Trusting Tweets: The Fukushima Disaster and Information Source Credibility on Twitter

Robert Thomson[1], Naoya Ito[2], Hinako Suda[1], Fangyu Lin[1], Yafei Liu[1], Ryo Hayasaka[1], Ryuzo Isochi[1], Zian Wang[1]
[1] Graduate School of International Media, Communication and Tourism Studies
[2] Research Faculty of Media Communication
Hokkaido University
{rob.thomson, naoya, hsuda}@imc.hokudai.ac.jp

ABSTRACT

This paper focuses on the micro-blogging service Twitter, looking at source credibility for information shared in relation to the Fukushima Daiichi nuclear power plant disaster in Japan. We look at the sources, credibility, and between-language differences in information shared in the month following the disaster. Messages were categorized by user, location, language, type, and credibility of information source. Tweets with reference to third-party information made up the bulk of messages sent, and it was also found that a majority of those sources were highly credible, including established institutions, traditional media outlets, and highly credible individuals. In general, profile anonymity proved to be correlated with a higher propensity to share information from low credibility sources. However, Japanese-language tweeters, while more likely to have anonymous profiles, referenced low-credibility sources less often than non-Japanese tweeters, suggesting proximity to the disaster mediating the degree of credibility of shared content.

Keywords

Anonymity, credibility, crisis communication, culture, deindividuation, Fukushima, social media, trust, Twitter.

INTRODUCTION

In April 2011, Hokkaido University researchers carried out a survey of 1,000 Japanese consumers, tapping attitudes towards and uses of various media directly after the devastating northeastern Japan earthquake of March 11th 2011 (Hokudai Earthquake Project, 2011). In one section, respondents were provided a list of information sources and asked to rate how reliable they felt those information sources were in the three days following the earthquake. Twitter - a global micro-blogging service which boasts more Japanese language messages out of any other language other than English - had the lowest rating, with a reliability rating of only 58.2 out of a possible 100 points (Figure 1).

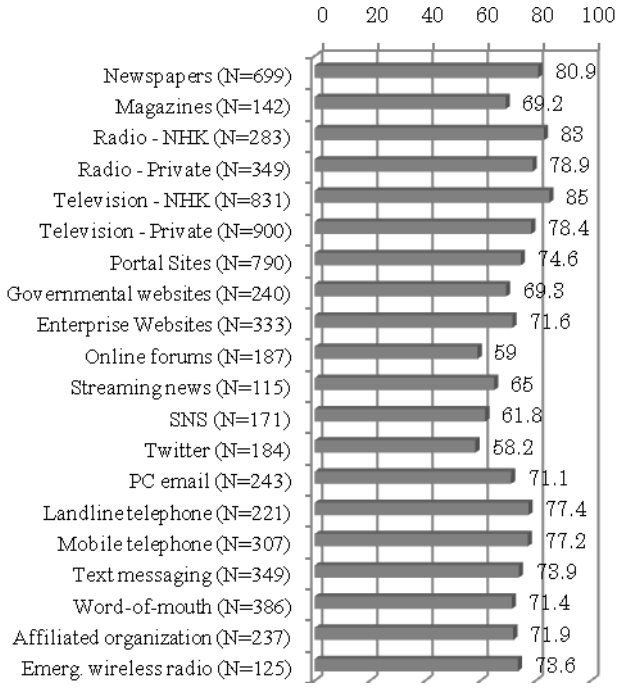
Echoing this public opinion are commentators such as Yu Ozeki, who hold that general panic surrounding disasters such as the great Japan earthquake and Fukushima Daiichi nuclear plant accident spreads into online spaces such as Twitter, bringing with it harmful unconfirmed rumors, a tide of un-useful information, and the drowning out of important information flowing from the disaster-hit areas (Ozeki, 2011). On the other hand, others hail Twitter as the next generation of innovation which mitigates many human casualties (Nishida, 2011); indeed, with much of the northeastern Honshu coast left without power and other essential infrastructure directly following the earthquake, a near-ubiquitous internet-enabled mobile phone ownership in Japan facilitated a rush of new users in Japan to Twitter for information (Saito, 2011).

Evidence from information credibility perception experiments, however, shows a definite tendency for people to view information shared on Twitter as less credible than other sources. Shmierbach and Oeldorf-Hirsch (2010) performed two separate studies to ascertain credibility perception of information on Twitter, compared with a newspaper website and anonymous blog. They found that the same piece of information, when posted according to the appropriate style afforded by each medium (long-form on the news website, short-form on the blog, and headline plus link only on Twitter), was rated as less credible when presented on Twitter.

Amongst this debate comes the obvious question: Is information shared on Twitter really that unreliable? To answer this central research question, an investigation was proposed whereby a sample of public tweets would be analyzed to establish source credibility. Of particular interest, from a *fuhyohigai* (collateral damage due to rumors) perspective, is information shared in relation to the Fukushima Daiichi nuclear power plant accident. That is to say,

despite little to no increase of background radiation in locales outside the mandated 30km exclusion zone around the nuclear plant (MEXT, 2011), such locales suffered collateral damage in the form of a substantial drop in tourist numbers, produce supply cuts, and misinformation (Hokudai Earthquake Project, 2011). Therefore, the credibility of information sources on Twitter regarding the Fukushima nuclear accident, both from within and without Japan, is naturally a point of widespread concern.

Figure 1. Please rate how reliable you think each of the following media were following the earthquake (N=1,000)



Respondents answered on a 5-point Likert scale where 0=totally unreliable, 5=totally reliable, score later converted to a 100-point scale (Hokudai Earthquake Project, 2011).

EVENT OF STUDY

Underpinning the global scale of the Fukushima Daiichi nuclear power plant accident is the nature and extent of the impact. The powerful magnitude 9.0 quake which hit off the coast of north-eastern Honshu in Japan on the 11th of March 2011 caused a succession of disasters, not the least of which was a massive tsunami engulfing much of the northeastern coast of Japan. Among the tsunami-affected areas on the coast was Okuma Town, home to the ageing Fukushima Daiichi nuclear power plant, located approximately 175km from the earthquake epicenter, and 240km from central Tokyo. The power plant's tsunami defense systems did not cope with the 15m high tsunami, causing auxiliary power systems to fail, rendering emergency coolant pumps inoperable (TEPCO, 2011). This led in turn to meltdowns in the 1st, 2nd and 3rd reactors, in turn causing considerable emissions of radioactive materials into the environment. Reaction to the ensuing disaster caused concern both within and outside of Japan: Exclusion zones around the plant affected up to 80,000 residents (The Japan Times Online, 2011a); radioactivity impacted the Japanese agriculture industry (NYTimes.com, 2011a, 2011b); fears over radioactive rain were voiced in Korea (Reuters, 2011); and anti nuclear policies around the world, in part responding to global protest, were hastily brought into question (The Japan Times Online, 2011b).

RELATED WORK

Research which touches on Twitter either as an information source, or on the information sources which comprise information on Twitter, are quickly increasing in number as social media use - especially during times of crisis - becomes more widespread. One important point which needs to be established from the outset is that during a disaster, the nature of the information environment extant on Twitter is different from times of relative calm. In their network analysis of over 300 million tweets in the weeks preceding and following the 2011 Japan earthquake, Toriumi et. al. (2011) found a stark increase in degree distribution directly after the earthquake, suggesting an increase in the importance of Twitter as an information source during a disaster. Indeed, Heverin and Zach (2010) found that during the 2009 police officer shootings in Washington, US, a large majority of tweets regarding the event were based on some form of third-party information (whether external links or re-tweets).

Regarding this "manifest" factor (Merton, 1968) of third-party information sharing - which arguably is a factor that has direct and obvious implications for the trustworthiness of information on Twitter - several researchers have looked specifically at information sources on Twitter during a disaster. When looking at information shared during the Red River Valley flooding in the US and Canada in April 2009, Starbird et. al. (2010) categorized tweets into

Source type		Ability to provide correct information	Intent to provide correct information	Credibility Level
Traditional Media (print, television, radio)		✓ Resources (human and physical), professionalism	✓ Longevity as business rests on provision of correct information	High
Public institutions		✓ Power of inquiry, Resources (human and physical), professionalism	✓ A charge to protect the public	High
Enterprises		✓ Resources (human and physical), professionalism	✓ Longevity as business rests on provision of correct information	High
Non-profit organizations		✓ Resources (human and physical), professionalism	✓ Providing a service	High
Individuals	**Freelance journalist**	✓ Take part in TEPCO press meetings, experience in reporting and research, professionalism	✓ Personal brand rests on providing correct information	High
	Academics, specialists, researchers	✓ Specialized knowledge, observations based on empirical data	✓ Status as professional	High
	Fukushima prefecture residents	✓ Local perspectives, on-the-ground reports	✓ Direct stakeholders in disaster outcomes	High
	Non- identified sources	✕ Unclear	✕ Unclear	Low
	Non-locals	✕ Reliance on second-hand information, amateurism	✓ Web citizenship	Low
	Conspiracy theorists	✓ Resources, time	✓ Strong conviction to convey truth	Low[1]
Alternative media		✕ Unclear	✕ Unclear	Neither of the above

Table 1. Credibility Assessment

three distinct categories: *original tweets*; *synthesis tweets* (defined as tweets which contain third-party information plus some form of commentary by the poster); and *derivative tweets* (whereby tweet and/or other third-party sourced information is passed on wholesale). They noted that in relation to the disaster, "official information remains important and is complemented, not usurped, by information generated by the public. People use and rely on official sources and other believable eyewitness accounts from which to source their information" (p. 9). This does not, however, suggest that Twitter is void of unhelpful information. In a questionnaire survey administered by Acar and Muraki (2011) and completed by 26 Miyagi prefecture residents who used Twitter following the 2011 Japan earthquake (Miyagi prefecture was one of the hardest hit locales in northeastern Honshu), users indicated that some widely re-distributed tweets containing unconfirmed rumors made the process of delineating helpful and un-helpful content difficult. Suggestions to improve reliability of information included the implementation of an official hashtag system in order to moderate the reliability of shared content on Twitter.

An area of social media use in relation to crisis communication which has not yet been explored, however, is that of "latent" factors such as cultural and psychological factors, which may affect source credibility. Due to the fact that the current event of study occurred in Japan, a characteristic of social media use in Japan which merits particular attention is that of anonymity online; online spaces in Japan are well known in cross-cultural studies to be characterized by anonymity (Barker & Ota, 2011; Takahashi, 2010; Thomson & Ito, forthcoming). Furthermore, anonymity in computer-mediated-communication has been shown to elicit a variety of behavioral outcomes, ranging anywhere from increased openness and communication (Chester & Gwynne, 1998) to a decrease in a sense of personal responsibility, leading to anti-social behavior (Davis, 2002). A central precept in such discussion is the concept of deindividuation (Postmes, Spears, & Lea, 1999); the shedding of social cues which would typically inhibit certain behaviors in the physical presence of others. The effect of anonymity on Japanese internet user online behavior is not conclusive however, ranging from cases of relatively benign effects, such as increased emphasis on

[1] While the category *conspiracy theorists* does, on paper, meet the criteria for highly credible information sources, the central conspiracy theory surrounding the great Japan earthquake (HAARP triggered earthquake theory proffered in Japan by Richard Koshimizu - http://richardkoshimizu.at.webry.info/) lacks imperfection, supposes an excessive rationality, and requires a "nihilistic degree of skepticism", which allows one to consider it to be an "unwarranted conspiracy theory" (Keeley, 1999).

diary writing in Mixi (one of Japan's leading social networking sites) (Barker & Ota, 2011), to severe criticism and flaming characteristic of 2-Channel, Japan's leading bulletin board site (Katayama, 2007).

Finally, a note regarding the evaluation of the credibility of information sources online is warranted. When evaluating the credibility of any given source of information, the evaluator is essentially entering into a form of persuasive communication. According to the Elaboration Likelihood Model of persuasion, an actor is persuaded by one of two routes to persuasion; a *central route* or a *peripheral route* (Petty & Cacioppo, 1986). If one is to take a central route to persuasion, one would assess information from a source in a logical way, evaluating the claims of the content of the communication (such as the truthfulness of the information) to determine the credibility of the source. The peripheral route, which is the default route for many when evaluating online information (Ohshima et al., 2011), involves turning to other cognitive clues to a source's credibility. Ohshima et. al., elaborating on Hovland (1953), suggest that such peripheral clues consist of "the degree to which a communicator is perceived to have adequate expert knowledge and trustworthiness" (Ohshima et al., 2011, p. 3). Yamagishi's (1998) treatise on the structure of trust elaborates on this further, whereby trust is conceptualized as an individual's perception towards an actor's *ability* and *intention* in a social exchange.

RESEARCH QUESTIONS

Taking the above related work into account, and if one considers that the active sharing of third-party information on Twitter is an act of recommending that information - thereby bestowing upon that information a degree of social value (Starbird & Palen, 2010) - then the obvious question which follows is:

RQ1 During times of crisis where the transfer of timely and correct information is of utmost importance, where are users obtaining their information, and are those sources credible?

Furthermore, in regards to potential effects of anonymity on user behavior online:

RQ2 Does anonymity on Twitter affect users' judgment when sharing third-party information during a crisis situation, and is this mediated by culture?

INVESTIGATIVE FRAMEWORK

For the current study several key concepts were developed in order to investigate the questions posed. First, echoing Starbird et. al.'s (2010) tweet categorization framework, the information environment extant on Twitter was conceptualized as being comprised of two distinct types of tweet: the *original tweet* and the *synthesis-derivative tweet*. Original tweets are defined as those which do not contain any obvious third-party information. That is, tweets of a generative nature; autobiographical narrative and commentary (Starbird et. al., 2010, p. 6). The synthesis-derivative tweet category includes synthesis tweets (third-party informational media "process[ed] and digest[ed] ... then redistribute[d] on the Twitter platform") and derivative tweets (re-tweets, re-sourced information, and links) (Starbird et. al., 2010, p. 7). In short, tweets in this single category include tweets which contain some form of reference to third-party information. This is a simplification of Starbird et. al.'s categorization scheme, adopted to simplify coding (i.e., a tweet has one of two states; third-party-sourced information present or not present). Naturally, information sources for synthesis-derivative tweets can be highly credible or otherwise.

Secondly, synthesis-derivative tweet sources are conceptualized as either *high* or *low* credibility. Taking an inevitable peripheral route to persuasion (inevitable owing to the immense breadth of topics, events, and subject matter referenced on Twitter, precluding a detailed logical evaluation of every single claim), Hovland's (1953) and Yamagishi's (1998) concepts of credibility and trust were employed, whereby a source's perceived ability and intention to provide correct information were taken as clues to their credibility. A full overview of this study's framework for assessing synthesis-derivative tweet source credibility can be found in Table 1.

It is worth noting at this point that the current study's measure of credibility is not without its limitations. What the framework in Table 1 essentially embodies is how a third-party observer might perceive the credibility (worthiness of trust (credibility, n.d.)) of a source, based on a combination of Hovland's and Yamagishi's concepts of credibility and trust. This falls short of a holistic concept of credibility however, in that it does not allow for the consideration of the development of credibility perception over time. That is to say, any particular Twitter user may be wary of, for example, information from "non-local individual source A" concerning the Fukushima accident, until their experience with that source deepens; source A may prove, despite initial impressions, to be a supplier of sound, truthful information (or otherwise). Indeed, Morris et. al. (2012) showed that considerably fewer users in their study

were concerned with the credibility of information originating from people they follow, compared with other methods of information discovery on Twitter (such as trending topics, Twitter search engine, and other search engines). In short, therefore, our framework of credibility assessment allows only for a detached, third-party appraisal of source credibility, as if tweets were encountered by users through trending topics, Twitter searches, and web searches; a form of "initial perceived credibility", if one will. Furthermore, for the evaluation of source credibility, ideological evaluations of credibility were relied upon, based on the ideological place of mass media, public institutions, and other institutions of knowledge in society. Equally conceivable is the possibility that any one individual's perception of, for example, public institutions' credibility, may deteriorate over time as the individual's experience with the institution deepens (such as, for example, when official radiation levels are not correctly announced in a timely manner). These limitations notwithstanding, our framework is nonetheless an attempt to attain a snapshot of what kind of information sources were cited in synthesis-derivative tweets, in order to better understand if the general perception of Twitter being less credible than other information sources is warranted, at least from the viewpoint of a point-in-time evaluation of initial perceived credibility (by a group of third-party observers).

By way of overview, Figure 2 below illustrates the two main factors this study investigates when addressing the central concern of information reliability on Twitter; that is, an appraisal of the manifest factor of source credibility, and an investigation into possible latent factors comprising psychological components which could affect the ratio of information from high and low credibility sources being shared on Twitter.

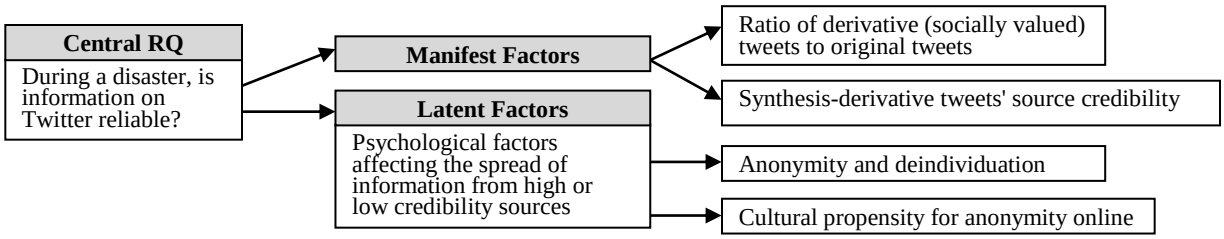


Figure 2. Investigative Framework

Considering the nature of Twitter profiles (limited space for biographical information, high propensity for handle-name usage among Japanese users), and for the sake of coding simplicity, anonymity in the present study was measured simply by whether or not a user specifies a location on their profile.

Following this framework, and considering previous findings in related works outlined above, the following hypotheses were advanced (all in relation to the Fukushima nuclear accident):

H1 Synthesis-derivative tweets make up the majority of tweets sent;

H2 The majority of information sources on which synthesis-derivative tweets are based comprise of highly credible sources;

H3 Individual Twitter users who do not show their location on their profile are more likely to share information from low credibility sources than those who do;

H4 Japanese language synthesis-derivative tweets are more likely posted by individual users who do not specify a location, and therefore are more likely to contain information from low-credibility sources.

METHOD

Data collection

The core data for the present study's analysis was collected from the Twitter public timeline between the 12th of March and the 13th of April 2011, via a keyword-based data collection system managed by HEROIC - a collaborative investigation of crisis communication by researchers the University of Colorado Colorado Springs and University of California-Irvine. To mitigate issues of high noise (unrelated tweets), analysis was limited to tweets which included the hashtag *#fukushima* (the hashtag is a function within Twitter whereby users are able to socially categorize tweets into relevant topics). Previous to the earthquake and subsequent nuclear accident, this hashtag was

used less than a few times a day (comprising local information relevant to the place name Fukushima), increasing at such a massive rate after the disaster so as to comprise a total of approximately 236,000 tweets during the month following the disaster (Butts et al., 2011). This indicates an extremely high level of relevancy to the accident. *#fukushima* also allowed a cross-language comparison of the data; Japanese terms such as *#genpatsu* (nuclear power) were not likely to be used as hashtags by non-Japanese language users. From this pool of *#fukushima* tweets, tweets from active users (those users who tweeted 5 or more times using *#fukushima* during the study period) totaling 88,283 tweets were extracted. The decision to limit the data set to high volume users was made with coding burdens in mind; in the end, a 5.6% random sample of those high volume tweeters equated to 4,950 tweets to code. This percentage was a result of first determining how many tweets each member of the research team could code within the time frame allotted (7 coders with 700 tweets per coder, plus 50 tweets used in the inter-coder reliability testing). This does, however, introduce the possibility that results may have ended up skewed towards higher or lower overall instances of high or low credibility information shared. That is, one could suggest that high-volume users might be more savvy in their judgment of sources, taking care not to share information from low-credibility sources. Equally conceivable, however, is that high-volume Twitterers find information on Twitter more credible in general, compared with less experienced users (Flanagin and Metzger, 2000), meaning that they may have more propensity to share information from sources which, in the present study's concept of credibility (initial perceived credibility), are deemed low-credibility.

Coding

Two coding passes were performed on the data: the first pass involved the entire sample, and separated users and tweets into general categories; the second pass was limited to categorizing synthesis-derivative tweets' information sources identified as *individuals* into credibility categories of *high*, *low*, and *neither of the above*, as per Table 1. For the first coding pass on the data, we followed closely Starbird et. al.'s (2010) categorization scheme, where the content of the tweets, user characteristics, and information sources were categorized into the following categories.

User types: Individuals, traditional media, alternative media, public institutions, enterprises, non-profit organizations, bots, unknown.

Location of user: Fukushima prefecture, prefectures adjoining Fukushima prefecture, within Japan (from locations separate from Fukushima and adjoining prefectures), overseas, within Japan (no further specific location specified), unknown.

Tweet type: Original, synthetic, and derivative.

Information source: Individual, traditional media, alternative media, public institutions, enterprises, non-profit organizations, bots, unknown.

The main method of determining a user's location was via the Twitter user profile "location" field. Noteworthy here is Hecht et. al.'s (2011) findings that on Twitter, 34% of users in the US enter either non-geographic information or no location information at all into their profile's location field. In the present study, non-geographic information, as defined in Hecht et. al. (2011) - e.g., *outer space*, *inside my head*, *earth*, all examples from the current study - was coded as unknown. Coders were, however, instructed to use other periphery clues to make inferences as to a user's location if either non-geographic information or no location information was provided in the Twitter profile location field. This included a cursory examination of a page of a user's tweet content (e.g., references to one's neighborhood) and location information found on a user's homepage or blog, if the URL was shared on the user's Twitter profile. In regards to synthesis-derivative tweets' source credibility evaluation, coders utilized a simple three point scale where 1=*low credibility*, 2=*nether high nor low*, 3=*high credibility*.

Coding was performed by seven researchers (700 individual tweets per coder), with inter-coder reliability tested over two separate test-runs of 20 and 30 tweets respectively. Overall intra-class correlation scores were acceptable (α= 0.98 on both test-runs), however issues were identified in the classification of *alternative media*, with reliability scores as low as α= 0.45. These issues were addressed in research meetings, and a coding manual was created to guide the coding process.

	N	%
Synthesis-derivative	3381	68%
Original	1150	23%
Unknown	419	9%
Total	**4950**	**100%**

Table 2. Tweets by Type

RESULTS

Synthesis-derivative tweets (N=3383) comprised almost 70% of tweets shared over the study period (Table 2). This confirms **H1** where it was

Source		N	%
Individuals		1189	34.8%
Official information	Traditional Media	965	28.2%
	Public Institutions	323	9.5%
	Enterprise	127	3.7%
	NPOs	128	3.7%
	Subtotal Organization	*1543*	*45.1%*
Other	Alternative Media	443	13.0%
	Bots	10	0.3%
	Unknown	234	6.8%
	Subtotal Other	*687*	*20.1%*
Total		**3419**	**100.0%**

Table 3. Synthesis-derivative Tweet Information Source

Credibility	Category	*N*	*%*
High	Freelance Journalists	450	37.8%
	Academics/Professionals	100	8.4%
	Fukushima Locals	88	7.4%
	Other high-trust	128	10.8%
	Subtotal - High	*766*	*64.4%*
Low	Un-identifiable sources, non-locals	193	16.2%
	Conspiracy theorists	72	6.1%
	Subtotal - Low	*265*	*22.3%*
Neither high nor low		158	13.3%
Total		1189	100%

Table 4. Individual Source Credibility

expected that there would be more sharing in the form of third-party information passed on, than purely original content.

H2 concerned the information sources of synthesis-derivative tweets, where it was expected that the majority of information sources on which synthesis-derivative tweets are based would comprise of highly credible sources. As shown in Table 3, individuals made up 34.8% (N=1189) of the information sources, with a remaining 45.1% (N=1543) being from some form of official source. A second coding pass on the data, however, found that 64.4% (N=766) of all information sourced from individuals was from highly credible individuals (Table 4). Therefore, when combined with other high credibility categories, a full 67.5% (N=2,309) of information sources making up synthesis-derivative tweets proved to be highly credible (Table 5), supporting the second hypothesis.

Hypotheses three and four concerned anonymity and between-languages comparisons in regards to the sharing of information from low credibility sources. The full sample (N=4950) consisted mainly of tweets in Japanese (60.4%), with English (29.9%) and other languages (9.7%) making up the remainder (Table 6). When looking at average credibility rating scores (1 to 3), in general, it was observed that indeed, supporting **H3**, individual Twitter

Credibility	Source	Category	*N*	Source *%*	Credibility *%*
High	Official sources	Traditional Media	965	28.2%	41.8%
		Public Institutions	323	9.4%	14.0%
		Enterprise	127	3.7%	5.5%
		NPOs	128	3.7%	5.5%
	Subtotal official info.		***1543***	***45.1%***	***66.8%***
	Individuals	Freelance Journalists	450	13.2%	19.5%
		Other high-trust	128	3.7%	5.5%
		Academics/Professionals	100	2.9%	4.3%
		Fukushima Locals	88	2.6%	3.8%
	Subtotal individuals		***766***	***22.4%***	***33.2%***
Subtotal High Credibility			***2309***	***67.5%***	***100%***
Low	Individuals	Un-identifiable sources, non-locals	193	5.6%	72.8%
		Conspiracy theorists	72	*2.1%*	27.2%
Subtotal Low Credibility			***265***	***7.8%***	***100%***
Neither high nor low	Other	Individuals	158	4.6%	18.7%
		Alternative Media	443	13.0%	52.4%
		Bots	10	0.3%	1.2%
		Unknown	234	*6.8%*	*27.7%*
Subtotal Other Credibility			***845***	***24.7%***	***100%***
Source Total			**3419**	***100%***	

Table 5. Synthesis-derivative Tweet Source and Credibility

users who did not show their location on their profile (M=2.60, SD=.677) posted synthesis-derivative tweets whose sources were of low credibility more often than those who did show their location on their profile (M=2.67, SD=.592), t(3151) = 3.06, p< .01, *d*=.110.

	N	%
Japanese	2990	60.4%
English	1492	29.9%
Other	478	9.7%
Total	4950	100%

Table 6. Tweets by Language

H4 had two elements: 1) Japanese language synthesis-derivative tweets are more likely posted by individual users who do not specify a location, and therefore 2) are more likely to contain information from low-credibility sources. In regards to the first element, tweet language did have a small effect on location disclosure (V=.145); a significantly higher percentage of Japanese language tweets were posted by users who did not specify their location (39.7%) than users who posted in other languages (24.0%), χ^2 (2, N = 3,381) = 71.012, p <.001. Despite this, however, regarding the second element of the hypothesis, average credibility ratings of the sources of synthesis-derivative tweets posted in Japanese (M=2.66, SD=.636) were no lower than synthesis-derivative tweet sources quoted by non-Japanese posters (M=2.62, SD=.595), t(2391) = 1.17, p= .241.

Interestingly, when the data was limited specifically to synthesis-derivative tweets posted by individual users whose sources were other individuals, *non-Japanese* language tweets' sources (M=2.18, SD=.892) were significantly less credible than Japanese language tweets' sources (M=2.50, SD=.789), t(981) = 3.94, p< .001, *d*=.380. In this way, hypothesis four was partly supported, although it is clear that the Japanese propensity towards anonymity online did not cause an increase in sharing of low credibility information.

DISCUSSION

While close to 70% of synthesis-derivative tweets were found to cite highly credible sources, it requires somewhat of a precarious value judgment as to whether this indicates positive or negative value for Twitter as a whole during disasters in general, and the Fukushima nuclear disaster in particular. One could equally point out, with valid concern, that a full one third of tweets based on third-party information utilized information from low credibility sources. This is further confounded by the possibility that many users who did cite low-credibility information, as defined in our model, were, as far as their experience with their source is concerned, in fact citing sources which to them may have been highly credible. Follow-up qualitative study is required here to tap into greater detail users' behavior and perception regarding credibility of information sources, which would help provide a clearer picture.

As **H4** states, there was an initial expectation to see a higher propensity for the sharing of information from low-credibility sources by Japanese tweeters (due to a higher level of profile anonymity). A probable explanation to this expectation not being met could simply be that individual Japanese users would, understandably, feel a greater responsibility towards the Twitter information environment in times of national disaster, more so than overseas users for whom the relevance of and responsibility towards the disaster is diluted due to distance. Indeed, for individuals who posted individual-sourced synthesis-derivative tweets, those who listed their location as somewhere in Japan averaged a higher source credibility (M=2.49, SD=.796) than those who indicated their location as outside of Japan (M=2.25, SD=.886), t(806) = 2.72, p< .01, *d*=.285. This adds another dimension to Starbird & Palen's (2010) study where proximity to the crisis was found to bear differences in information shared on Twitter during a crisis.

Worth noting is the absence of false rumors in tweets shared with the *#fukushima* hashtag. As noted previously, Acar and Muraki (2011) report that Twitter users in disaster-hit Miyagi Prefecture found false rumors to be an issue when using Twitter as an information source. Locals suggested instigating an official hashtag system, presumably so that tweets could be screened for false rumors. One might suggest, however, that such a system of official intervention into an informal information ecosystem which operates on the concept of informal, social, user-generated media, is unlikely to garner the support required for its widespread adoption. Rather, a system which makes the most of the organic, collaborative filtering nature of user-based participation on Twitter would be of more value. Such a system which utilizes the fact that false rumors are more widely questioned than truths, for example, has been suggested previously by researchers in the aftermath of the Chile earthquake of 2010 (Mendoza, Poblete, & Castillo, 2010). One might suppose that the absence of false rumors associated with the *#fukushima* hashtag could be due to this organic collaborative filter in action. That is, non-hashtagged tweets on the whole only reach nodes within a sender's network and any subsequent networks they are forwarded on to. Therefore, they can only be questioned by a limited cluster(s) of individuals, to whom the sender is presumably connected to in some way (and therefore is presumably trusted enough by others in their networks so as to encourage further sharing). Hashtagged tweets, however, are immediately exposed to a much larger audience (who are not necessarily connected to the user

in any meaningful trust-relationship way), inviting a much larger probability of being questioned and dismissed by a large cross-section of users from a diverse range of clusters. Further research is required here.

CONCLUSION

In this paper, a common perception that information shared on Twitter is unreliable was addressed. Analysis was focused on information shared and categorized by the hashtag *#fukushima*, and for this particular hashtag, it was found that close to 70% of synthesis-derivative tweets (tweets containing some form of third-party information; the a majority of tweets being of this type) were based on highly credible sources. Furthermore, anonymity, an oft-cited factor in the degradation of information online, did not appear to have a detrimental effect on Japanese tweeters (who showed a relatively high propensity for anonymity). Rather, proximity to the crisis seemed to moderate an increased tendency to share information from highly credible sources. Further work is required in order to capture a snapshot of source credibility which allows for the consideration of the dynamism of credibility perceptions over time.

LIMITATIONS

For this paper, analysis was limited to synthesis-derivative tweets only. The reasoning behind this is that it was considered that information passed on by users - i.e., socially filtered information - constitutes more value to the information ecosystem as a whole when compared with original tweets which were not passed on. This does not, however, mean to suggest that original tweets, which are presumably read at least by some fellow users, do not hold any power to influence and/or cause difficulties for locals and other users during a disaster situation.

Furthermore, as briefly mentioned earlier in the paper, due to the breadth and depth of topics shared in the tweet sample, it was not possible for coders to achieve a central route to persuasion (i.e., assessing the truthfulness of each and every claim of each individual tweet). As such, we looked at simply initial perceived credibility, based on the ideological place in society of those high-credibility sources as defined in this study.

Finally, it needs to be acknowledged that characteristics particular to the hashtag *#fukushima* may limit the generalizability of these observations. In particular, more universal hashtags such as *#earthquake*, *#tsunami*, etc., were conceivably used by a considerably larger subset of users, thus may be prone to more a greater concentration of synthesis-derivative tweets based on low-credibility sources or vice-versa (and also spam, which was not found in this study's sample of tweets).

ACKNOWLEDGMENTS

Many thanks for the kind cooperation of Leysia Palen and Kate Starbird of Project EPIC at the University of Colorado Boulder. Thanks also go to the HEROIC research effort at the University of California Irvine and the University of Colorado Colorado Springs (PIs Carter T. Butts and Jeannette Sutton) for supplying the data sample, the collection of which was supported under NSF awards CMMI-1031853 and CMMI-1031779. Thanks also to our reviewers for their insightful comments.

REFERENCES

Acar, A., & Muraki, Y. (2011). Twitter for crisis communication: lessons learned from Japan's tsunami disaster. *International Journal of Web Based Communities*, *7*, 392. doi:10.1504/IJWBC.2011.041206

Barker, V., & Ota, H. (2011). Mixi Diary versus Facebook Photos: Social Networking Site use among Japanese and Caucasian American Females. *Journal of Intercultural Communication Research*, *40*(1), 39.

Butts, C. T., Sutton, J., Spiro, E. S., Greczek, M., Fitzhugh, S., & Pierski, N. (2011). Hazards, Emergency Response, and Online Informal Communication Project Data. University of California, Irvine and University of Colorado, Colorado Springs.

Chester, A., & Gwynne, G. (1998). Online Teaching: Encouraging Collaboration through Anonymity. *Journal of Computer-Mediated Communication*, *4*(2). Retrieved from http://jcmc.indiana.edu/vol4/issue2/chester.html

credibility. (n.d.). Dictionary.com Unabridged. Random House, Inc. Retrieved January 06, 2012, from Dictionary.com website: http://dictionary.reference.com/browse/credibility

Davis, J. P. (2002). The Experience of "Bad" Behavior in Online Social Spaces:A Survey of Online Users. *Microsoft Research*.

Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of Internet information credibility. Journalism and Mass Communication Quarterly, 77, 515-540.

Hecht, B., Hong, L., Suh, B., & Chi, E. H. (2011). Tweets from Justin Bieber's heart: the dynamics of the location field in user

profiles. *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11 (pp. 237–246). New York, NY, USA: ACM. doi:10.1145/1978942.1978976

Heverin, T., & Zach, L. (2010). Microblogging for crisis communication: an examination of Tiwtter use in response to a 2009 violent crisis in Seattle-Tacoma. *Paper presented at the Seventh International ISCRAM Conference, Seattle, Washington*. Retrieved from http://www.thomasheverin.com/uploads/4/6/5/8/4658640/heverin_iscram_2010.pdf

Hokudai Earthquake Project. (2011). *2011 General Consumer Survey*.

Hovland, C. (1953). *Communication and persuasion: Psychological studies of opinion change,*. New Haven: Yale University Press.

Katayama, L. (2007, April 19). 2-Channel Gives Japan's Famously Quiet People a Mighty Voice. *Wired.com*. Retrieved October 23, 2011, from http://www.wired.com/culture/lifestyle/news/2007/04/2channel

Keeley, B. L. (1999). Of Conspiracy Theories. *The Journal of Philosophy*, *96*(3), 109-126.

Mendoza, M., Poblete, B., & Castillo, C. (2010). Twitter Under Crisis: Can we trust what we RT? Retrieved October 23, 2011, from http://research.yahoo.com/node/3255

Merton, R. K. (1968). *Social theory and social structure*. Free Press.

MEXT. (2011, June 1). Reading of environmental radioactivity level by prefecture. Retrieved from http://www.mext.go.jp/component/english/__icsFiles/afieldfile/2011/06/01/1306685_060119.pdf

Morris, M. R., Counts, S., Hoff, A., Roseway, A., & Schwarz, J. (2012). Tweeting is Believing? Understanding Microblog Credibility Perceptions. *Proceedings of CSCW 2012*. Presented at the CSCW 2012, Seattle, WA, USA: ACM. Retrieved from http://research.microsoft.com/apps/pubs/default.aspx?id=155374

Nishida, M. (2011, May 11). The internet increases disaster response capabilities (in Japanese). Retrieved October 18, 2011, from http://astand.asahi.com/magazine/wrnational/2011042800007.html

NYTimes.com. (2011a, September 24). Japanese Rice's Radiation Levels Prompt More Tests. Retrieved October 17, 2011, a from http://www.nytimes.com/2011/09/25/world/asia/japan-testing-rice-for-radiation.html?ref=foodsafety

NYTimes.com. (2011b, July 18). Radiation-Tainted Beef Spreads Through Japan's Markets. Retrieved October 17, 2011, b from http://www.nytimes.com/2011/07/19/world/asia/19beef.html?_r=3

Ohshima, H., Yamamoto, Y., Yanbe, Y., Takahashi, R., Jatowt, A., Nakamura, S., & Tanaka, K. (2011). Web information credibility (in Japanese). *Information Science and Technology*, *61*(1), 2-7.

Ozeki, Y. (2011, May 11). The risks of web media, as shown by disaster (in Japanese). Retrieved October 18, 2011, from http://astand.asahi.com/magazine/wrnational/2011042000017.html

Petty, R., & Cacioppo, J. (1986). *Communication and persuasion : central and peripheral routes to attitude change*. New York: Springer-Verlag.

Postmes, T., Spears, R., & Lea, M. (1999). Social identity, normative content, and "deindividuation" in computer-mediated groups. In N. Ellemers & R. Spears (Eds.), *Social identity: Context commitment content* (pp. 164-183). Oxford, UK: Blackwell Science Ltd.

Reuters. (2011, April 7). Some South Korean schools close over radioactive rain concerns. Retrieved October 17, 2011, from http://www.reuters.com/article/2011/04/07/us-japan-quake-korea-idUSTRE7360KL20110407

Saito, T. (2011, April 18). A review of Nielson's March 2011 mixi, Facebook and Twitter survey: Earthquake causes spike in social media users (in Japanese). *Alternative Blog*. Retrieved October 27, 2011, from http://blogs.itmedia.co.jp/saito/2011/04/mixi-twitter-fa-0f8b.html

Schmierbach, M., & Oeldorf-Hirsch, A. (2010). A little bird told me, so I didn't believe it: Twitter, credibility, and issue perceptions. *Paper presented at the annual meeting of the Association for Education in Journalism and Mass Communication*. The Denver Sheraton, Denver, CO.

Starbird, K, & Palen, L. (2010). Pass It On?: Retweeting in Mass Emergencies. *7th International Information Systems for Crisis Response and Management Conference*. Presented at the ISCRAM 2010, Seattle, WA, USA. Retrieved from http://www.cs.colorado.edu/~palen/starbirdpaleniscramretweet.pdf

Starbird, Kate, Palen, L., Hughes, A. L., & Vieweg, S. (2010). Chatter on The Red: What hazards threat reveals about the social life of microblogged information. *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, 241-250.

Takahashi, T. (2010). MySpace or Mixi? Japanese engagement with SNS (social networking sites) in the global age. *New Media & Society*, *12*(3), 453-475.

TEPCO. (2011, April 9). TEPCO : Press Release | Results of the investigation regarding tsunami arrived in Fukushima Daiichi Nuclear Power Station and Fukushima Daini Nuclear Power Station. Retrieved October 17, 2011, from http://www.tepco.co.jp/en/press/corp-com/release/11040910-e.html

The Japan Times Online. (2011a, April 22). No-go zone trespassers face fines, arrest. Retrieved October 17, 2011, a from http://search.japantimes.co.jp/cgi-bin/nn20110422a1.html

The Japan Times Online. (2011b, May 28). G-8 differ in reactions to Fukushima. Retrieved October 17, 2011, b from http://search.japantimes.co.jp/cgi-bin/nb20110528x1.html

Thomson, R., & Ito, N. (forthcoming). The effect of relational mobility on SNSuser behavior: A study of Japanese dual-users of Mixi and Facebook. *The Journal of International Media, Communication and Tourism Studies*, *14*.

Toriumi, F., Shinoda, K., Kurihara, S., Sakai, T., Kazama, K., & Noda, I. (2011). *How Disaster Changes Social Media (in Japanese)*. Paper presented at the JWEIN '11 Conference.

Yamagishi, T. (1998). *The Structure of Trust: Evolutionary games of mind and society (in Japanese)*. University of Tokyo Press.