

ORIGINAL ARTICLE

Historicizing New Media: A Content Analysis of Twitter

Lee Humphreys¹, Phillipa Gill², Balachander Krishnamurthy³, & Elizabeth Newbury¹

1 Department of Communication, Cornell University, Ithaca, NY, USA

2 Department of Computer Science, Stoneybrook University, Stoneybrook, NY, USA

3 AT&T Labs Research, Florham Park, NJ, USA

This paper seeks to historicize Twitter within a longer historical framework of diaries to better understand Twitter and 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 styles are the most popular narrative styles on Twitter. Despite important differences between the historical diaries and Twitter, this analysis reveals long-standing social needs to account, reflect, communicate, and share with others using media of the times.

doi:10.1111/jcom.12030

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, 1968). More recently, scholars have looked to historical communication systems to reveal insights into contemporary communication issues (Bolter & Grusin, 2000; Gitelman, 2006; John, 1998; Marvin, 1988; McCarthy, 2010; Milne, 2010; Park & Pooley, 2008). Many have critiqued an overemphasis on the new, suggesting that all media engage with older media and social practices (Bolter & Grusin, 2000; Gitelman, 2006; Marvin, 1988). In this research vein, this study seeks to historicize contemporary microblogging practices, focusing on the service Twitter.

Twitter is a popular microblogging service that allows people to share updates, news, and information (known as “tweets”) with people in their Twitter network and beyond. With over 200 million registered users (Halliday, 2011) and 15% of online Americans using Twitter (Smith & Brenner, 2012), Twitter is one of the most popular social media. Research has compared the content of Twitter to earlier kinds

Corresponding author: Lee Humphreys; e-mail: lmh13@cornell.edu

of social media like blogs (Java, Song, Finin, & Tseng, 2007) and social network sites (Gruzd, Takhteyev, & Wellman, 2011). Recent analyses 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.

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 predecessors, earlier technologies for personal writing may help us to better understand Twitter's popularity and use today. Many have argued that the everyday writings of men and women not only reflect cultural production but also produce culture itself (Hansen, 1994; Miller, 1998; Williams, 1966). By situating Twitter as a form of everyday writing and within a longer historical framework of diaries, we seek to better understand not only Twitter but broader communication systems, practices, and patterns.

While today the term "diary" might conjure up thoughts of locked notebooks where people pour their innermost 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, some diaries, particularly from antebellum period in the United States, were written with the intent to be shared and were made up of relatively brief writings (Culley, 1989; Hansen, 1994). It is these kinds of writings that warrant a closer comparison with contemporary microblogs.

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 of tweets to more systematically ground our historicization in the actual content of the Twitter messages themselves (Krippendorff, 2004). 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.

Systematic analyses of small everyday writings can reveal large insights into the larger cultural milieu (Miller, 1998). Indeed, we argue that the precedents we see in historical writings suggest that the writing and sharing of everyday events may serve a social purpose. Publicly sharing the chronicling of life may contribute to the social bonds between people and communities. The diary and Twitter provide a platform

for people who might traditionally be excluded from public discourse to have a voice in representing themselves as well as their perceptions of the world around them. By analyzing the content of Twitter based on literature of historical personal writings, we aim to understand how people account, reflect, and share with others using media of the times.

Literature review

English language diaries from the 17th, 18th, and 19th centuries were often semipublic in nature (Bloom, 1996; Culley, 1989; Fothergill, 1974), sometimes considered a kind of public record. Some diaries, particularly those by women, were often shared during the diarist's lifetime to maintain family and social relations (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 (Fothergill, 1974). Sometimes husbands or wives 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 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 to improve 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 where we might share a scrapbook or a photo album with a visiting friend, historically diaries were often meant to be shared with visitors as a way of documenting and sharing important events in the family and community.

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 suggests Twitter users not only write to the people following them on Twitter but also write to themselves (Marwick & Boyd, 2010). Like historical diarists, some Twitter users report writing to express themselves and not necessarily to an intended or known audience. Researchers also suggest that the mere act of writing and keeping a diary implies an audience (Ede & Lunsford, 1984; Ong, 1975; Rosenweig & 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. How exactly Twitter users articulate an audience within their tweets has yet to be systematically analyzed.

Limited length

One of the distinguishing technological affordances of microblogging is the limited length of messages. Twitter is limited to 140 characters. Historical parallels suggest that technological limitations are not new and 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 in New England (McCarthy, 2000). About 2 × 4 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. 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 middle-class women's diaries during the 18th and 19th centuries often focused on life events such as births, deaths, marriages, travels, visits, illnesses, and work (Hansen, 1994; Sjöblad, 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 (e.g., Hansen, 1994; Sennet, 1992) and the content of the diaries of the time often reflects this. Particularly for women whose work occurred

in the domestic sphere, the home was not a separate private sphere (Hunter, 1992). Hansen's (1994) research on antebellum New England argues that a separate social sphere existed outside of the public-private dichotomy in which much historical activity occurred. For example these kinds of activities are reflected in Elizabeth Fuller's 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 today.

Sept 8—I spun three skeins today.

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 spheres. The historical chronicling of social events as well as travel and visits demonstrate Hansen's (1994) argument that such activities were not in the public or private sphere, but occurred in the social sphere. Particularly for the working and middle classes, the social sphere was characterized by communal activities centered around the church that brought people together and through the exchange of social labor, such as caring for the sick, visiting, and exchanging work (Hansen, 1994). It is important to note that according to Hansen's research, the social sphere of this time was gender-integrated. Indeed forgoing the public-private dichotomy may prove quite helpful for understanding practices on Twitter.

Analyses of Twitter suggest the majority of messages describe both personal and professional topics (Java et al., 2007; Naaman, Boase, & Lai, 2010; Smith & Rainie, 2010). Interviews with Twitter users suggest people are aware of multiple audiences for their tweets and self-censor their topics accordingly, avoiding information that may be too personal or too controversial (Marwick & Boyd, 2010). Indeed participants 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. Similarly, 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 typically 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 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 diarists 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, MA:

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 accounts 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 a miscarriage in an equally terse narrative (Clark-Flory, 2009). Twitter has become a place for 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.

Accounting diaries differ stylistically from more reflective and introspective diaries (Culley, 1989; Sjödblad, 1998). Reflective diaries, especially for women and young girls, became a place for them to have a voice and provided them an opportunity to discuss and explore their inner thoughts (Brumberg, 1997; Culley, 1985, 1989; Simmons & Perlina, 2002). Victorian diaries by girls were considered 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 their actions. These reflective diaries often focused not only on the activities of the writer but on the writer's own consciousness (Culley, 1989). The reflective narrative style in diaries often communicated thoughts, beliefs, and perspectives of the diarist him-or herself and often included a social or moral evaluation as well.

Research suggests that users acknowledge different purposes for writing tweets that may influence what as well as how they write (Marwick & Boyd, 2010). That said, the narrative style on Twitter has yet to be examined on a content level. Therefore this study seeks to explore how the narrative styles on Twitter resemble early accounting diaries or more reflective diaries.

Method

This study examined Twitter by conducting a content analysis of publicly available tweets that was derived from historical diary literature. Historically some diarists wrote only about themselves while others wrote about their friends, family, and community. Therefore, in this study we asked: (a) Who are the actors identified in tweets? Topics of diaries have ranged from religion to work to health to mundane everyday activities and events to introspective and contemplative thoughts. Thus,

we asked: (b) What is the prevalence of various diary topics on Twitter, specifically those topics commonly found in middle-class accounting diaries? Along with various topics, the narrative style of historical diaries includes daily chronicling or accounting as well as more introspective exploration and commentary. In addition to accounting and reflective styles, the interactive nature of Twitter can include new functions that may not have been prevalent or possible in historical diaries, such as information seeking and web content sharing (Naaman, Boase, & Lai, 2010). Therefore, we asked: (c) What is the prevalence of functions or style of tweets? Together these three research questions guided the content analysis.

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 Krishnamurthy, Gill, and Arlitt (2008). In total, we collected information for 101,069 tweets (message and user). In the combined samples ($N = 101,069$), we were able to get URL data for two-thirds of the sample, from which country of origin may be interpreted. After the United States, the top-10 countries were Japan, Germany, United Kingdom, Brazil, Holland, France, Spain, Belgium, Canada, and Italy. We then distinguished between tweets which had been submitted via the website ($n = 60,673$) and tweets submitted via text message ($n = 7,568$) to see if tweet mode influenced content, in the way pocket diaries differed from larger book diaries. The remainder of the sample was submitted via custom applications like Twitterific ($n = 24,387$), instant message ($n = 7,388$), and Facebook ($n = 1,053$). We randomly sampled 1050 web and 1050 text message English-language tweets to include in our content analysis. This random sample included 30 tweets that were half in English and half Spanish, so our final sample included 1024 web tweets and 1046 text tweets. We report only those statistically significant differences.

Measures

To answer our research questions, we analyzed Twitter messages along three main criteria: (a) who was talked about in the tweet (Tweet Subject), (b) the topic of the tweet (Tweet Topic), and (c) the narrative style of the tweet (Tweet Style) (see Table 1 for codebook).

Tweet subject

Historically, there were shifts 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 or people being discussed in the message.

Table 1 Reliability Coefficients

Variable	Alpha	Example
Tweet subject		
First person singular	.91	<i>I love twitter</i>
Third person singular	.85	<i>Gloria loves twitter</i>
Third person plural	.74	<i>Teachers love twitter</i>
Audience	.74	<i>Do you guys love twitter?</i>
None	.86	<i>Twitter is so slow</i>
Tweet topic		
Weather	.74	<i>It's snowing in April</i>
Family	.72	<i>My dad loves baseball</i>
Media	.81	<i>Just watched American Idol</i>
Food/beverage	.91	<i>Enjoying a beer on the porch</i>
Religion	.72	<i>I'll keep you in my prayers</i>
Health	.78	<i>I have a splitting headache</i>
Sleep	.81	<i>It's 4am and I should be sleeping</i>
Activities	.79	<i>Just updated my blog</i>
Home activities	.73	<i>Making lasagna for dinner</i>
Work activities	.78	<i>Finished the report for my boss</i>
Outside of home & work activities	.80	<i>Running errands at the mall</i>
Tweet style		
Accounting	.71	<i>Drivin' to work</i>
Commentary	.78	<i>I hate "family" dinners</i>
Information seeking	.73	<i>What's the best pizza in NYC?</i>
Content sharing	.80	<i>http://tinyurl.com/34abql</i>
Response	.71	<i>@cutegirl sure thing!</i>

A single Twitter message could have multiple subjects. For example, the Tweet Subject for the tweet, "going to the McDonalds with Sue" would be coded as first person (i.e., author) and third person singular (i.e., Sue). Tweet subjects were coded as 1st Person, 3rd Person Singular, 3rd Person Plural, Audience, and None (see Table 1 for examples of each code).

First person subjects were indicated with personal pronouns such as "I" or "me" or "we," or when 1st person was implied but not directly stated. Tweets were coded as *3rd person singular* if a real (nonfictional) person's specific name, username, title, or initials were 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 acknowledged an audience within the tweet itself. 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. Messages without subjects or whose subjects were animals or fictional characters were coded as *none*.

Tweet topic

Tweets were also coded for several topical areas. Based on the historical diaries, we coded for several “mundane” topics. Multiple topics could occur in one tweet, but not every tweet included a coded topic. We coded tweets as *weather* if they directly commented on or mentioned weather or weather-related conditions such as snow or temperature. We coded tweets as *family* if they directly mentioned one’s family relation such as sister, husband, child, etc., or the 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. 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 tweet’s 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 *media* if they mentioned media or communication technology, including: TV, movies, music, internet, computers, mobile phones, digital gaming (e.g., Xbox, World of Warcraft), Twitter (or Twitter-related word), books, magazines, and comic books. URLs in and of themselves 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 (Sjöblad, 1998). Activities were defined as tweets about an action, about doing something, and about being somewhere. Cognitive activities (e.g., wanting, thinking, and missing) were not coded as activities. 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 tweets that discussed activities occurring in or around the home such as cooking, cleaning, laundry, gardening, home or vehicle repair, as well as preparing for or being en route for home. *Work-related activities* discussed activities related to work or school such as 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 would be coded into multiple sub areas if appropriate. For example, “working from 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 the convergent and ubiquitous nature of contemporary media use.

Tweet style

The style of a tweet was defined as its broad purpose or form. This was loosely based on the various functions or styles of diaries over the centuries (Motz, 1987). A single Tweet could have multiple 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 reported on or shared 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 a reflection, emotion, opinion, or evaluation. Tweets were coded as *content sharing* if they contained content from other websites, authors, or sources. 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.

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 4 hours per week for between 2 1/2 months to 5 months on 628–950 messages outside of the final sample 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. Specifically, the inclusion of the narrative style “response” emerged during reliability training. When coders had reached acceptable levels of reliability (Krippendorff’s $\alpha < 0.70$) (Lombard, Snyder-Duch, & Bracken, 2002), they coded the random samples of web and text tweets ($n = 2070$) 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%, we calculated Krippendorff’s Alpha for each category to ensure acceptable levels of inter coder reliability (Hayes & Krippendorff, 2007; Lombard et al., 2002). Values ranged from 0.71 to 1 (see Table 1). All discrepancies of the multicoded tweets (i.e., the reliability sample) were discussed and consensus coded before including them in the analysis.

Table 2 Frequency of Subjects, Topics, and Styles by Text Versus Web-Based Tweets

Variable	Percent of All Tweets (<i>N</i> = 2,070)	Percent of Text Message Tweets (<i>N</i> = 1,046)	Percent of Web-Based Tweets (<i>N</i> = 1024)
Tweet subject			
First person	66	65	67
Third person singular	22	22	22
Third person plural	12	11	12
Audience	19	17*	22*
None	15	15	15
Tweet topic			
Activities	41	42	40
Home activities	5	4**	7**
Work activities	6	5*	7*
Outside of home & work activities	15	22***	8***
Media	35	27***	44***
Food/beverage	9	10**	7**
Health	6	7	5
Weather	4	4	4
Sleep	3	2*	4*
Family	3	2	3
Religion	1	1	1
Tweet Style			
Commentary	75	75	75
Accounting	68	62***	74***
Content sharing	11	3***	18***
Information seeking	4	4	4
Response	4	4	4

Note: Chi-squares were calculated for text and web-based tweets.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Findings

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%, $n = 1,375$) (see Table 2). 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 an 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 less likely to identify audience than tweets submitted via the web, $\chi^2 (1, n = 390) = 5.612, p < .05$.

Additionally, we found that 46% of all tweets ($n = 950$) involved someone other than the author him- or herself. Tweets subjects were not mutually exclusive, so while 66% of tweets involved the first person, 41% of those tweets ($n = 387$) also involved another person in some way (either as audience, third person singular, or plural) in addition to the first person. These suggest 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 groups in their tweets.

Tweet topic

On the basis of the topics identified from the historical diaries, we were able to code 68% of our sample ($n = 1,415$) for the tweet topic (see Table 2). The most frequent topic was activities, making up 41% ($n = 852$) of our sample. We were only able to further sub code 61% ($n = 521$) of the activities into home activities, work activities, or activities outside of home or work. The most popular kind of activity tweet was outside of home and work ($n = 310$), which was more likely to be sent via text message than through the website, $\chi^2 (1, n = 310) = 79.454, p < .001$. However, both home and work activities were more likely to be submitted via the web, respectively $\chi^2 (1, n = 108) = 9.479, p < .01$ and $\chi^2 (1, n = 128) = 6.235, p < .05$. Media tweets, the second most popular topic in our sample, were also more likely to be sent via the web than through text message, $\chi^2 (1, n = 734) = 59.481, p < .001$; and were also a common activity in our sample. 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 24% of our sample ($n = 491$). Religion was the least frequent topic of the tweets ($n = 14$), while food and beverage tweets ($n = 182$) were the most frequent after activities and media. Food and beverage tweets were also more likely to be submitted via text message than the web, $\chi^2 (1, n = 182) = 7.837, p < .01$; while sleep tweets were more likely to be submitted via the web than the mobile, $\chi^2 (1, n = 68) = 6.530, p < .05$.

Tweet style

Overwhelmingly, most tweets involved some sort of commentary or reflection 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 were each 4% of the sample. Content sharing and accounting were more likely to be submitted via the web than through text message, $\chi^2 (1, n = 222) = 123.371, p < .001$ and $\chi^2 (1, n = 1414) = 33.781, p < .001$ respectively.

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 the content of a random sample of publicly available tweets inspired by various characteristics of historical diaries that historians have identified (Culley, 1989; Fothergill, 1974). In particular, we explored the resemblance of the actors identified in the tweet, the prevalence of various historical diary topics on Twitter today, and the narrative style of tweets.

Overwhelmingly tweets in our sample were about people. Often they were about the author him-or herself, but this was certainly not always the case. The focus of diaries has varied historically (Fothergill, 1974). Very early religious diaries and Victorian era diaries focused primarily on the author him-or herself, while more social accounting diaries focused more on the diarist's activities and interactions with others (Culley, 1985; Fothergill, 1974; Ulrich, 1991). Our data suggest that Twitter tends to combine elements of historical diaries in terms of who is written about. In our sample, the most prevalent kind of actor identified in the tweet was the Twitterer him or herself (first person); however, Twitter is not only a place for people to talk about themselves. Indeed almost half of the tweets in our sample involved someone besides the author him-or herself, whether it be another person or group of people including members of the Twitter community. Only 15% of tweets did not include any mention or implication of the person 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 thoughts and development nor those diaries or ledgers that marked work-related tasks only but did not involve any people (Fothergill, 1974). The social accounting diaries provide a helpful background because they help to illuminate what it means to share short, mundane messages. Writing down the daily short messages not only chronicles one's experiences and those of the community, but the social chronicling reflects the fact that we are part of communities that share meaning and reproduce our connections through everyday practice and routine. These findings suggest microblogging is not merely social in its allowance for the presentation of self (Hogan, 2010), but social in its constitution and communication as well.

Some historical diaries focused on everyday events and activities rather than the emotional or spiritual development of the diarist (Culley, 1985). 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. Conservatively, we were only able to sub code a portion of activities into work, home, and outside of work and home. However, our inability to confidently code so many of our activities speaks to the continued blurring of public, private, and social spheres in our contemporary environment.

While 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 historical diaries that merely 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 (Culley, 1985). It seems that the majority of tweets combine the narrative styles of different kinds of historical diaries. This reinforces findings that social media allow for curation (Hogan, 2010), where people share information through these services by choosing which information to share but also by adding their own comment to it.

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 important difference is the degree of interactivity on Twitter compared to diaries. While there is evidence that people sent diaries back and forth and would even write in each other's diaries (Culley, 1985; Sjödblad, 1998), the near simultaneity of Twitter affords much greater interactivity than diaries. This difference can facilitate vast amounts of feedback more quickly, as evidenced in the presence of information-seeking and response tweets. While these styles of tweet were the least 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). That said, just because Twitter affords greater interactivity than diaries does not mean that it necessarily is more interactive than diaries. Indeed some research suggests that most tweets are ignored (Geere, 2010).

Another important 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 (e.g., Golder & Macy, 2011). While the diarists may have been socially linked within communities, the diaries themselves were likely not connected. 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 subject, topics, and narrative styles on Twitter, we also explored how the mode of tweeting (web or text) was related to the content and style of tweets. The differences we found here suggest that the context of tweeting may influence the kinds of tweets people write, much like pocket diaries did historically. The fact

that accounting was also more likely to occur via the web may reflect use of Twitter within routinized patterns of daily life. Indeed large data analyses of Twitter suggest strong temporal patterns in use (Golder & Macy, 2011; Honeycutt & Herring, 2009). The fact that content sharing was more likely to occur via the web is not surprising due to the difficulty of sharing URLs via a text message, especially if one has to transform it into a short URL to fit into the 140-character limit. The relationship we found between tweet mode 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 media 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-messaged tweets is still an important comparison.

It is important to note, however, 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, 1974) suggests further a long-standing desire to chronicle new events and experiences and to share them with others. Future research should further examine how tweets sent via mobile applications differ in content and style from other modes to explore how technological affordances, context, and norms differ.

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 how often tweets mention the audience or share content may not reflect current Twitter practices. It is likely that current trends in retweeting would be underestimated based on our findings regarding content sharing in 2008.

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, 1974; 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.”

From a methodological perspective it might have been ideal to have an equivalent diary sample; however, this was not logistically possible. While there are extensive databases of diaries (such as the New American Women’s Letters and Diaries), collections often span many years and even centuries. While we took a random

sample of the tweets in our study, there is no equivalent corpus from which to sample English diaries. Additionally, randomization will not correct for the fact that certain diaries are more likely to be saved over time than others, thus quantification of existing diaries is less helpful. Lastly, our unit of analysis is the tweet itself and not the Twitterer. Given our sample, it would have been difficult to analyze random specific diaries' entries. Future research might compare a Twitterer's archive of tweets over their lifetime with a diarist's entries as they may be more directly comparable. In addition, this kind of close textual analysis is more methodologically congruent with historical and archival methods from which most of the historical diary literature comes. Nevertheless, drawing on historical literature regarding diaries and diary practices provides a helpful framework for beginning to historicize and understand Twitter though further analyses should certainly be done.

Despite these limitations, the fact that we were able to account for almost 70% of our sample of tweets with our historically derived topics, and that accounting and commentary-style tweets constituted 95% of our sample suggests that 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 this 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, 1985; Sjödblad, 1998). Today African Americans and Latinos have higher rates of adoption of Twitter than Whites (Hargittai & Litt, 2011; Smith, 2011), suggesting perhaps that Twitter, like historical diaries, may be an important outlet for minority voices in the United States. 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 deepen our understandings of media technology (e.g., Bolter & Grusin, 2000; Gitelman, 2006; Milne, 2010). Rather than condemn the accounting and reflecting practices on Twitter as narcissistic (Sarnow, 2009), by placing them into a longer discussion of media and communication we can begin to understand Twitter's popularity. While there are important differences regarding breadth of audience and degree of interactivity between Twitter and historical diaries, the similarities to historical diaries suggest long-standing social needs to account, reflect, communicate, and share with others.

Acknowledgments

The authors would like to thank the New Media & Society group at Cornell for comments on an earlier version of the article. The authors would also like to thank the editor and reviewers for their insightful comments.

References

- Barnouw, E. (1968). *A history of broadcasting in the United States: The golden web, 1933 to 1953*. Oxford, England: 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: The University of Massachusetts Press.
- 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*. Proceedings of HICSS-43. Washington, DC: IEEE.
- Brumberg, J. J. (1997). *The body project: An intimate history of American girls*. New York, NY: Random House.
- Clark-Flory, T. (2009, September 26). Tweeting a miscarriage. Retrieved from <http://www.salon.com>.
- Culley, M. (Ed.) (1985). *A day at a time: The diary literature of American women from 1764 to the present*. New York, NY: 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, England: Oxford University Press.
- Geere, D. (2010, October 11). It's not just you: 71 percent of tweets are ignored. *Wired*. Retrieved from www.wired.com.
- Gitelman, L. (2006). *Always already new: Media, history, and the data of culture*. Cambridge, MA: MIT Press.
- Golder, S. & Macy, M. (2011). Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333(6051), 1878–1881. doi: 10.1126/science.1202775
- Gruzd, A., Takhteyev, Y. & Wellman, B. (2011). Imagining Twitter as an imagined community. *The American Behavioral Scientist*, 55(10), 1294–1318. doi: 10.1177/0002764211409378
- Halliday, J. (2011). Guardian activate 2011 conference live blog. Guardian Activate 2011 Conference. Retrieved from <http://www.guardian.co.uk/media/pda/2011/apr/28/guardian-activate-2011-new-york>
- Hansen, K. V. (1994). *A very social time: Crafting community in Antebellum New England*. Berkeley, CA: University of California Press.
- Hargittai, E. & Litt, E. (2011). The tweet smell of celebrity success: Explaining variation in Twitter adoption among a diverse group of young adults. *New Media & Society*, 13(5), 824–842. doi: 10.1177/1461444811405805
- Hayes, A. & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1), 77–89.
- Hogan, B. (2010). The presentation of self in the age of social media: Distinguishing performances and exhibitions online. *Bulletin of Science, Technology & Society*, 30(6), 377–386.

- Honeycutt, C. & Herring, S. (2009). *Beyond microblogging: Conversation and collaboration via Twitter*. Proceedings of HICSS-42. Washington, DC: IEEE. doi: 10.1109/HICSS.2009.602
- 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. doi: 10.2307/2713180
- Java, A., Song, X., Finin, T., & Tseng, B. (2007, August). 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.
- Krishnamurthy, B., Gill, P. & Arlitt, M. (2008). *A few chirps about Twitter*. Proceedings of WOSN-08, 19–24. New York: ACM. doi: 10.1145/1397735.1397741
- Lombard, M., Snyder-Duch, J. & 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, England: 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/>
- Miller, S. (1998). *Assuming the position: Cultural pedagogy and the politics of commonplace writing*. Pittsburgh, PA: University of Pittsburgh Press.
- Milne, E. (2010). *Letters, postcards, and email: Technologies of presence*. New York, NY: Routledge.
- Moore, R. J. (2009, October 5). Twitter data analysis: An investor's perspective. Techcrunch. Retrieved from <http://www.techcrunch.com>
- 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. Proceedings of SIGCHI, 1739–1748. New York: ACM. 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. doi: 10.2307/540918
- Naaman, M., Boase, J., & Lai, C. H. (2010). Is it really about me? Message content in social awareness streams. Proceedings of CSCW2010, 189–192. New York: ACM. doi: 10.1145/1718918.1718953
- Niles, R. (2009, June 25). Michael Jackson's death and its lessons for online journalists covering breaking news. *Online Journalism Review*. Retrieved from <http://www.ojr.org>
- O'Sullivan, C. (2005). Diaries, online diaries, and the future loss to archives; or blogs and the blogging bloggers who blog them. *The American Archivist*, **68**(Spring/Summer), 53–73.
- Ong, W. J. (1975). The writer's audience is always a fiction. *PMLA*, **90**(1), 9–21. doi: 10.2307/461344

- Park, D. W. & Pooley, J. (2008). *The history of media and communication research: Contested memories*. New York, NY: Peter Lang.
- Rosenzweig, R. & Thelen, D. (1998). *The presence of the past: Popular uses of history in American life*. 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, PA: University of Pittsburgh Press.
- Sjödblad, C. (1998). From family notes to diary: The development of a genre. *Eighteenth-Century Studies*, 31(4), 517–521. doi: 10.1353/ecs.1998.0041
- Smith, A. (2011). *Twitter update 2011*. Washington, DC: Pew Internet & American Life.
- Smith, A. & Brenner, J. (2012). *Twitter use 2012*. Washington, DC: Pew Internet & American Life.
- Smith, A. & Rainie, L. (2010). *8% of online Americans use Twitter*. Washington, DC: Pew Internet & American Life.
- Tsotsis, A. (2011, September 21). With 4B SMS tweets processed monthly, Twitter now lets users share photos via text. Techcrunch. Retrieved from <http://techcrunch.com>
- Ulrich, L. (1991). *A midwife's tale: The life of Martha Ballard, based on her diary, 1785–1812*. New York, NY: Vintage Books.
- Williams, R. (1966). *Culture and society*. New York: Harper and Row.

Copyright of Journal of Communication is the property of Wiley-Blackwell and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.