workshop on Online Social Networks*. ACM. 7–12.

[37] Lampe, C., Ellison, N. B., and Steinfield, C. (2006). A face(book) in the crowd: Social searching vs. social browsing. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work (CSCW '06)*. New York: ACM. 167–170.

[38] — (2007). A familiar face (book): Profile elements as signals in an online social network. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. New York: ACM. 435–444.

[39] — (2008). Changes in use and perception of Facebook. In *Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work (CSCW '08)*. New York: ACM. 721–730.

[40] Lampinen, A., Lehtinen, V., Lehmuskallio, A., and Tamminen, S. (2011). We're in it together: Interpersonal management of disclosure in social network services. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems (CHI '11)*. New York: ACM. 3217–3226. `http://doi.acm.org/10.1145/1978942.1979420`.

[41] Lampinen, A., Tamminen, S., and Oulasvirta, A. (2009). All my people right here, right now: Management of group co-presence on a social networking site. In *Proceedings of the ACM 2009 International Conference on Supporting Group Work (GROUP '09)*. New York: ACM. 281–290.

[42] Lenhart, A. and Madden, M. (2007). Teens, Privacy & Online Social Networks. Technical report, Pew Internet and American Life Project. `http://www.pewinternet.org/Reports/2007/Teens-Privacy-and-Online-Social-Networks.aspx`.

[43] — (2007). Teens, Privacy and Online Social Networks: How Teens Manage Their Online Identities and Personal Information in the Age of Myspace. Technical report, Pew Internet and American Life Project. `http://www.pewtrusts.org/uploadedFiles/wwwpewtrustsorg/Reports/Society_and_the_Internet/PIP_Teens_Privacy_SNS_Report_Final.pdf`.

[44] Lenhart, A., Purcell, K., Smith, A., and Zickuhr, K. (2010). Social Media and Young Adults. Technical report, Pew Internet and American Life Project. `http://pewinternet.org/Reports/2010/Social-Media-and-Young-Adults.aspx`.

[45] Lewis, K. (2011). The co-evolution of social network ties and online privacy behavior. *Privacy Online: Perspectives on Privacy and Self-Disclosure in the Social Web*, 91–109.

[46] Lewis, K., Kaufman, J., and Christakis, N. (2008). The taste for privacy: An analysis of college student privacy settings in an online social network. *Journal of Computer-Mediated Communication*, 14(1):79–100.

[47] Liu, H. (2007). Social network profiles as taste performances. *Journal of Computer-Mediated Communication*, 13(1). `http://jcmc.indiana.edu/vol13/issue1/liu.html`.

[48] Madden, M. (2012). Privacy Management on Social Media Sites. Technical report, Pew Internet and American Life Project. `http://www.pewinternet.org/~/media//Files/Reports/2012/PIP_Privacy_management_on_social_media_sites_022412.pdf`.

[49] Madejski, M., Johnson, M., and Bellovin, S. M. (2012). A study of privacy settings errors in an online social network. In *Proceedings of the 2012 Inaugural Structural Engineering Biennial Conference, "From Theory to Practice" (SESOC 2012)*.

[50] Marwick, A. E. and boyd, d. m. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1):114–133.

[51] Richiardi, L., Pivetta, E., and Merletti, F. (2012). Recruiting study participants through Facebook. *Epidemiology*, 23(1):175.

[52] Schneier, B. (2010). A taxonomy of social networking data. *Security & Privacy, IEEE*, 8(4):88–88.

[53] Skeels, M. M. and Grudin, J. (2009). When social networks cross boundaries: A case study of workplace use of Facebook and Linkedin. In *Proceedings of the ACM 2009 International Conference on Supporting Group Work (GROUP '09)*. New York: ACM. 95–104.

[54] Stutzman, F. D. (2006). An evaluation of identity-sharing behavior in social network communities. *International Digital Media and Arts Journal*, 3(1):10–18.

[55] Stutzman, F. D. and Kramer-Duffield, J. (2010). Friends only: Examining a privacy-enhancing behavior in Facebook. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems (CHI '10)*. New York: ACM. 1553–1562. http://doi.acm.org/10.1145/1753326.1753559.

[56] Sweeney, L. (1997). Guaranteeing anonymity when sharing medical data, the Datafly system. In *Proceedings of the AMIA Annual Fall Symposium*. American Medical Informatics Association. 51–55.

[57] Thelwall, M. (2008). Social networks, gender and friending: An analysis of Myspace member profiles. *Journal of the American Society for Information Science and Technology*, 59(8):1321–1330.

[58] Tsotsis, A. (2011). Mark Zuckerberg explains his law of social sharing. *TechCrunch*, July 6.

[59] Tufekci, Z. (2008). Can you see me now? Audience and disclosure regulation in online social network sites. *Bulletin of Science Technology and Society*, 28(1):20–36. http://bst.sagepub.com/cgi/content/abstract/28/1/20.

[60] Wang, N., Grossklags, J., and Xu, H. (2013). An online experiment of privacy authorization dialogues for social applications. In *Proceedings of the 16th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW 2013)*. San Antonio, TX.