Historicizing New Media: a content analysis of Twitter

Lee Humphreys

Philipa Gill

Balachana Krishnamurthy

Elizabeth Newbury

Abstract

This paper seeks to historicize Twitter within a longer historical framework of diaries to better understand not only Twitter but broader communication practices and patterns. Based on a review of historical literature regarding 18th and 19th century diaries, we created a content analysis coding scheme to analyze a random sample of publicly available Twitter messages according to themes in the diaries. Findings suggest commentary and accounting style tweets are the most popular narrative styles on Twitter. Despite important differences between the diaries and Twitter, this analysis suggests long-standing social needs to account, reflect, communicate, and share with others.

Historicizing New Media: a content analysis of Twitter

Communication technologies and services are typically characterized by their "newness". Yet new is a relative term. Research often compares "new" communication technologies with their recent technological predecessors. For example, television was compared with radio when it was first introduced (Barnouw, 1966). More recently scholars have looked to historical communication systems and services to reveal insights into contemporary communication issues (Bolter & Grusin, 2000; Gitelman, 2006; John, 1998; Marvin, 1988; McCarthy, 2010; Milne, 2010; Park & Pooley, 2008). Indeed many have critiqued the overemphasis on the new, suggesting that all media engage with older media and social practices (Gitelman, 2006; Grusin & Bolter, 2000; Marvin, 1988). In this research vein, the current study seeks to historicize contemporary micro-blogging practices.

Twitter is a popular social media service that allows people to share updates, news, and information (known at "tweets") with people in their Twitter network and beyond. With over 200 million registered users (Halliday, 2011) and 13% of online Americans using Twitter (Smith, 2011), Twitter is one of the most popular social media available. Research has compared Twitter to earlier kinds of social media like blogs (Java, Finn, Song, & Tseng, 2007) and social network sites (Gruzd, Takhteyev, & Wellman, 2011). Recent analyses of microblogging suggest that the brevity and broadcastability of messages are important affordances of microblogging (boyd, Golder & Lotan, 2010; Java, et al., 2007). Twitter, however, is not the first service to allow people to share short messages with other people.

It can be revealing to make comparisons with communication practices from even earlier historical periods. One can look beyond blogs and social network sites to find other historical examples that demonstrate remarkable resemblance to microblogs. Placing microblogs into a

longer historical context helps to reveal what is new and not new about microblogging. Therefore this article situates Twitter and other status updating services into a historical context of personal writing for public consumption. Although Twitter may seem unique from its immediate communication technological predecessors, earlier technologies for personal writing may help us to better understand Twitter's popularity and use today.

While today the term "diary" might conjure up thoughts of locked notebooks where people pour their inner most thoughts, this has not always been the case. Indeed the diary as a location for private confessional is but one of many uses for diaries (Culley, 1989; Fothergill, 1974). Historically, certain kinds of diaries were written with the intent to be shared and were made up of relatively brief writings (Culley, 1989). It is these kinds of writings that warrant a closer comparison with contemporary examples of personal shared writings.

By historicizing Twitter within a longer historical framework of diaries, we seek to better understand not only Twitter but broader communication systems, practices and patterns. For example, the desire to communicate mundane activities may serve a phatic communicative purpose, where the act of communicating serves to reinforce social connection. Broadcasting or sharing personal information with many others may also serve to reinforce connections within the community or network. The boundaries between personal and professional spheres can also be further revealed by exploring pre-industrial personal writings when such spheres were not clearly delineated (Sennet, 1992). Reviewing scholarship about diaries in order to analyze Twitter not only reveals insights into Twitter but also the diaries themselves by potentially revealing contemporary and long-standing social and communicative practices.

Historically some diarists write only about themselves while others write about their friends, family, and community. Therefore in this study we analyze the resemblance of the actors identified in tweets to those discussed in historical diaries (R1). Topics of diaries have ranged from mundane everyday activities and events to introspective and contemplative thoughts. Therefore we examine the prevalence of various historical diary topics on Twitter today (R2). In addition to various topics, the narrative style of historical diaries has changed from daily chronicling or accounting to more introspective exploration and commentary, therefore we also explore the kinds of functions or style of the tweets relative to historical diaries (R3).

We begin by reviewing literature about historical diaries and shared personal writings in order to contextualize Twitter. Based on this literature, we developed a content analysis of publicly available tweets in order to explore the kinds of information that are shared publicly on Twitter and to see if the tweets contained similar topics to the topics historians have identified in diaries from hundreds of years earlier. Content analysis is particularly useful here because it differentiates this study from previous research linking old and new media (e.g. Bolter & Grusin, 2000; Milne, 2010), which draw on case studies to reveal genealogy or parallels with historical media. Content analysis allows us to quantitatively analyze the content (Krippendorff, 2004) of tweets to more systematically ground our historicization in the actual content of the Twitter messages themselves. After we present the findings of the content analysis, we discuss the similarities between Twitter and historical writings, as well as the characteristics of Twitter that do not have precedents in historical writings.

Literature Review

Throughout history, there have been many different kinds of personal writings, in addition to letters, that were meant to be shared. In particular, certain kinds of historical diaries

were written with an understanding that they would and should be read by others. Indeed, historians suggest throughout the 18th and 19th centuries many diaries were semi-public and intended to be shared with an audience (Culley, 1989; Motz, 1987). "It is only relatively recently (within the last hundred years) that the content of the diary has been a record of private thoughts and feelings to be kept hidden from others' eyes." (Culley, 1989, p. 15). Not all historical diaries were meant to be shared. Early religious diaries in the 17th century were often private and thought to encourage pious behavior (Fothergill, 1974). These differed from secular personal writings which tended to be of two categories: introspective or accounting (Motz, 1987). Throughout the early and mid 19th century, "accounting" diaries were quite popular and tended to document travel or everyday events in the household and community (Culley, 1989). Importantly, these diaries were also often intended to be shared. Because of their audience, content, narrative style, and limited length, these historical diaries share many characteristics of modern microblogs and are thus worthy of closer review.

Audience

Diaries from the 17th, 18th and 19th centuries were often semi-public in nature (Bloom, 1996; Culley, 1989; Fothergill, 1974). Some were even considered a kind of public record. Some accounting diaries, particularly those by women, were often shared during the diarist's lifetime to maintain family and communication networks (Culley, 1989). These journals chronicled the life events of the family and could be mailed to extended family members who lived far away. For example, when young women would marry and move away from their families, some would keep journals and then send them to their families as a way of maintaining kin ties. Travel diaries were also very commonly shared upon returning from travel or sent to others while away (Fothergill, 1974). Sometimes husbands or wives of military men would keep journals while

traveling and send them to relatives as a way of staying connected (O'Sullivan, 2005; Sjödblad, 1998) . Travel diaries were also used as a means of reporting on the experiences of what happens abroad to an audience back home (Fothergill, 1974; Sjödblad, 1998). Sometimes people would rewrite their diary entries as a means of improving their writing style and penmanship, especially if they were planning to share their diaries (Hunter, 1992). In rural areas, diarists would share their journals with friends who visited (O'Sullivan, 2005). Much like today we might share a scrapbook or a photo album with a visiting friend, historically diaries were meant to be shared with visitors as a way of documenting and sharing important events in the family and community.

By the end of the 19th century diaries were less accounting diaries and more introspective. However, even these introspective diaries would be read by those other than the diarist him or herself. It was not uncommon for Victorian girls to maintain diaries that were then later read by their parents or read aloud to the entire family (Hunter, 1992). These young women would also share their diaries with close personal friends. Sometimes they would write together and even write in each others' diaries as a way of developing intimate relations with each other (ibid.). In eras where emotional intimacy was not readily articulated, the shared diary was a means through which people could socially connect with others. Historical diaries were much more public and shared forms of communication than what we typically may think of diaries today.

There are also technological indications of the potential audience of historical diaries. Indeed diaries did not include privacy features until the mid-19th century (Hunter, 1992; McCarthy, 2000). Ties and locks were not common on personal diaries until around 1860. Prior to this, diaries resembled books, which could be easily shared and read by others.

The degree of publicness of some historical diaries was of course more limited than the broadcastability and accessibility of microblogs today. Nevertheless, even today some Twitterers selectively choose who can see their microblogs and limit access to only a few (Moore, 2009). Research on Twitter users suggests that the audience they have in mind when they tweet is primarily to the people following them on Twitter but also writing to themselves (Marwick & boyd, 2010). Like historical diarists, some Twitter users report writing to express themselves but necessarily to an intended or known audience. How exactly the Twitter users articulate an audience in their tweets has yet to be systematically analyzed.

Limited length

One of the distinguishing technological affordances of micrologging is the limited length of messages. Twitter is limited to 140 characters. Historical parallels suggest that this technological limitation can be a welcomed restriction for many chroniclers of life events. With advancements in paper production and printing, small leather-bound journals called pocket diaries became popular in the mid-19th century (McCarthy, 2000). About 2x4 inches in size, these small journals could easily be tucked in a pocket or in a waistband; and were thus more mobile than earlier journals had been. Because of the physical size of the diaries, users were also limited in how much they could write, typically keeping their entries to only two or three sentences. For example, on January 26th, 1873, Jane Fiske wrote in her pocket diary: "Cold disagreeable day. Felt very badly all day long and lay on the sofa all day. Nothing took place worth noting," (McCarthy, 2000, p. 274). The limited size of pocket diaries was not necessarily a liability for diarists of the 19th century; instead the limited size was a welcomed constraint. Historian Molly McCarthy (2000) writes:

The space afforded by the pocket diary may have been limited, but it saved journalists

with only minutes to spare from having to write long entries. And diarists appeared

thankful for both the opportunity pocket diaries offered as well as the limitations they

imposed. (p. 295)

Pocket diaries technologically limited writers to brief prose, much like microblogs do today. In

the mid-19th century, when leisure time was still something only afforded to the upper classes,

pocket diaries imposed a welcome limitation on the amount of writing that literate middle class

diarists had to record events and activities (McCarthy, 2000).

In addition to the physical size of the diary and its pages, there were other technological

constraints which limited how much authors could write (McCarthy, 2000). For example, the

printing of lines on paper constrained the space on which people wrote. Similarly when diaries

include printed dates on the pages, it constrained the amount of space a person could write for

that day. While people could go outside the lines or write in other areas of the page, these

additions to the blank pages technologically influenced how and how much people wrote.

Content

The content of particularly women's diaries during the 18th and 19th centuries often focused

on life events such as births, deaths, marriages, travels, visits, illnesses, and work (Sjödblad,

1998). It was not until the end of the 19th century that secular diaries became much more

introspective and confessional in nature (Culley, 1989; Hunter, 1992; Motz, 1987). Particularly

for early women diarists, the content of their journals could be characterized by their mundane

and even repetitive nature (Culley, 1985). Some of these diaries would cover topics "ranging

from information on the weather and daily chores to remarks on the health and well-being of

friends and family" (McCarthy, 2000, p. 275). These diaries documented the activities of

everyday life. By the end of the 19th century this also included media use. As leisure reading

became more popular in the mid 19th century, young girls were encouraged to use their diaries to record which books they had read (Hunter, 1992).

The distinctions between home and work or public and private were often blurred during the early 19th century and the content of the diaries of that time often reflects this. Particularly for women whose work occurred in the domestic sphere the home was not a private sphere (Hunter, 1992). For example, Elizabeth Fuller wrote in her diary about her work spinning fabric, as well as visits from friends to her family in 1792:

> Sept.6- I spun three Skeins.
>
> Sept 7- Fidelia Mirick here a visiting to-day.
>
> Sept 8- I spun three skeins to-day.
>
> Sept 9- I spun three Skeins. Pa & Ma went to Mr. Richardson's a visiting.
>
> Sept 10- I spun three skeins. (Culley, 1985, p. 75)

Fuller's diary chronicles her work and social life. The historical chronicling of social or familial events as well as travel often blurred home and work spheres. Indeed Fothergill (1974) argues that Samuel Pepys' late 17th century diary is the first example of a diary blurring the public and private in which he "treats no subject as either too intimate or too impersonal to be mentioned" (p. 13). This suggests that people throughout history have chronicled their everyday personal and public activities and experiences in much of the way that Twitter allows people to do today. Interesting to note, there is a Twitter account for Samuel Pepys, tweeting his diary entries from the 17th century.

This historical and contemporary personal chronicling often reflects the blurring of work and home life. Analyses of Twitter find the majority of messages describe the activities and experiences of the Twitterers themselves (Java, Finin, Song, & Tseng, 2007; Naaman, Boase, &

Lai, 2010; Smith & Rainie, 2010). An example of this kind of tweet might be: "Ran 3 miles today. Time to eat, then study and work on website." Communication technologies are often charged with the blurring of public and private spheres (Marvin, 1988). More recently, home computers, the internet, and mobile technology have contributed to shifting boundaries between work and home (Haddon, 2006). While modern communication technologies have been charged with the blurring of public and private spheres, particularly for women, we see this blurring to be historically true as well. This blurring is not just a modern phenomenon but a historical one as well. Twitter user research suggests that people are aware of multiple audiences for their tweets and will self-censor their topics accordingly, avoiding information that may be too personal or too controversial (Marwick & boyd, 2010). Indeed users in Marwick & boyd's study recounted the tension between self-revealing tweets and informative tweets, that is, tweets about the user him or herself and tweets about a particular topic or development of interest. Despite Twitter users acknowledging this tension, we do not know how this tension manifests in the tweets themselves. In contrast, we cannot necessarily know *how* diarists from history have chosen the topics of their daily writings, we only know *what* they have chosen to write about.

Narrative style: Reflection versus accounting

The narrative style of diaries in the 18th and 19th centuries was often matter-of-fact and truncated (Culley, 1985; McCarthy, 2000). Early secular diarists did not often write their feelings, thoughts, or beliefs into the narratives of their journals. Instead, these diaries had brief and episodic entries that used terse and concrete language (Culley, 1989). The narrative style of these diaries reflects a matter-of-fact chronicling of life events. "The earliest [diaries] of course contain very short entries, rendering everyday facts, and it is not until the eighteenth century that we find journals where a personal voice breaks the enumeration of happenings," (Sjödblad, 1998,

p. 517). Early diaries tended to use a narrative style that reported on everyday events rather than reflecting on them. For example, below is an excerpt from Mary Vial Holyoke's diary from 1770 in Salem, Massachusetts:

> Apr. 7. Mr. Fiske Buried.
>
> 23. Went with Mr. Eppes to Mrs. Thomas. Took Down Beds.
>
> 26. Put Sals Coat in ye frame.
>
> 27. Made Mead. At the assembly.
>
> May 14. Mrs. Mascarene here and Mrs. Crownsheild. Taken very ill. The Doctor bled me.
>
> Took an anodyne. (Culley, 1985, p. 5)

This curt style of narrative regardless of topic is similar to some kinds of narratives we see in microblogging today. Holyoke chronicles the death of eight of her own children in the same perfunctory manner as the excerpt above. Similarly, a young active microblogger tweeted that she was having miscarriage in an equally terse narrative (Clark-Flory, 2009). Twitter has become a place of people to announce and discuss deaths of famous people (sometimes prematurely) (Niles, 2009). Mundane and tragic life events are matter-of-factly reported through writings of historical diarists and microbloggers alike.

Particularly toward the end of the nineteenth-century, more reflective and introspective diaries emerge (Culley, 1989; Sjödblad, 1998). The diary, especially for women and young girls, is a place for them to have a voice and provide them an opportunity to discuss and explore their inner thoughts (Brumberg, 1997; Culley, 1985, 1989; Simmons & Perlina, 2002). Diaries in the Victorian era were often begun by girls as a character-building exercise (Hunter, 1992). Like early religious diaries (Fothergill , 1974), it was thought that the diary could serve as an opportunity for girls to reflect on daily activities and consider the morality of her actions. These

more reflective diaries often focused just not on the activities of the writer but on the writer's own consciousness (Culley, 1989). Other actors do not figure prominently in the narratives of these personal reflective diaries (Hunter, 1992). This differs from accounting diaries that chronicle the activities of the diarist as well as others in the family or community (Culley, 1985).

While these more reflective and introspective diaries were not necessarily always intended to be shared during the diarist's lifetime, nevertheless there is evidence to suggest that these diaries were indeed read by those other than the writers themselves. Youth in the late 19th century were not often granted the same privacy privileges as adults, so Victorian parents would often read their children's diaries, sometimes to the whole family (Hunter, 1992). Researchers also suggest that the mere act of writing and keeping a diary implies an audience (Ede & Lunsford, 1984; Ong, 1975; Rosensweig & Thelen, 1998). This audience may be a future self or those in later generations, but the act of writing may presume to some degree the potential of an audience.

Due to the public nature of Twitter, an audience of some degree may be presumed (Marwick & boyd, 2010), but the narrative style on Twitter has yet to be examined. In particular this study seeks to explore how the narrative styles on Twitter resemble early accounting diaries or more reflective diaries.

Current study

The current study sought to further explore the similarities between historical diaries and Twitter by conducting a content analysis of publicly available tweets. Because there is no equivalent corpus of diaries to Twitter, a direct comparison is not possible. Nevertheless, based upon the previous literature of historical diaries, this study seeks to systematically explore the kinds of messages on Twitter.

To answer our research questions, we analyzed Twitter messages along three main areas: a) who was talked about in the tweet (aka Tweet Subject), b) what was the topic of the tweet (Tweet Topic) and c) what was the narrative style of the tweet (Tweet Style)?

Tweet Subject. Historically, there was a shift in *who* was discussed in diaries from the self and community to primarily the self (Culley, 1989). For the purposes of this study we defined the Tweet subject as the person [people] who is being discussed in the message. This is the person giving rise to the specified response, feeling, or action described in the message. The subject can be the person doing the activity of the tweet. It can also be the person to whom an activity is being done. A single Twitter message could be coded as having multiple subjects. It should be noted that the subject must be a human being, whether he or she is alive or dead. Messages whose subjects are animals, inanimate objects, or fictional characters were coded as "None." Tweet subjects were coded as 1st Person, 3rd Person Singular, 3rd Person Plural, Audience, and None.

1st Person" subject may be indicated with personal pronouns such as "I" or "me," or implied but not directly stated. Tweets were coded as "3rd Person Singular," if a real (non-fictional) person's specific name, username, title, or initials are mentioned. Tweets were coded as 3rd person plural if groups of people with a particular knowledge, skill, profession, or characteristic were mentioned. This category also included a specific group's name. Organizations or sports teams were considered 3rd person plurals. Tweets were coded as "Audience" if they articulated or acknowledge an audience for the tweet within the tweet. This was done in three main ways: a) including the username of another Twitter member in the Tweet, b) including terms which implied a general Twitter community such as Twitterers, followers, or tweeties, and c) including indefinite pronouns without a specified referent such as you or your.

Tweet Topic. Tweets were also coded for several topical areas. Based on the historical diaries, we coded for several "mundane" topics. Topics were not mutually exclusive, nor was every tweet coded for a topic. We coded tweets as *weather* if they directly commented on or mentioned weather or weather-related conditions such as snow, rain, sun, wind, & temperature. This did not include weather-related paraphernalia such as hats, gloves, umbrellas, boots, sleds, etc. We coded tweets as *family* if they directly mention one's family relation such as sister, husband, child, etc. These tweets also include family of people in the author's personal network (e.g. a friend's sister). *Food/Beverage* tweets were coded as such if they mentioned the consumption, preparation, or ingestion of food and/or beverages. Tweets were coded as *religious* if they mentioned a religious activity, specific religion, prayers, or religious leaders and had overall religious reverence (e.g. tweets with "oh my god" or OMG were not coded as religious). Tweets were coded as *health* if they mentioned or commented on the health of a person or health-related behaviors of individual people, including exercise, stress, and sickness. Dieting or drinking excessively were also considered health-related. News about health was not considered health. We coded tweets for whether they describe states or activities related to *sleep* such as asleep, sleep, slept, sleepy, nap, tired (as in sleepy), yawn, waking, awake, or "zzz." The sheer mention of words did not necessarily determine a tweets topic. For example, a tweet that included the term "I'm sick and tired of…" was neither coded as health-related nor sleep-related.

Late 19th Century young diarists were encouraged to track their readings of books, considered popular media at the time (Hunter, 1992), so we coded tweets for general *media* if they mentioned media or communication technology, including: TV, movies, music (genre, song, or specific band/artist), internet, computers, mobile phones, digital gaming (e.g. xbox, ps3, wii, world of warcraft), Twitter (tweet, or twitter related word), books, magazines, graphic novels

(comic books). URLs were not coded as media unless the tweet made reference to them (e.g. "check out this website…").

We also coded for discussion of *activities* to resemble the kinds of everyday acts that were reported in historical diaries (Sjödblad, 1998). Activities were defined as tweets about an action, about doing something, about being active. Being at a place was considered an activity, however, tweets describing cognitive activities (e.g. wanting, thinking, contemplating, missing, and needing) were not coded as activities. Activities could occur in the future, present or past. Historically diarists would document activities regarding home life, work, and the community (Culley, 1985), therefore activities were further sub-coded as home-related, work-related or activities outside of home and work. *Home-related activities* were coded as such if they discussed activities that occur in or around the home such as cooking, cleaning, laundry, gardening, home or vehicle repair, as well as preparing for/being en route for home. *Work-related activities* were coded as such if they discussed activities related to work or school such homework, reports, bosses, teachers, students, and work. We coded activity tweets as *outside of home and work* if activities typically occur not in the home or traditional work environments such as shopping, going to concerts/theater or eating and drinking out, as well as running errands, and exercising. Travel or transit unless specified as work-related was coded as an activity outside of home and work. Activity tweets could be coded into multiple sub-areas if appropriate. For example, "working at Starbucks today" would be coded as both work-related and outside of home and work. Tweets related to computer or media activities were not coded into the sub-areas unless the location or use was explicitly identified as work- or home-related or outside of home and work due to convergent and ubiquitous nature of contemporary media use.

Tweet Style. The style of a tweet is defined as its broad purpose or form. This was loosely based on the various functions or styles of diaries over the centuries (Motz, 1987). These categories were not mutually exclusive. A single Tweet could have all styles, but all tweets had at least one of the following styles: Accounting, commentary, content sharing, information seeking, and response. Tweets were coded as *accounting* if they report on or share current or recent information and activities of the person in the tweet, or convey changes in status. Tweets were coded as *commentary* if they expressed an introspection, opinion, or an evaluation. Jokes and rhetorical questions were considered commentary as well as emotional expressions, including emoticons and punctuation such as exclamation makes, ellipses and all caps. Tweets were coded as *content sharing* if they contain content from other websites, authors or sources. Any "hyperlinked" URL or quote was considered content sharing. Sources did not have to be identified for a tweet to be considered content sharing. Tweets were coded as *information seeking* if they sought out information, answers, opinions, or feedback about the Tweet subject or topic. Tweets were coded as *response* if they indicate a response to a prior conversation. This was often indicated with the @username convention, though not always. Responses, at times, were identified by their use of undefined referents, such as "them," "that" (noun), or "it," or by common answers, such as "yes," "no," "sure," or "gotcha," both of which indicate prior communicative exchanges.

Sample

With Twitter's permission, we collected an initial sample of users whose tweets appeared in the public timeline. Friends of these users were then crawled using a constrained breadth first search technique. (We maintained a running median, m, of the number of friends all users had and only collected the first m friends for each user.) A second set of sample tweets was collected

by repeatedly querying the public timeline over three weeks from January 22 to February 12, 2008. Additional details regarding the sampling strategy can be found in Authors (2008). In total, we collected information about 101,069 tweets (message and user). From this study, we distinguished between tweets which had been submitted via the website and tweets submitted via text message because we wanted to see if tweet mode included a different kinds of information. We randomly sampled 1050 web and 1050 text message tweets to include in our content analysis. We coded 1026 web tweets and 1046 text tweets after excluding tweets in a language other than English.

Coding Procedures

The content analysis involved two pairs and two triads of independent undergraduate coders who were extensively trained on particular categories. Coding teams trained with the first author on specific categories for about four hours per week for between two and a half months to five months on 628 to 950 messages for each of the variables before reliable coding was attained. Reliably coding such short messages proved challenging due to lack of contextual cues; therefore, during these trainings, the codebook was often 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 the random samples of web and text tweets (n=2072) from the initial 101,069 public tweets collected. Coders double-coded (or triple-coded, if the coding team was a triad) 24% of this random sample (n=499). In order to account for coder drift, the first 300 tweets and the last 199 tweets were double/triple coded. Based on this 24%, Krippendorff's Alpha was calculated for each coder category to ensure acceptable levels of inter-coder reliability (Hayes & Krippendorff, 2007; Lombard et al., 2002;) and ranged from 0.70 to 1 (see Table 1). All discrepancies of the multi-

coded tweets (i.e. the reliability sample) were discussed and consensus coded before including them in the analysis.

Findings

We analyzed the tweets for three major areas: tweet subject, tweet topic, and tweet style in order to identify who was identified in the message (tweet subject), what was discussed (tweet topic) and how was it discussed (tweet style).

Tweet Subject

The majority of Tweets included information from the point of view of the first person. This category of tweets accounted for about two thirds of the sample (66.4%, n=1,375). Over 20% of the sample mentioned another individual or person (third person singular). About 12% of the sample mentioned a group or organization (third person plural). Almost 20% (n=390) indicated some kind of audience for their message, either by including a Twitter username in the tweet or indefinite referent such as "you". Lastly, 15% of the tweets (n=310) did not identify or articulate any person or people giving rise to the specified response, feeling, or action described in the message. Tweets sent via text message were negatively correlated with tweets which identified an audience (r=-.05, p<.05). This and all correlations reported are two-tailed tests. Tweet mode did not correlate with any other tweet subject categories.

Tweets subjects were not mutually exclusive, so while 66% of tweets involved the first person, 41% of those tweets involved another person in some way (either as audience, third person singular, or plural) in addition to the first person (see table 3). Additionally, we found that 46% of tweets involved another person beyond the author him or herself. This suggests that people on Twitter are not just talking about themselves or their organizations in the first person but are invoking or actively discussing other people and organizations in their Tweets.

Tweet Topic

Based on the topics identified from the historical diaries, we were able to code 68% of

our sample (n=1,399) for the tweet topic. Of the topics we coded for, the most frequent was the

reporting of activities, accounting for 41% (n=852) of our sample. We sub-coded activities into

home activities, work activities or activities outside of home a work when able. We were only

able to further code 61% (n=521) of the activities into home activities, work activities or

activities outside of home a work. The largest kind of activity tweeted was outside of home and

work. These tweets was significantly correlated with mobile tweets (r=.19, p<.001). Media and

activities were correlated (r=.07, p<.01); however, unless it was explicitly stated, tweets that

mentioned media were not presumed to be either work, home, or outside of home and work.

Media convergence and ubiquity today cannot easily be place or context defined, but media use

was prevalent nonetheless in our sample. Indeed of the 331 activity tweets not further sub-coded

into home, work, or outside of home and work, 69% (n=229) of them were coded as media-

related.

Each of the additional tweet topics were found to be represented in the sample though to

lesser and varying degrees. Overall these historically informed topics (food/beverage, health,

weather, sleep, family, religion) accounted for 26% of our sample (n=550). Religion was the

least frequent topic of the tweets while food and beverage tweets were the most frequent after

activities and media.

Home and work activities were negatively correlated with tweets from the mobile (r=-

.07, p<.01 and r=-.06, p<.05 respectively) whereas activities outside of home and work were

positively correlated with tweets from the mobile phone (r=.20, p<.001). Media tweets were also

negatively correlated with mobile tweets (r=-.17, p<.001) as were sleep-related tweets (r=-.06,

p<.05). Food and beverage tweets were positively correlated with mobile tweets (r=.06, p<.01). The remaining tweet topics were not significantly associated with tweet mode.

Tweet Style

Overwhelmingly, most tweets involve some sort of commentary with 75% of our sample coded in this category. Sixty-two percent of tweets were accounting-style tweets that reported on current or recent information and activities about the subject of the tweet. Taken together, over 95% of the sample (n=1,975) involved either accounting and/or commentary-style tweets. Content sharing tweets made up 11% of the sample, while information seeking and responses each accounted for 4% of the sample.

Mobile tweets were negatively correlated with content sharing (r=-0.24, p<.001). This is not surprising as it can be quite difficult share URLs via a text message, especially if one has to transform a long URL into a short URL so as to fit into the 140-character limit. None of the other tweet styles were significantly correlated with the mode of tweet.

There was a relationship between who was identified as the tweet subject and the tweet function or style (see Table 3). In particular first person tweets were most strongly correlated with accounting (r=0.40, p<.001), and negatively correlated with all of the other tweet styles. Also those tweets which identified or acknowledge an audience were significantly positively correlated with information seeking tweets (r= 0.25, p<.001) and responses (r=0.32, p<.001). Tweets with no subjects were correlated with both commentary and content sharing (r=0.06, p=.005 and r=0.10, p<.001, respectively). Third person singular tweets were also correlated with content sharing (respectively, r=.03, p<.05,). Third person plural tweets were not significantly correlated with any of the tweet styles.

Discussion

This study sought to explore the content of Twitter as it relates to the content and style of historical diaries. While we cannot make direct quantitative comparisons between a corpus of historical diaries over the centuries and public tweets, this study systematically analyzed a random sample of publicly available tweets to look for similarities and contrasts to the ways historians have characterized various kinds of historical diaries. In particular we explored the resemblance of the actors identified in the tweet to those discussed in historical diaries (R1), the prevalence of various historical diary topics on Twitter today (R2), and the narrative style of tweets as they related to accounting or introspective-style diaries (R3).

Historically diaries suggest that there have been cycles in diary use regarding who is the focus of the diaries. Very early religious diaries focused on the author him or herself as a means of maintaining piety, while later diaries became more accounting diaries which would more focus on the diarist's activities and the activities of other people (Fothergill, 1976; Culley, 1986; Ulrich, 1991). Towards the second half of the 19th century, however, there was another shift toward more introspective but secular personal writings (Culley, 1986), which also focused more on the individual diarist. Our data suggests that Twitter combines elements of both 18th and 19th century diaries. In our sample the most prevalent kind of actor identified in the tweet with the Twitterer him or herself (first person). However Twitter is not only a place for people to talk about themselves. Indeed over 45% of all the tweets in our sample involved someone beside the author him or herself, while 15% of tweets were not about people at all. Thus Twitter would seem to most resemble those diaries which were considered social histories for the group, rather than those diaries which only recount the individual's own thought's and development.

Some historical diaries focused on everyday events and activities rather than the emotional or spiritual development of the diarist (Culley, 1986). Our sample suggests that the

topics discussed on Twitter slightly resemble late 18th century secular diaries. The limited number of topics that we coded for accounted for 68% of our sample, with activities and media being the most popular topics discussed. It is not surprising that on a medium such as Twitter, that media itself would be a large topic of discussion.

While the everyday topics such as activities and media were popular on Twitter, the fact that so much of our sample included commentary suggests an important divergence from merely historical diaries that accounted or reported the events of the day. With 75% of our sample including some commentary and 62% including an account, our Twitter sample suggests a blending of both reflective and accounting practices. Like secular diaries of the late 18th century, Twitterers seem to be writing about new information of the day but they are also adding commentary to their messages like the diaries of the later 19th century. So it seems that the majority of tweets combines the narrative styles of different kinds of historical diaries.

Another important divergence of tweets in this sample from historical diaries is information seeking and response. While some historical diarists would certainly share their diaries and even occasionally write in one another's diaries, there is little evidence to suggest that diarists actively asked questions to one another through their journals. Indeed young Victorian girls would write about people in their diaries rather than writing directly to them in their diaries (Hunter, 1992). While information seeking and responses accounted for the most rare narrative styles, there presence at all reflects the highly interactive nature of Twitter that historical diaries could never achieve. This difference marks a unique characteristic of Twitter and one which helps to account for the networks' potential and vast influence.

Despite the similarities we found between the historical diaries and Twitter, there are some very important differences beyond what we have analyzed here. First and foremost the size

of the potential audience on Twitter is significantly larger than with historical diaries. This difference suggests that a Twitter message not only can reach more people but also potentially reach a more socially and geographically diverse audience than analogue diaries. Another importance difference is the degree of interactivity on Twitter compared to the diaries. While there is evidence that people would send diaries back and forth, the near simultaneity of Twitter allows for much greater interactivity than the diarists would have had who were reliant on the postal system. This difference can facilitate vast amounts feedback more quickly, as evidenced in the presence of information seeking tweets. While this style of tweet as not the most frequent in our sample, some have argued that an increasingly important contribution of online social networks like Twitter is the ability to ask a question of one's network and get feedback from hundreds if not thousands of people instantaneously (Morris, Teevan & Panovich, 2010). This difference between levels of interactivity on Twitter and the diaries is very important and begins to highlight how and why Twitter is so popular today.

Another importance difference with the diaries is that Twitter itself is a network, suggesting that information can be connected, linked and sorted to see patterns such that the whole of Twitter is more valuable than the sum of its parts. While the diarists may have been socially linked within communities, the diaries themselves were not. One could argue that the diaries become linked or networked when they become part of a digital archive, however, unlike Twitter the diarists themselves are not able to use or leverage the collection or network of personal writings during their lifetime.

In addition to the comparisons between diaries and Twitter, we also explored how the mode of tweeting was related to the content and style of tweets. The differences we found in the sample regarding web versus text message suggest that the context of tweeting may influence the

kinds of tweets people write. It is important to note, however, that the mobile-based tweets in this sample were only text messages and did not include mobile Twitter applications as these were still relatively uncommon when we collected our sample. Nevertheless, it is not surprising that tweets about activities outside of home and work including travel would be associated with text-messaged tweets due to the mobility of the device at hand. The popularity of historical travel diaries (Fothergill, 1976) suggests further a long-standing desire to chronicle new events and experiences to share them with others.

The negative relationship we found between text message tweets and content sharing may also be a product of the time of the sample. Increasingly mobile devices and tablets make it easier to share stories and information read online through social network sites like Facebook and Twitter. Nevertheless, smartphones still represent a minority of the global mobile phone market and as of September 2011 four billion text message tweets were sent monthly on Twitter (Tsotsis, 2011), which suggests analyzing text message based only is still an important comparison. Future research should further examine how tweets sent via mobile applications differ in content and style from other modes.

Limitations

There are several limitations of our study. First, our sample was taken from 2008 at the beginning of the exponential growth of Twitter (Marwick & boyd, 2010). In particular, the use of retweets and directed messages (i.e. using @username to ensure another Twitter user saw the tweet) were just beginning to become established practices on Twitter. Therefore our findings regarding audience and content sharing may be particularly less than current Twitter practices.

Another potential limitation is the fact that we only coded for eight topics on Twitter. Clearly there are many more topics to be identified on Twitter. We initially coded for travel and

news as they were also identified as topics for diaries (travel significantly more so than news) (Fothergill, 1976; Motz, 1987). However, we were not able to gain reliability on either measure despite months of training. This was primarily due to the lack of contextual cues in the tweets. Nevertheless, while we were unable to account for travel as a single category, we were able to account for it within "activities outside of home and work".

Concluding thoughts

Despite these limitations, the fact that we were able to account for almost 70% of tweets with our historically-derived topics, and that accounting and commentary-style tweets constituted 95% of our sample suggests that perhaps the kinds of things that people wrote about historically are indeed similar to the kinds of things people write about today. The mundane what-I-had-for-breakfast tweets can be better understood by placing them into a historical context. The chronicling and sharing of everyday events was historically and is today a means of reinforcing social connections with others. The commentary we find on Twitter today suggests its potential ability to give voice to those who may not have other outlets for expression. Indeed the diary has often been an outlet for women and minorities to find and develop their voice in society (Culley, 1986; Sjödblad, 1998). Today African Americans and Latinos have higher rates of adoption of Twitter than whites (Smith, 2011), suggesting perhaps that Twitter like historical diaries may be an important outlet for minority voices in the US. Future research should specifically examine minority uses of Twitter and other social media.

This study contributes to a growing literature that seeks to historicize contemporary communication practices and broaden our definitions of media technology (e.g. Bolter & Grusin, 2000, Gitelman, 2006; Milne, 2010). Rather than condemning Twitter as a narcissistic (Sarnow, 2009), by placing it into a longer discussion of media and communication we can begin to

understand its popularity. While there are important differences regarding breadth of audience

and degree of interactivity between Twitter and historical diaries, the similarities to historical

diaries

Table 1: Reliability Coefficients

	Variable	Kappa
Tweet Subject	First Person Singular	.91
	Third Person Singular	.85
	Third Person Plural	.74
	Audience	.74
	None	.86
Tweet Topic	Weather	.74
	Family	.72
	Media	.81
	Food/Beverage	.91
	Religion	.72
	Health	.78
	Sleep	.81
	Activities	.79
	Home Activities	.73
	Work Activities	.78
	Outside of Home & Work Activities	.80
Tweet Style	Accounting	.71
	Commentary	.78
	Information Seeking	.73
	Content Sharing	.80
	Response	.71

Table 2: Frequencies

	Variable	Percent of all Tweets
Tweet Subject	First Person	66.4% (n=1,375)
	Third Person Singular	22% (n=457)
	Third Person Plural	11.9% (n=246)
	Audience	18.8% (n=390)
	None	14.9 (n=310)
Tweet Topic	Activities	41% (n=852)
	Home Activities	5% (n=108)
	Work Activities	6% (n=128)
	Outside of Home & Work Activities	15% (n=313)
	Media	35.5% (n=735)
	Food/Beverage	9% (n=182)
	Health	7% (n=137)
	Weather	3.9% (n=308)
	Sleep	3% (n=68)
	Family	2.8 (n=57)
	Religion	1%(n=14)
Tweet Style	Commentary	75% (n=1,554)
	Accounting	62% (n=1,286)
	Content Sharing	11% (n=222)
	Information Seeking	4% (n=83)
	Response	4% (n=82)

Total Tweets coded =   2,070

Table 3: Tweet Subject

		Other*		
		Absent	Present	
First person	Absent	27.7% (n=310)	40.7% (n=387)	33.7% n=697
	Present	72.3% (n=810)	59.3% (n=563)	66.3% n=1373
	Total	100% n=1,120	100% n=950	100% n=2,070

* Other was determined by whether the tweet mentioned either Audience or third person singular or plural. Percentages based on columns.

Table 4: Tweet Subject & Tweet Style Correlations:

	Commentary	Update	Content Sharing	Information Seeking	Response
1st person	-.06**	.40***	-.15***	-.12***	-.07***
3rd person singular	-.01	.01	.02*	-.01	.02
3rd Person plural	.04	.000	.01	-.01	-.02
Audience	.04	-.22***	.02	.25***	.32***
None	.06**	-.22***	.10***	-.05*	-.06**

* Correlation is significant at the 0.05 level (2-tailed)

** Correlation is significant at the 0.01 level (2-tailed)

*** Correlation is significant at the 0.001 level (2-tailed)

References

Barnouw, E. (1968). *A history of broadcasting in the United States: the Golden Web, 1933 to 1953*. Oxford, UK: Oxford University Press.

Bloom, L. Z. (1996). "I write for myself and strangers": private diaries as public documents. In S. Bunkers & C. Huff (Eds.), *Inscribing the daily: Critical essays on women's diaries* (pp. 23-37). Amherst, MA.

Bolter, R., & Grusin, J. D. (2000). *Remediation: Understanding new media*. Cambridge, MA: MIT Press.

boyd, d., Golder, S., & Lotan, G. (2010). Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. In *Proceedings of HICSS'10*. New York, NY: ACM Press.

Brumberg, J. J. (1997). *The body project: An intimate history of American girls*. New York: Random House.

Carr, N. (2009). Woodrow Wilson was the First Twitterer: The New York (Real) Times. *Encyclopedia Britannica Blog*. Retrieved from:

Clark-Flory, T. (2009, September 26). Tweeting a Miscarriage.   Retrieved December 30, 2009, from

http://www.salon.com/life/broadsheet/feature/2009/09/26/miscarriage_tweet/index.html.

Culley, M. (Ed.). (1985). *A day at a time: The diary literature of American women from 1764 to the present*. New York: The Feminist Press at the CUNY.

Culley, M. (1989). "I Look at Me": Self as Subject in the Diaries of American Women. *Women's Studies Quarterly, 17*(3 & 4), 15-22.

Ede, L., & Lunsford, A. (1984). Audience Addressed/Audience Invoked: The Role of Audience in Composition Theory and Pedagogy. *College Composition and Communication, 35*(2), 155. doi:10.2307/358093

Fothergill, R. A. (1974). *Private chronicles: A study of English diaries.* London: Oxford University Press.

Gitelman, L. (2006). *Always already new: Media, history, and the data of culture.* Cambridge, MA: MIT Press.

Gruzd, A., Takhteyev, Y., and Wellman, B. (2011). Imagining Twitter as an Imagined Community. *American Behavioral Scientist*, *55*(10): 1294-1318.

Haddon, L. (2006). Empirical studies using the domestication framework. In T. Berker, M. Hartmann, Y. Punie & K. J. Ward (Eds.), *Domestication of media and technology* (pp. 103-122). Maidenhead: Open University Press.

Halliday, J. (2011). Guardian Activate 2011 Conference Live Blog. Guardian Activate 2011 Conference. Retrieved 2011, from http://www.guardian.co.uk/media/pda/2011/apr/28/guardian-activate-2011-new-york

Hayes, A., & Krippendorff, K. (n.d.). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*(1), 77-89.

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

Hunter, J. H. (1992). Inscribing the self in the heart of the family: Diaries and girlhood in late-Victorian America. *American Quarterly, 44*(1), 51.

Java, A., Song, X., Finin, T., & Tseng, B. (2007, August 12). Why we Twitter: Understanding microblogging usage and communities. Paper presented at the Joint 9th WEBKDD and 1st SNA-KDD Workshop '07, San Jose, CA.

John, R. R. (1998). *Spreading the news: the American postal system from Franklin to Morse.* Cambridge, MA: Harvard University Press.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology.* Thousand Oaks, CA: Sage.

Lombard, M., Snyder-Duch, J. and Bracken, C. C. (2002), Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability. *Human Communication Research, 28*(4): 587–604. doi: 10.1111/j.1468-2958.2002.tb00826.x

Marvin, C. (1988). *When old technologies were new.* Oxford: Oxford University Press.

Marwick, A. E., & Boyd, D. (2010). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society, 13*(1), 114-133. doi:10.1177/1461444810365313

McCarthy, M. (2000). A pocketful of days: Pocket diaries and daily record keeping among Nineteenth-Century New England women. *The New England Quarterly, 73*(2), 274-296. doi:10.2307/366803

McCarthy, M. (2010). Redeeming the almanac: Learning to appreciate the iPhone of early America. *Common-Place, 11*(1): Common Reading. Retrieved from http://www.common-place.org/vol-11/no-01/reading/

Milne, E. (2010). *Letters, postcards, and email: technologies of presence.* New York: Routledge.

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

Morris, M. R., Teevan, J., & Panovich, K. (2010). What do people ask their social networks, and why ? A survey study of status message Q & A behavior. In *CHI'10 Proceedings*. New York, NY: ACM Press. doi:10.1145/1753326.1753587

Motz, M. F. (1987). Folk expression of time and place 19th-Century midwestern rural diaries. *Journal of American Folklore, 100*(396), 131-147.

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

Niles, R. (2009, June 25). Michael Jackson's death and its lessons for online journalists covering breaking news. *Online Journalism Review*. Retrieved October 23, 2011, from http://www.ojr.org/ojr/people/robert/200906/1755/

Ong, W. J. (1975). The writer's audience is always a fiction. *PMLA*, 90(1), 9-21.

O'Sullivan, C. (2005). Diaries, on-line diaries, and the future loss to archives; or blogs and the blogging bloggers who blog them. *The American Archivist, 68*(Spring/Summer), 53-73.

Park, D. W., & Pooley, J. (2008). *The history of media and communication research: Contested memories.* New York: Peter Lang.

Rosenzweig, R., & Thelen, D. P. (1998). *The presence of the past: Popular uses of history in American life, Volume 2*. New York: Columbia University Press.

Sarnow, D. (2009, March 11). On Twitter: Mindcasting is the new lifecasting. *LATimes.com*. Retrieved from: http://latimesblogs.latimes.com/technology/2009/03/on-twitter-mind.html

Sennet, R. (1992). *The fall of the public man.* New York, NY: Norton & Co.

Simmons, C., & Perlina, N. (2002). *Writing the siege of Leningrad: Women's diaries, memoirs, and documentary prose.* Pittsburgh: University of Pittsburgh Press.

Sjödblad, C. (1998). From family notes to diary: the development of a genre. *Eighteenth-Century Studies, 31*(4), 517-521.

Smith, A. (2011). *Twitter update 2011.* Pew Internet & American Life: Washington DC.

Smith, A., & Rainie, L. (2010). *8 % of online Americans use Twitter.* Pew Internet & American Life: Washington DC.

Tsotsis, A. (2011). With 4B SMS tweets processed monthly, Twitter now lets users share photos via text. *Techcrunch.* Retrieved October 30, 2011, from http://techcrunch.com/2011/09/21/with-4b-sms-tweets-processed-monthly-twitter-now-lets-users-share-photos-via-text/

Ulrich, L. (1991). *A midwife's tale: the life of Martha Ballard, based on her diary, 1785-1812.* New York: Vintage Books.