How much is too much? Privacy issues on Twitter

Lee Humphreys, Phillipa Gill, and Balachander Krishnamurthy

Key Words: privacy, social media, Twitter, content analysis

Authors' Notes:

Lee Humphreys, Ph.D. is an Assistant Professor of Communication at Cornell University. Phillipa Gill is a Ph.D. student in Computer Science at the University of Toronto. Balachander Krishnamurthy, Ph.D. is a Research Scientist at AT&T Labs Research. Please direct all correspondences to Lee Humphreys, Dept. of Communication, Cornell University, 305 Kennedy Hall, Ithaca, NY 14853 or lmh13@cornell.edu

Abstract

Social media provide many opportunities to connect people; however, the kinds of personally identifiable information that people share through social media is understudied. Such public discussions of personal information warrant a closer privacy discussion. This paper presents findings from a content analysis of Twitter in which the amount and kinds of personally identifiable information in Twitter messages were coded. Findings suggest that the majority of time Twitterers do write about themselves. Overwhelmingly, Twitterers do not include identifiable information such as phone numbers, email and home addresses. However, about a quarter of tweets do include information regarding when people are engaging in activities and where they are. This kind of information may have privacy implications when found in the same tweet or if coupled with other kinds of publicly available information.

How much is too much? Privacy issues on Twitter

Social media provide many people a new way to connect with friends, family and colleagues. In particular, social network sites are frequently used to communicate with people known to one another through offline connections (Ellison, Steinfield, & Lampe, 2007). For example, as of August 2009, Facebook was the fifth most frequented website in the US (ComScore, 2009). These services can help to reinforce social bonds and manage social identities (d. m. boyd, 2004; Lange, 2007; Liu, 2007). Research has shown that there can be benefits that come from sharing personal information in social and public ways (e.g. boyd, 2004; Ellison et al., 2007; Hampton & Wellman, 1999).

In addition to the benefits of using social network sites, there may be risks associated with using such services. For example, research has begun exploring what kinds of personally identifiable information (e.g. phone numbers, email address, postal address, social security numbers, etc.) people share through services such as Facebook and MySpace (Kolek & Saunders, 2008; Lenhart & Madden, 2007). The misuse of personally identifiable information obtained online can raise many privacy concerns such as identity theft or even discrimination (Lyon, 2001). Therefore this study seeks to explore the kinds of personally identifiable information that people publically share by analyzing the content of a representative sample of public Twitter messages. Twitter is a popular micro-blogging and social network service that allows people to share messages of 140 characters in length. As of September 2009, Twitter had over 50 million unique users (Moore, 2009). While Twitter allows people to share information among friends or "followers", the default privacy setting on Twitter is that all messages are public, that is, anyone who signs up for Twitter may see them. In addition, all public tweets may be posted to a public timeline website which showcases the twenty most recent tweets. Profiles

on Twitter are relatively short compared to Facebook, therefore the bulk of the information about a person is communicated through their Twitter messages or tweets. This study explores the kinds of personally identifiable information that public tweets disclose.

Beyond personally identifiable information, sharing other kinds of personal information on Twitter may put people at risk to be taken advantage of. For example, in June 2009 Israel Hyman, an Arizona-based video podcaster, tweeted that he was looking forward to his family vacation to Saint Louis where they would be visiting family friends for the week. He tweeted again when they had successfully arrived in Missouri. While they were away, their house was broken into and several thousand dollars of computer and video equipment were stolen (Van Grove, 2009). According to one news report, Hyman said, "We don't know for sure if that's what caused the break it in, but it sure gives you pause to think about what you're publicly going to broadcast on the internet," ("Man Robbed After Posting His Vacation On Twitter", 2009). While this may have been an isolated event, it does raise questions about who has access to personal information and how that might put people at risk (Mills, 2009).

Concerns about sharing information regarding where people are and when are not necessarily a new phenomenon. People have often tried to keep the fact that they are on vacation discreet from potential vandals or thieves, whether it be through cancelling their mail or newspaper service or even getting a house sitter. Social media, however, allow people to share their locations with thousands of people with the click of a button. Such broadcastability may have important safety implications. There are offline examples of broadcasting personal time and location information and the risks associated with it. For example, funeral notices in newspapers can broadcast where and when family members will be and there have been examples of people’s homes being broken into while they are at funeral services (Wolfe, 1992). Most funeral

announcements request that flowers and cards be sent to the funeral home rather than the home of the family to avoid broadcasting the family's home address. These examples suggest that personally identifiable information is not the only kind of personal information shared that can have privacy implications. Incidental information such as when and where people may be can also have privacy implications. Time and location may constitute a second tier of personally identifiable information, which while seemingly mundane and minor can raise potential safety concerns when publically broadcasted and shared.

Prominence of Twitter

Twitter is one of the fastest growing social network sites on the web today, with 8 million users joining monthly (Moore, 2009). Twitter is most frequently used by young adults. Twenty-five to 34 year olds make up the largest percentage of Twitter users (Lenhart & Fox, 2009). This differs somewhat from other social networking services. For example, Pew reported that median age of Twitterers is several years older than the median age of MySpace or Facebook users but younger than LinkedIn users (Lenhart & Fox, 2009). From its inception, Twitter was cross-platform, meaning that users could submit their messages via the web, instant messenger or SMS ("short messaging service" or text message). This may have contributed to the fact that Twitter users tend to be "more mobile in their communication and consumption of information" than the average internet user," (Lenhart & Fox, 2009, p. 3).

Previous studies of Twitter have explored the kinds of messages people post (Mischaud, 2007; Naaman, Boase, & Lai, 2010), the degree of interactivity within messages (d. boyd, Golder, & Lotan, 2010; Honeycutt & Herring, 2009), the network size of Twitterers and the frequency of tweets (Krishnamurthy, Gill, & Arlitt, 2008; Moore, 2009). Twitter ostensibly asks users, "What are you doing?", but research suggests that users do not always tweet about what

they are doing (Mischaud, 2007; Naaman et al., 2010). People use Twitter to share information about themselves as well as to share information publicly available elsewhere on the web, such as breaking news or interesting media such music, videos, blogs, etc. Honeycutt and Herring (2008) found that 41% Tweets in their sample were shared information about the author him or herself. Similarly Naaman, Boase, & Lai (2010) found that about half of Twitter messages were about the author him or herself while the rest were about other people or things. These studies suggest that Twitter users are not only talking about themselves directly; but even if just half of the messages are about themselves that still means that Twitter users are sharing 12 million tweets per day about themselves (Liew, 2009). Sometimes of course messages that do not directly reference the user can still share information about the user's tastes, interests, and preferences (Liu, 2007). Given the rise of GPS and mobile technologies which may encourage sharing of location information (Humphreys, 2007), it is important to take a step back and examine personally identifiable information as well as a second tier of identifiable information including when and where people are. This is the first study to the best of our knowledge that explores the kinds of personally identifiable information that people post on Twitter.

Social Media & Sharing

Much research has explored the ways people share information about themselves online and the privacy implications (see Joinson & Paine, 2007 for an overview). Time and again, research has shown that people will disclose more personal information online than they will face-to-face (Joinson & Paine, 2007). Not only do people readily self-disclose in online experimental settings, (e.g. (Tidwell & Walther, 2002), but they often also disclose personal identifiable information when this is requested by a website (Metzger, 2004). The personal

information revealed in Twitter messages, however, are at the complete discretion of users, so long as they conform to the 140-character limit.

While Twitter differs from social network sites like Facebook and MySpace in its format, it can be helpful to look privacy attitudes and behaviors on these sites in order to better situate this study. A study of the attitudes towards privacy and Facebook use by Acquisti & Gross (2006) found while privacy concerns predicted Facebook use for older people, it did not predict use for students, suggesting that even when young adults were concerned about privacy issues they were still likely to be active and contributing members of Facebook. Lennart and Madden (2007) found that as many as two-thirds of teens on social network sites report to have changed their profile settings so that they are not visible to the entire public. In addition, younger teens and females were likely to engage in privacy-protecting behaviors than were older teens and males (Lennart & Madden, 2007). Given that Twitter users tend to be older than Facebook users (Lennart & Fox, 2009), this may suggest that Twitters users may not take engage in as many privacy-protecting behaviors. We see evidence of this in recent reports about the lowering percentage of new Twitter users who change their privacy settings from public to protected when they first join (Moore, 2009). Older Facebook users may be less likely to change their privacy settings compared to younger users. Kolek & Saunders (2008) found that only 11% of their sample of college students using Facebook had restricted access to their profiles so that the researchers could not examine the content of them. It is unclear, however, whether the remaining 89% of the profiles were completely publically accessible or restricted to in-network members (which the researchers may have been part of if they sampled their own university). The later may have been partially true since Facebook's default privacy setting is based on the user's network affiliations (Facebook, 2008).

Surveillance & social media

The rise of social media more broadly brings about many issues with regards to privacy and surveillance. Privacy and surveillance are often presented as counter points when discussing issues of personal information and new technology. Privacy has been defined as the ability to control what information about oneself is available to others (Westin, 2003). When one cannot control what information about oneself others know, one may be open to surveillance by others. Lyon defines surveillance as "any collecting or processing of personal data, whether identifiable or not, for the purposes of influencing or managing those whose data have been gathered," (2001, p. 2). When it comes to social media, or all information technology for that matter, the challenge of privacy and surveillance are complicated because one does not know who exactly has access to your personal information or how your information may be used in unintended ways.

Privacy concerns can arise due to the digital storage of personal information. Alterman (2003) suggests that there are three kinds of concern associated with the storage of data. First is that someone "will legitimately gain access to information about you and utilize it to locate and harass or harm you in some manner" (Alterman, 2003, p. 140). The second kind of privacy concern is that information you gave up for one reason or purpose will be accessed or used for purposes that you could not have thought about or did not approve. The third kind of concern that Alterman (2003) suggests is that your data will be illegitimately or even illegally accessed and used which will then put your at risk for embarrassment or worse. While the third is unlikely to occur through the public timeline of Twitter, the first two privacy concerns, which Alterman suggests arise from the storage of data, are quite relevant on Twitter in that followers on Twitter could potentially use tweets to harm or harass you. Given that many Tweets are public, one may

not even have to be a Twitter follower to gain lawfully access to a person's tweets. In fact, recently Google & Bing announced that they would be including public postings from Facebook and Twitter in their search results (Chapman, 2009).

Inherent in social media is the dual activity of the production and consumption of social information. Users generate content for other users. There are recent examples of how social media both facilitate and rely on surveillance mechanisms (e.g. Albrechtslund, 2008; Andrejevic, 2007; Author own cite; Zimmer, 2008). Social media's participatory value is partially from the consumption of others' social information. Andrejevic (2005) argues lateral surveillance, the asymmetrical, nontransparent monitoring of citizens by one another, is also an important component of the networked society. With the advent of the internet and interactive media, people have similar technological capabilities previously held exclusively by corporate and state entities. As such, citizens can monitor other citizens' behavior through nonreciprocal forms of watching. Everyday people can search for information about other citizens without their knowledge or permission. For example, people may use Twitter assuming that only their "followers" will read their messages, when in fact, any message posted by a user who has not changed their default privacy settings may be accessed by anyone else.

Default settings can be very powerful in influencing behaviors online. Shah and Kesan (2003) suggest that default settings are one of the three key governance characteristics of code. They argue that there are two reasons why people do not change the default settings online. First they suggest that users may be uninformed that it is possible to change default settings or what the ramifications of the various settings may be. Second, Shah and Kesan (2003) suggest that people may not have the technological know-how to change their default settings. These two factors may be increasingly relevant on Twitter as its user base grows and may include less tech

savvy people. In fact, Moore (2009) found in January 2007 that almost 40% of new users to Twitter changed their privacy settings from the default public setting to the protected setting. By August 2009 this number had dropped to less than 8% of new Twitter users changing their privacy settings away from the default (Moore, 2009). This number differs greatly from what Lennart and Madden (2007) found regarding teens' privacy settings on social network sites. We do not know, however, if this difference is due to a) the fact that Twitter is a different kind of social network site than MySpace or Facebook and thus norms and technological affordances encourage more public sharing of personal information, b) the fact that Lennart & Madden (2007) relied on self-report measures whereas Moore (2009) scraped actual web data, or c) the potential that Twitter's older demographic may not be as concerned about privacy as social network site users.

While there are privacy concerns about who has access to personal information, this does not necessarily dissuade users from sharing information through communication technology. People can derive benefit from sharing where they are in public and when. Previous research suggests that sharing such time and location information can encourage social connectivity, as well as, facilitate face-to-face meetings (Barkhuus et al., 2008; Humphreys, 2007, forthcoming-a). For example, Dodgeball was a mobile social network whose function was to facilitate the sharing of people's locations with their members of their social network. Not only did such messaging reinforce social bonds among members, but this online and mobile communication facilitated offline face to face meetings in urban public settings (Humphreys, 2007). Dodgeball members did not have to explicitly communicate time in their messages. They could just send a text message with their location, and because it was time-stamped, members of Dodgeball network would know where they were and when. This points to an important similarity on

Twitter. If users share their locations in real time, they do not have to give temporal information because all Tweets are time-stamped as well. An important difference between Dodgeball and Twitter is that the default setting on Dodgeball was that only Dodgeball "friends" would receive location information (Humphreys, 2007). That said, sharing location information through Dodgeball still generated information on the website about which members had recently checked in at which locations. This information became publicly available on the Dodgeball website and could open users up for potential privacy concerns (Author own cite).

Many different kinds of information about people can be coupled together to make up a person's digital or internet footprint. An internet footprint is any information that a person has "created which is online, widely available, and specifically linked to author's real name" (Garfinkel & Cox, 2009, p. 2). This is an important concept because it suggests that while small bits of communication may in and of themselves provide little to no information about a person's identity and/or behaviors, in aggregate these bits of information together make up an overall footprint of the person that may tell a much deeper, more intimate story. Even though people may take moves in order to avoid "real" identification through their online information, such as through online pseudonyms or encryption services, there are an increasing number of technological measures to counter these moves including photo recognition software and computer recovery capabilities (Garfinkel, 2001; Garfinkel & Cox, 2009). Individual messages or bits of information may not seem very incriminating or important but in a digital age can be retrieved and aggregated with other kinds of information such as credit card purchasing habits or mobile internet searches to inform when and where people are, or even where people will be in the future.

Methodology

In order to examine the level of personally identifiable information that was disclosed on Twitter, we conducted a content analysis of a sample of tweets. With Twitter's permission, two of the authors collected a sample of tweets on the public timeline over three weeks from January 22 to February 12, 2008, at four distinct times of day (2:00, 8:00, 14:00, and 20:00 Mountain Time) and extracted the users that posted the statuses in these timelines. We collected details of the current user as well as a partial list of users being followed by the current user. To further the crawl, the first *m* users followed by the current user were added to the set of users to crawl (Manku, Rajagolana, & Lindsay, 1998). If the current user followed fewer than *m* users, all users are added to the set of users to crawl. Additional details regarding the sampling strategy can be found in (Krishnamurthy et al., 2008). In total, information from 101,069 tweets was collected.

The tweets were coded for whether or not they included: 1) personally identifiable information, 2) information regarding location, 3) proper names, 4) information regarding the time of day (not including the time stamp that is on all tweets), and 5) information about the author him or herself. Personally identifiable information was defined as information that could be directly tied to or associated with an individual such as email, phone number, or address (Gandy Jr., 1993). Examples of tweets not from our sample with personally identifiable information include: "@thereal [ten digit phone number] jXXXXXXXXXXXfilms@gmail.com" and "is looking for a StarCraft key...anyone? cXXXXXXXk@gmail.com please and thank you so much, now I can play against my son. hopefully I'll win".[i]

Tweets were coded as having location information if they included information regarding the *location of people*, i.e.: country, state/territory (e.g. east coast), city, and specific locales (e.g. coffee shop or restaurant, airport, highway, building name, etc). Examples of tweets including location information include "I'm back in the states...bout to get off the plane and go clear

customs and immigration" and "I'm going out to Home Depot to buy stuff for my kids project." Tweets that mentioned locations but did not suggest the location of a person were not be coded as including location information. Examples that mention places but would not be coded according to our definition include: "New York is not for the lighthearted" and "Off to never never land". Also our coding did not include self-referential locations like home and work.

Tweets were coded as including proper names if they referred directly to or mention people's real names or usernames. This included usernames, first names, last names, first name and last initial, first initial and last name, and both first and last name. Examples of this include: "Savannah is TOO RIGHT...Josh and I Tweet waaaay to often...only cause its so simple and fun through txt mssg" and "@SweetPea_woke up with vertigo. Apparently long family history. A little freaked out about how this might affect our life if it continues." Most Twitter profiles also include name information of the user him or herself, but we wanted to know how often users mentioned other peoples' names in their tweets as there may be privacy implications for a third party who was mentioned.

Tweets were coded as including time if they mentioned when activities occur or referenced *specific* times such as today, tomorrow, tonight, morning, afternoon, week, weekend, just (as in "recently"), now, this or next [followed by specific time or date], [specific date], Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday, etc. This did not include general time-related words such as everyday or again. Salutations such as "Good Morning/Afternoon/Evening" were counted as time-related. A length of time: i.e. "10 day spiritual retreat" did not denote a time related tweet, but those that mention a specific day, such as "in 10 days", would count because a "10 day retreat" does not reveal as much about where to find someone, or know someone is away, as a "retreat in 10 days" would. Examples of time

related tweets include: “Prep for a long ride tomorrow. Saddle height up and pasta for dinner. And we went to the pub. Oops!” and “is counting down the minutes until she can be in bed again... at 4am. ;(”

Tweets were also coded for whether or not they were about the author him or herself. Sometimes this was indicated with personal pronouns such as “I” or “me,”. Examples include: “i need to sleep more or at least stay away from tsetse flies.” Sometimes, however, the author’s role was implied, but not directly stated, for example: “at sears buying another 30 lbs dumbbell.” In this case, one could put “I am…” or the username before the message and see if it makes sense. If it did, then the message was coded as being about the author him- or herself. Explicit or implied plural pronouns such as “we,” “us,” and “our” were also coded as indicating the message was about author.

The content analysis involved three independent undergraduate coders who were trained for 4-5 hours per week for two and a half months with the first author. In total, the coders trained on 728 messages, in which all discrepancies were discussed and consensus was reached. Often during these trainings, the codebook was further refined to account for additional insights that emerged during the training process. When coders had reached acceptable levels of reliability (Lombard, Snyder-Duch, & Campanella Bracken, 2002), they coded a random sample of 2097 messages from the initial 101,069 tweets collected. They tripled coded 24% of this random sample (n=499). Based on this 24%, Cohen’s kappa was calculated for each category to ensure acceptable levels of intercoder reliability (Lombard et al., 2002) and ranged from 0.71 to 1.0 (see Table 1). Since there were three coders, kappas were calculated for each dyad and then averaged. All discrepancies were coded based on majority and determined by whichever two of the three coders agreed.

Results

The content analysis indicated several trends regarding the kinds of messages people publicly post on Twitter as they relate to potential privacy implications. More specifically, we present results regarding the number of messages that mention information regarding the author him or herself, messages that share personally identifiable information, messages about the location of people, messages that mention a proper name or messages about when people's activities occur. In addition, we present results about messages that share several of these characteristics.

Overwhelmingly, the content of tweets do not include *personally identifiable information*. Only 0.1% of messages in our sample (n=2) mentioned an email address, phone number or postal address. In our sample one of these messages included a phone number and the other included an email address, but importantly both included a proper name of another individual besides the user. Twitter messages that included information regarding the *location* of the Twitter user were more prevalent. In our sample, 12.1% of tweets (n=253) mentioned the location of a person. We have not yet gone through these tweets to determine if the location is past, present or future, but certainly all could have privacy implications. Certainly present or future discussions of location of people could raise safety concerns, but past locations may also indicate routines wherein such location information may occur in the future as well. Twitter messages that included a *proper name*, including Twitter screennames, accounted for 22.7% of our sample (n=475). Messages that included when user activities occurred (*time)* accounted for 20.1% of our sample (n=421). The majority of our sample also were coded as including information specifically regarding the activities or opinions of the Twitter user him or herself (*self*), accounting for 66% of our sample (n=1398) (see Table 2).

Within our sample, only 0.6% of messages (n=14) included information about the author, location, time and proper name. Only 3% of messages (n=63) were about the user him or herself *and* include location and time information. Only 2 messages (.01%) in our sample included location and time information in addition to a proper name, but did not include messages about the author him or herself. Fifteen percent of messages in our sample are about both the author and mention time (n=325). Ten percent of messages in our sample mention a location and are about the author (n=209). This suggests that when users mention time and location they are likely to be talking about themselves.

Discussion

Messages confined to 140 characters may seem and even be innocuous but they can still have important ramifications. Our study suggests that Twitter users do not often explicitly mention when and where they are in publicly available tweets. People almost never share personally identifiable information on Twitter. It was good to see that only 3% of our sample shared information regarding the users themselves and information about a time and place. Still, while that 3% may seem like a small number within the flood of 24 million tweets per day, it suggests that as many as 360,000 tweets per day may share location and time information of the Twitter user him or herself with the public.

Not only can people explicitly share where they are at particular points in time in their tweets, but when coupled with other information, either from Twitter Profiles themselves or other information publicly available through the web or even the phonebook, even sharing one of these kinds of information may put people at risk to be taken advantage of. For example, even though Hyman Israel did not tweet his home address, a phone book could have easily provided such information to the would-be robbery. It is not necessarily difficult to find out where exactly

someone one works or lives, through publicly available information offline or online such as profiles, company websites, or even another tweet.

This raises an interesting point. We coded each tweet as an individual unit of analysis. Therefore we did not couple tweets by the same user, nor did we couple these with profile information. Reading tweets over time from the same person could expose many more of habitual practices regarding location and time than just one individual tweet would. Similarly information in tweets may take on more significance when coupled with a Twitter user's profile information which includes "name", "location", and "bio". In addition to one's universe of tweets and profile information, a single tweet can also easily be coupled with other information online to allow for much greater privacy concerns than any individual message alone (Gandy Jr., 1993).

There are several ways to combat concerns about publicly sharing personal information. Of course Twitters could merely avoid sharing any such potentially personal information, but this is an unlikely and unhelpful scenario. Much of what makes Twitter a valuable service are the communications between people that rely on sharing personal information. A more helpful suggestion may be to be careful about with whom users share this information– that means, changing privacy settings on social media sites or being more careful whom you let "follow you" or friend you. As it is now, anyone can sign up to get Twitterers' messages without people's approval unless they have actively changed their privacy setting.

Changing privacy settings can protect you from lateral surveillance issues of everyday people taking advantage of you. But even when you're sharing this information with personal friends, you're actually sharing it with Twitter and Facebook first who then shares it with your friends. Thus participating in social media in general opens oneself up to the gaze of the

marketers so that any personal information shared, even if it's to a private group of close relations, can still be commoditized through the system itself (Andrejevic, 2007; Author own cite). As social media like Twitter are integrated into search engine results (Chapman, 2009), not only will such information be easier to find by those not on Twitter, but the information itself would be presented alongside other kinds of publicly available information.

While these are potential concerns, we don't want to be alarmist in our discussion of Twitter use. We all live in particular places and are still constrained by the fact that there are 24 hours in the day; as such time and location will continue to be relevant issues in our lives that we will continue to talk about. Our location is still easy to attain, and take advantage of, based on someone seeing us walk out the door on the way to work. In the past 25 years, theorists have argued space and time have become increasingly flexible (Harvey, 1990). Flows of space and time are characteristic of a networked society where geographic boundaries are quickly overcome (Castells, 2000). Nevertheless, in our daily lives we are still constrained in many ways by place and time. Thus communicating about time and space are fundamental ways in which we organize and orient ourselves (Casasanto, 2009). Therefore it is not terribly surprising that we see people communicating time and space information through social media. But what raises concerns is the *broadcastability* of such information through social media, so that we do not always know who has access to this information and what their intentions may be to act on it.

Conclusion

People may know to be careful about sharing phone numbers, home addresses and email addresses publicly, however, a second tier of personally identifiable information may raise concerns about who has access to information about where and when people are. One of the important points about location information is that when people say where they are or where they

will be, it also indicates where they are not. People do not have to explicitly say their home will be most likely be vacant when they are on vacation. Communicating that the entire family is in Disneyworld for the week, also says that the entire family is not at home in New York that week. Broadcasting such information may alert would-be robbers of where to go. Sometimes it's not always about where a person isn't but where he or she is that is cause for concern. For example, a US Representative tweeted that he had arrived in Baghdad and as a result put many people's lives at risk (Savvas, 2009).

Of course not all location information may lead to such legal or dire concerns. Sometimes people say they're at one place when they are actually at another, for example, going out when you say that you are home sick. Anecdotal evidence suggests that social awkwardness and embarrassment can arise when people use social media to communicate their actual locations whether it be through tweets or pictures that place people in certain locations at certain times when they said they would be elsewhere (Author own cite). We know that people often engage in white lies to avoid socially awkward situation with using their mobile devices (Birnholtz, Guillory, Hancock, & Bazarova, 2010). Now there are increasing ways to broadcast oneself to a variety of audiences, thus potentially increasing the opportunities for people to get caught in their white lies.

Sharing personally identifiable information through social media is not necessarily problematic. There are benefits that come from sharing personal information in social and public ways (Hampton & Wellman, 1999). People who use these sites are certain to expect that their postings will be read by someone. However, we do not fully understand Twitter users' conceptions of their audience. Posting to sites such as Twitter or Facebook implies some sort of audience, whether it be friends, family, colleagues, etc. Research indicates that people can be

quite adept at negotiating audiences and identity through social media (Lange, 2007). Nevertheless, future research should explore whether micro-bloggers conceptualize audience differently than do users of traditional social network sites (Tufekci, 2008) or blogs (Nardi, Schiano, & Gumbrecht, 2004).

Location is an important element through which we orient ourselves in the world and has helps us to connect with friends and kin. Therefore, communication and sharing about locations will not change, but what might change is the ability to control who has access to that information. Access and control of personal information is an important area of research which must be continually and critically examined. Default settings are a primary means of influencing and shaping online behavioral practices. Altering default settings may help to ensure that people can enjoy the benefits from sharing information through social media while protecting themselves from unwanted exposure.

This article only explores the actual content of messages but additional location and personal information may be gathered by GPS and mobile Google search that collect time and location of users. Even if the public does not have access to such personal information, marketers do. Educating consumers about the ways in which personal information can be used for alternative purposes is an important step in media literacy.

Future research should explore privacy attitudes and behaviors of social media users in order to determine explore everyday privacy protecting behaviors that people engage in to manage the tension between the sharing and maintaining privacy. As social media continue to proliferate, privacy protecting behaviors and conventions will emerge. As such, this is a critical time to examine and in turn influence these social and technological conventions.

End Notes

[1] The use of the "@" symbol before a username is a Twitter convention that allows users to direct public messages toward a specific user. When a Twitter messages includes an @username, the tweet message will be posted in that user's Twitterfeed as well as potentially on the public timeline.

References

Acquisti, A., & Gross, R. (2006). *Imagined communities: Awareness, information sharing, and privacy on the facebook.* Paper presented at the 6th International Workshop, PET 2006, Cambridge, UK.

Albrechtslund, A. (2008). Online social networking as participatory surveillance. *First Monday, 13*(3), article 6.

Alterman, A. (2003). "A piece of yourself": Ethical issues in biometric identification. *Ethics and Information Technology, 5*(3), 139-150.

Andrejevic, M. (2005). The work of watching one another: Lateral surveillance, risk, and governance *Surveillance and Society, 2*(4), 479-497.

Andrejevic, M. (2007). *iSpy: Surveillance and power in the interactive era*. Lawrence, KS: University Press of Kansas.

Barkhuus, L., Brown, B., Bell, M., Hall, M., Sherwood, S., & Chalmers, M. (2008). *From awareness to repartee: Sharing Location within Social Groups.* Paper presented at the CHI Florence, Italy.

Birnholtz, J., Guillory, J., Hancock, J. T., & Bazarova, N. (2010). *"on my way": Deceptive texting and interpersonal awareness narratives. .* Paper presented at the Proceedings of the ACM Conference on Computer Supported Collabortive Work (CSCW2010), Savannah, GA.

boyd, d., Golder, S., & Lotan, G. (2010). *Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter.* Paper presented at the Proceedings of HICSS.

boyd, d. m. (2004, April). *Friendster and publicly articulated social networking.* Paper presented at the CHI, Vienna, Austria.

Casasanto, D. (2009). Space for thinking. In V. Evans & P. Chilton (Eds.), *Language, cognition and space: The state of the art and new directions* (pp. 453-478). London: Equinox Publishing.

Castells, M. (2000). *The rise of the network society* (2nd ed.). Malden, MA: Blackwell Publishers.

Chapman, G. (2009, October 21). Microsoft, Google integrating Twitter into search results. Retrieved October 22, 2009, from http://www.google.com/hostednews/afp/article/ALeqM5ip8IjVKg2IeMFrm6ChMp15fEiN0A

ComScore, M. M. (2009). *Top 50 U.S. Web Properties for August 2009*. Reston, VA.

Ellison, N., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends:" Exploring the relationship between college students' use of online social networks and social capital. *Journal of Computer-Mediated Communication, 12*(4), article 1. Retrieved Novemer 1, 2009, from http://jcmc.indiana.edu/vol12/issue4/ellison.html.

Facebook, Inc. (2008). *Facebook Principles.* Retrieved November 4, 2009, from http://www.facebook.com/policy.php.

Gandy Jr., O. H. (1993). *The panoptic sort: A political economy of personal information*. Boulder, CO: Westview Press.

Garfinkel, S. (2001). *Database nation: Death of privacy in the 21st century*. Sebastopol, CA: O'Reilly & Associates, Inc.

Garfinkel, S., & Cox, D. (2009). Finding and Archiving the Internet Footprint. Paper presented at *First Digital Lives Research Conference: Personal Digital Archives for the 21st Century*. London, England.

Hampton, K., & Wellman, B. (1999). Netville online and offline: Observing and surveying a wired suburb. *American Behavioral Scientist, 43*(3), 475-492.

Harvey, D. (1990). *The condition of postmodernity*. Malden, MA: Blackwell Publishers.

Honeycutt, C., & Herring, S. (2009). Beyond microblogging: Conversation and collaboration via Twitter. *Proceedings from the Forty-Second Hawai'i International Conference on System Sciences (HICSS-42)*, 1-10.

Humphreys, L. (2007). Mobile social networks and social practice: A case study of Dodgeball. *Journal of Computer Mediated Communication, 12*(1), article 17. Retrieved Novemer 1, 2009, from http://jcmc.indiana.edu/vol13/issue1/humphreys.html.

Joinson, A. N., & Paine, C. B. (2007). Self-disclosure, privacy & the Internet. In K. M. Adam N. Joinson, Tom Postmes (Ed.), *The Oxford handbook of Internet psychology* (pp. 237-252). New York: Oxford University Press.

Kolek, E. A., & Saunders, D. (2008). Online disclosure: An empirical examination of undergraduate Facebook profiles. *NASPA, 45*(1), article 2.

Krishnamurthy, B., Gill, P., & Arlitt, M. (2008, August 18). A few chirps about Twitter. *Proceedings from ACM SIGCOMM Workshop on Online Social Networks*, 19-24.

Lange, P. (2007). Publicly private and privately public: Social networking on YouTube. *Journal of Computer Mediated Communication, 13*(1), article 18. Retrieved November 1, 2009, from http://jcmc.indiana.edu/vol13/issue1/lang.html.

Lenhart, A., & Fox, S. (2009). *Twitter and status updating* Washington DC: Pew Internet and American Life.

Lenhart, A., & Madden, M. (2007). *Teens, privacy & online social networks*. Washington DC: Pew Internet and American Life Project.

Liew, J. (2009, September 4). Are there more Facebook status updates or Twitter tweets? Retrieved October 20, 2009, from http://lsvp.wordpress.com/2009/09/04/are-there-more-facebook-status-updates-or-twitter-tweets/

Liu, H. (2007). Social network profiles as taste performances *Journal of Computer Mediated Communication, 13*(1), article 13. Retreived November 1, 2009, from http://jcmc.indiana.edu/vol13/issue1/liu.html.

Lombard, M., Snyder-Duch, J., & Campanella Bracken, C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research, 28*(4), 587-604.

Lyon, D. (2001). *Surveillance society: Monitoring in everyday life*. Buckingham: Open University Press.

Man Robbed After Posting His Vacation On Twitter. (2009). Retrieved October 19, from http://www.wpxi.com/news/19648421/detail.html

Manku, G. S., Rajagolana, S., & Lindsay, B. G. (1998). Approximate meadians and other quantiles in one pass and with limited memory. *SIGMOD, 27*(2), 426-435.

Metzger, M. (2004). Privacy, trust, and disclosure: Exploring barriers to electronic commerce. *Journal of Computer Mediated Communication, 9*(4), article 1. Retrieved Novemer 1, 2009, from http://jcmc.indiana.edu/vol9/issue4/metzger.html.

Mills, E. (2009, June 8). Twitter user says vacation tweets led to burglary. Retrieved October 19, 2009, from http://news.cnet.com/8301-1009_3-10260183-83.html

Mischaud, E. (2007). *Twitter: Expressions of the whole self.* Masters Thesis, London School of Economics, London, UK.

Moore, R. J. (2009, October 5). Twitter data analysis: An investor's perspective. Retrieved November 1, 2009, from http://www.techcrunch.com/2009/10/05/twitter-data-analysis-an-investors-perspective/.

Naaman, M., Boase, J., & Lai, C.-H. (2010). Is it really about me? Message content in social awareness streams. *Proceedings of the ACM Conference on Computer Supported Collabortive Work* (CSCW2010), Savannah, GA.

Nardi, B. A., Schiano, D. J., & Gumbrecht, M. (2004). Blogging as social activity, or, would you let 900 million people read your diary? *Proceedings of the ACM Conference on Computer Supported Collabortive Work* (CSCW2004), Chicago.

Savvas, A. (2009, February 10). Twittering US politician endangers lives in Baghdad. Retrieved November 4, 2009, from http://www.computerweekly.com/Articles/2009/02/10/234715/twittering-us-politician-endangers-lives-in-baghdad.htm

Shah, R., & Kesan, J. P. (2003). Manipulating the governance characteristics of code. *Info 5*(4), 3-9.

Tidwell, L. C., & Walther, J. (2002). Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations: Getting to know one another a bit at a time. *Human Communication Research, 28*(3), 317-348.

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.

Van Grove, E. (2009). Twitter your way to getting robbed. Retrieved October 19th, from http://mashable.com/2009/06/01/twitter-related-burglary/

Westin, A. F. (2003). Social and political dimensions of privacy. *Journal of Social Issues, 59*(2), 431-453.

Wolfe, F. (1992, May 29). As families attend funerals burglars hit vacant homes. *Washington Times,* p. B2.

Zimmer, M. (2008, March 2). The externalities of search 2.0: The emerging privacy threats when the drive for the perfect search engine meets Web 2.0. *First Monday, 13*(3). Retrieved November 1, 2009, from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2137/1943.

Table and Figures

*Table 1: Intercoder Reliability**

Category	*Kappa*
Personally identifiable information	1.00
Location	0.71
Proper Names	0.89
Time	0.78
Self	0.92

*Since there were three coders, kappas were calculated for each dyad and then averaged. All discrepancies were coded based on majority and determined by whichever two of the three coders agreed.

Table 2: Overall Frequency of Categories

Category	*Frequency*
Personally identifiable information	0.1% (*n*=2)
Location	12.1% (*n*=253)
Proper Names	22.7% (*n*=475)
Time	20.1% (*n*=421)
Self	66.0% (*n*=1398)
Total Tweets coded	2097