Toward a Social-Technological System that Inactivates False Rumors through the Critical Thinking of Crowds

Yuko Tanaka
Stevens Institute of Technology
yuko.tanaka@stevens.edu

Yasuaki Sakamoto
Stevens Institute of Technology
ysakamot@stevens.edu

Toshihiko Matsuka
Chiba University
matsukat@muscat.l.chiba-u.ac.jp

Abstract

Critical thinking is an important part of media literacy. It allows people to find facts among rumors and to inactivate false information. Such abilities are essential when social media is flooded with rumors during disaster response. We envision a social-technological system in which critical thinking is crowd-sourced: Individuals benefit from others' criticisms of false information, and the system inactivates the spread of false information. To test the plausibility of this system, we examined the effect of exposure to criticisms on people's decision to spread rumors in social media. When people were exposed to criticisms before rumors, the proportion of responses aimed at stopping the spread of rumors was significantly larger than when people were exposed to rumors before criticisms. We identified some psychological factors that could explain this effect. Based on our results, we discuss practical implications for developing a social-technological system that harnesses the critical thinking of crowds.

1. Introduction

Recent advances in social media technologies, such as Twitter and Facebook, have dramatically changed the environment in which we obtain and exchange information. In social media, users indirectly collaborate by observing and learning from each other. These technologies, which have become integral to our everyday lives, allow us to generate contents and influence trends. Given the growing use and participatory nature of social media, critical thinking is considered an important element of media literacy that individuals in a society should possess [10].

Critical thinking is defined "reasonable reflective thinking focused on deciding what to believe or do" [9]. Aiming at promoting critical thinking as an element of media literacy, a great deal of research has focused on the role of teaching critical thinking to students in educational environments [6, 10, 11, 12, 19, 32]. It is no doubt important to teach and enhance critical thinking skills of each student.

Nonetheless, the approach to promoting critical thinking in our society as a whole may not work because not everyone has access to critical thinking education and critical thinking is domain- and subject-specific, as much research has pointed out [14, 25]. Thus, even if an individual is lucky enough to acquire general critical thinking skills, she may not have the necessary knowledge to think critically about a specific topic. Moreover, since cognitive processes involved in critical thinking are effortful [13, p. 20], they can be inhibited under mental pressure, such as during responses to disasters. For these reasons, we need an alternative approach that overcomes these limitations.

We envision a social-technological system in which critical thinking is crowd-sourced. This system allows individuals to make good use of the critical thinking of others. Not everyone needs to be a critical thinker all the time. For example, people who are poor critical thinkers under mental pressure can observe and rely on criticisms of others who are good at thinking critically under pressure. In this way, the system can disregard questionable information as long as some individuals are good critical thinkers in a specific domain. By sharing at a particular time criticisms of critical thinkers to those who are less critical at that moment, the system as a whole can behave like a critical thinker and mitigate the spread of false information.

The purpose of the current work is to test the plausibility of this system by examining the effect of exposure to criticisms on people's decision to spread rumors that are associated with the criticisms. Based on our results, we discuss practical implications for building a social-technological system that enables users to benefit from the critical thinking of others. In addition, we examine psychological factors associated with exposure to criticism in order to understand why exposure to criticisms changes the rumor-spreading behavior.

In particular, we investigate the following research questions:

RQ1: Does exposure to criticisms change rumor-spreading behavior?

RQ2: What psychological factors relate to the effect of exposure to criticisms on rumor-spreading behavior, if there is a significant effect?

Before presenting our experiment, the results, and our practical implications, we review relevant past work and formulate our hypotheses.

2. Background

2.1. Spread of Rumors in Social Media

On March 11, 2011, a 9.0 magnitude earthquake hit northeastern Japan. The Great East Japan Earthquake triggered powerful tsunamis and a series of aftershocks, causing devastating damages including radiation leaks. Social media played a major role in obtaining and transmitting information to understand the situation during the disasters. An example is Twitter, which allows its users to send and read text messages, known as "tweets", of up to 140 characters, and to forward a tweet to followers through a single click, known as "re-tweeting."

As social media allows immediate information sharing, communications during disasters increasingly relies on social media. For example, many people used Twitter to post, share, and obtain information during responses to the disasters caused by the Great East Japan Earthquake. Information on Twitter helped the discovery and rescue of some victims. Social media can no doubt play an important role in sharing information and coordinating disaster response. At the same time, social media can facilitate the dissemination of inaccurate information, which can cause widespread panic. In fact, after the Great East Japan Earthquake, Twitter immediately was flooded with false information [17, 30]. The Japanese government called attention to false rumors in social media, but the problem persisted.

There is a need to better prepare citizens and officials for disaster response using social media technologies. We contribute to this need through the current work and through our project, in which we develop a social-technological system that facilitates the critical thinking of crowds.

Twitter, which has been around since 2006, is relatively new, but the spread of false rumors during disasters is not new [20, 21, 28]. Prasad (1950) found that in the past 1,000 years the same types of rumors related to earthquakes appear again and again in different locations [21]. Since then, social psychologists have made progress in identifying the psychological factors associated with the rumor-spreading behavior [7]. Nevertheless, false rumors still circulate. The spread of false rumors seems to be becoming more severe as technologies advance. For instance, false rumors spread through the Internet during response to the Sichuan earthquake in 2008 [29], and through social media during response to the 2010 earthquakes in Haiti and Chile earthquake [16, 18]. Although social media can help people make sense of their situation during a disaster, social media can also become a rumor mill and create social problems.

2.2. Critical Thinking to Combat Rumors

A widely used definition of rumor is "*unverified and instrumentally relevant information statements in circulation that arise in contexts of ambiguity, danger, or potential threat and that function to help people make sense and manage risk*" [7, p.13]. Rumor is different from gossip, which is defined as an evaluative statement about someone's private lives.

Past rumor studies have revealed that psychological factors, such as accuracy, anxiety, and importance of rumors, affect rumor transmission [1, 2, 8, 22, 24, 31]. Analyzing the spread of rumors after WWII, Allport & Postman [1] proposed that the spread of rumor "will vary with the importance of the subject to the individuals concerned times the ambiguity of the evidence pertaining to the topic at issue." Later, Anthony [2] introduced anxiety as another key element in rumormongering. For example, anxious students were more likely to report that they heard a rumor [2]. The likelihood of sharing a rumor was related to how anxious the rumor made people feel [22]. Furthermore, Chorus [5] extended the rumormongering model of Allport and Postman [1] by adding critical thinking ability; as the critical thinking ability increases, the spread of rumors decreases.

Related to critical thinking, a number of studies have paid attention to the role of denial or rebuttal messages in impeding the transmission of rumor [3, 4, 8, 15]. For example, Shibutani [27] reviewed work that examined how individual differences, including differences in people's critical thinking ability, contributed to the rumor behavior. Consistent with Shibutani's review described above, we noticed that, while many people spread rumors, others tried to stop the spread of false rumors by posting tweets criticizing the rumor-tweets in the case of the Great East Japan Earthquake. The presence of false rumors and criticisms that counter the false rumors on Twitter during disasters suggests that there are three kinds of

people in such a situation: People who originate false information, people who detect and criticize questionable information, and people who diffuse rumors, criticisms, or both. Twitter is a social-technological environment, in which these three kinds of people generate and diffuse information collectively.

Are criticisms on Twitter effective in mitigating the spread of false rumors? Can exposure to criticisms minimize the spread of rumors? We hypothesize that:

H1: Exposure to criticisms reduces people's intent to spread rumors.

If *H1* is supported, it lends support to our idea of a social-technological system that minimizes the spread of false rumors by making good use of the critical thinking of crowds.

Assuming that *H1* is supported, what psychological factors are related to this effect?

H2: Exposure to criticisms lowers perceived accuracy, anxiety, and importance of rumors.

By examining *H2*, we can help fine-tune the design of the social-technological system, which we envision.

In the current work, we test these hypotheses in an experiment. There are two important differences between the present study and the traditional social psychological research of rumor: First, we use several false rumors, as stimuli, that were spread on Twitter, a social media environment; second, we examine the rumor-spreading behavior in social media, not in face-to-face environments.

3. Experiment

3.1. Participants

In total, 87 undergraduate and graduate students (40 male, mean age 20.2 years) from Chiba University in Japan participated individually. They received course credit and a gift card in the amount of 500 Japanese yen (about $6.3). The experiment was conducted between October 19, 2011 and November 1, 2011. Data collection was completed within eight months after the Great East Japan Earthquake. Chiba University was located in one of the areas affected by the disasters.

3.2. Stimuli

First, we collected 10 *rumor-tweets* related to the disasters following the Great East Japan Earthquake including tweets about the nuclear accident, electric power issues, and supplying disaster areas. Rumor was defined as a tweet that had been criticized by one or more other tweets. An experimenter checked that each rumor-tweet included wrong, inaccurate, or suspicious information. No rumor-tweet offered any evidence to support its information. We also collected the 10 *criticism-tweets* that criticized the corresponding rumor-tweets (see Appendix for complete list of stimuli). Criticism was defined operationally as a tweet that denied, refuted, or doubted another tweet by citing the tweet. Each tweet was posted in Japanese on Twitter between March 11, 2011 and September 7, 2011.

Next, we converted each of these 20 tweets to a 700×162 pixels image in the PNG format (see Figure 1). With the purpose to examine the difference between rumor-tweets and criticism-tweets in terms of the text information, we controlled other aspects such as user name and user image as follows: The user name associated with each tweet was generated by randomly combining alphabet and number. Every stimulus had the same user image consisting an egg-shape on a red square background. Each image also contained the actual date when the original tweet was posted.

We created each criticism-tweet by adding the word "RT" (an abbreviation for Re-Tweet), the user name of the corresponding rumor-tweet, and part of the rumor-tweet to the criticism (see Figure 1, bottom). The maximum number of characters in each tweet image was 140 in Japanese.

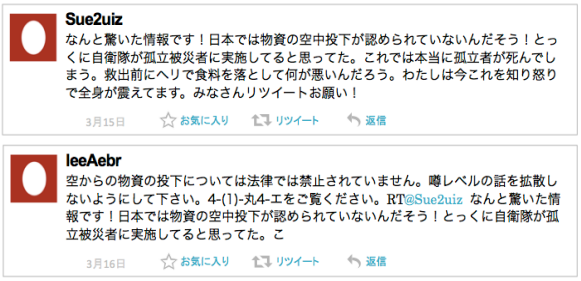

Figure 1. Example of stimuli

Top) A rumor-tweet: "*Air drop of supplies is not allowed in Japan! I though it has already been done by the Self-Defense Forces. Without it, the isolated people will die! I'm trembling with anger. Please retweet!"* Bottom) A corresponding criticism-tweet: *"Air drop of supplies is not prohibited by the law. Please don't spread rumor. Please see 4-(1)-丸 4-エ."*.

3.3. Experimental Design

The experiment involved a 2 × 2 factorial design in which there are two groups. Each group had two types

of stimuli (Table 1). The two groups were the *rumor-first group* and the *criticism-first group*. Of the 87 participants, 40 participants were allotted to the rumor-first group in which the 10 rumor-tweets were given followed by the 10 criticism-tweets. The other 47 participants were allotted to the criticism-first group in which the criticism-tweets were given followed by the rumor-tweets. The order of rumor-tweets was randomized within 10 rumor-tweets. The order of criticism-tweets was randomized in the same way.

Table1. Experimental design

Participants were assigned to the rumor-first group or criticism-first group.

Phase	Rumor-first group (n = 40)	Criticism-first group (n = 47)
1	Rumor × 10	Criticism × 10
2	Criticism × 10	Rumor × 10
3	Demographic information	
4	Debriefing	

3.4. Procedure

Participants accessed the experiment through the Internet using a computer. They were instructed to answer all questions within 50 minutes, and to answer each question in order. The experiment consisted of four phases in the following order:

1. Rumor-tweets: Participants answered the following eight questions about each of the 10 rumor-tweets:

(1) Familiarity – Have you heard this information? (Yes, No)
(2) Anxiety – How anxious did you feel when you heard this information? (1 Not at all, 7 Highly anxious)
(3) Importance – How important do you think this information is? (1 Not at all, 7 Highly important)
(4) Intended receiver – Who should know this information? (Family, Friend, Victims, Many Japanese, Many people abroad, Anyone, Other)
(5) Intent to spread – How many people do you to send this information to? (Open-ended)
(6) Self-accuracy – How accurate do you think this information is? (1 Not at all, 7 Highly accurate)
(7) Estimated transmission – How many people do you think have already known this information at present? (Open-ended)
(8) Others-accuracy – How accurate would others think this information is? (1 Not at all, 7 Highly accurate).

2. Criticism-tweets: Participants answered the same eight questions used in the rumor-tweets phase about each of the 10 criticism-tweets. The order of the rumor-tweets phase and the criticism-tweets phase was reversed in the criticism-first condition.

3. Demographic information: Participants answered demographic questions, questions about how severely they felt affected by the Great East Japan Earthquake, and questions about their familiarity with Twitter and mass media.

4. Debriefing: We explained the purpose of the experiment to each participant. It was emphasized that the tweets in the experiment might be false, and that the spread of false rumor was becoming a social problem following the disaster. For further information, we recommended useful books and websites that examined the false rumor-tweets related to the disaster.

4. Results

Seventy-nine participants (90%) lived in Chiba prefecture and eight participants lived in neighboring prefectures of Chiba. 72 participants (83%) were affected by the disasters to some degree. Forty-eight of 87 participants (55%) had Twitter account. To answer the two research questions mentioned above, we will focus on the results of the intent to spread tweets (measured by Q5) and psychological factors (measured by Q2, Q3, Q6, and Q7).

4.1. Proportion of Rumor-Stopper

H1: Exposure to criticisms reduces people's intent to spread rumors.

As to the responses about 10 rumor-tweets and 10 criticism-tweets by the 87 participants, the range of the intent to spread in the original data set was from 0 to 100,000 billion. However, since the number of users who created their profiles is 383 million as of January 1, 2012 [26], it is practically impossible to spread tweet information to 100,000 billion. Thus, the responses over 383 million were defined as outliers and removed from the data. The proportion of outliers to all responses of the intent to spread was 1.7%. Table 2

shows the number of outliers and non-outliers in each condition.

According to the intent to spread, responses in the rumor condition were categorized into three groups: rumor-stopper, weak rumor-spreader, and strong rumor-spreader. Rumor-stopper is a response aimed at stopping the spread of rumor-tweets. Weak rumor-spreader is a response to spread a rumor to a relatively small number of people, from one to no more than 10,000. Strong rumor-spreader is a response to spread a rumor to a relatively large number of people more than 10,000. Table 2 shows the numbers of responses, the averages, and the medians, of the intent to spread rumor-tweets and criticism-tweets in these three categories.

We performed Chi-square analyses in order to examine the effect of criticism exposure on the spread of rumor-tweets. Figure 2 shows the proportions of these three groups in the rumor-first group and the criticism-first group. In terms of rumor-stopper, the result of chi-square test showed that the proportion of rumor-stopper in the criticism-first group (49.3%) was significantly larger than rumor-first group (32.1%) (x^2 = 26.11, df = 1, p < 0.001). In terms of strong rumor-spreader, the chi-square test of the proportion of strong rumor-spreader in the criticism-first group (25.0%) slightly decreased than the rumor-first group (30.3%) (x^2 = 2.97, df = 1, p = 0.85).

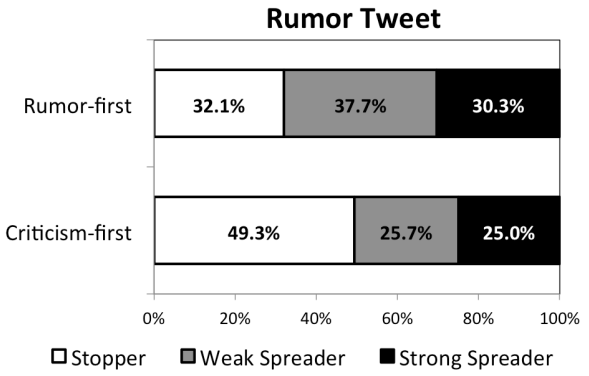

Figure 2. The proportion of stopper, weak spreader, and strong spreader in rumor-tweets

Table 2. Intent to spread in the rumor-first group and criticism-first group

Stopper represents a response that intended to stop a tweet. Weak-Spreader represents a response that intended to spread a tweet to a relatively small number of people in the range of 1 to 10,000. Strong Spreader represents ta responses that intended to spread a tweet to a relatively large number of people more than 10,000. Outliers were defined as the responses that intended to spread a tweet to more than 383 million people that was an impracticable number on Twitter at the point when the experiment was conducted.

	Rumor-first group (n = 40)		Criticism-first group (n = 47)	
	Rumor-tweets	**Criticism-tweets**	**Rumor-tweets**	**Criticism-tweets**
Number				
Stopper	126	115	227	192
Weak Spreader	148	120	118	153
Strong Spreader	119	158	115	120
Total responses	393	393	460	465
Outliers	7	7	10	5
Average				
Weak Spreader	2,257.3	3,663.7	3,642.6	3,335.9
Strong Spreader	30,072,689.1	27,896,993.7	21,651,078.3	25,546,083.3
Median				
Weak Spreader	100	1,000	1,000	1,000
Strong Spreader	9,000,000	1,000,000	1,000,000	1,000,000

4.2. Psychological Differences between Rumor-First Group and Criticism-First Group

H2: Exposure to criticisms lowers perceived accuracy, anxiety, and importance of rumors.

In order to examine the differences in the proportion of rumor-stopper between the rumor-first group and criticism-first group, we analyzed the four psychological factors: self-accuracy, others-accuracy, anxiety, and importance.

A two-way analysis of variance (ANOVA) of self-accuracy, with order of tweet (rumor-first vs. criticism-first) as a between-subject factor by tweet type (rumor vs. criticism) as a within-subject factor, revealed a significant main effect of order F (1, 1736) = 122.8, $p < 0.001$, a significant main effect of tweet type, F (1, 1736) = 64.1, $p < 0.001$ (see the first figure from the left in Figure 3). The order × tweet type interaction effect was also significant, F (1, 1736) = 22.2, $p < 0.001$. Analysis of simple main effect of order on rumor-tweets showed that self-accuracy in the rumor-first group (M = 3.4, SD = 1.7) was higher than the criticism-first group (M = 3.1, SD = 1.6), F (1, 868) = 5.3, $p < 0.05$. Analysis of simple main effect of order on criticism-tweets showed that self-accuracy in the rumor-first group (M = 4.7, SD = 1.6) was higher than the criticism-first group (M = 3.6, SD = 1.7), F (1, 868) = 82.2, $p < 0.001$.

As for others-accuracy, a two-way ANOVA revealed a significant main effect of order, F (1, 1736) = 10.2, $p < 0.01$, and a significant order × tweet type interaction effect, F (1, 1736) = 10.1, $p < 0.01$, but the main effect of tweet type did not reach statistical significance (see the second figure from the left in Figure 3). Analysis of simple main effect of order showed a significant effect on criticism-tweets, F (1, 868) = 22.3 $p < 0.001$, but no significant effect on rumor-tweets.

A two-way ANOVA of importance revealed a significant main effect of order, F (1, 1736) = 64.5, $p < 0.001$. Neither the main effect of tweet type nor the order × tweet interaction did not reach statistical significance. To examine the order effect on rumor-tweets, we performed a one-way ANOVA of importance. The main effect of order was significant, F (1, 868) = 18.8, $p < 0.01$ (see the first figure from the right in Figure 3).

The results of these ANOVAs on self-accuracy, others-accuracy, anxiety, and importance showed that the factors related to the order effect on rumor-tweets were self-accuracy, anxiety, and importance: self-accuracy, anxiety, and importance rates about rumor-tweets were higher in the rumor-first group than criticism-first group.

4.3. The Effects of Criticism Exposure on Intent to Spread Rumors

The result of the different proportion of rumor-stopper between rumor-first group and criticism-first group showed revealed that the proportion of rumor-stopper increased from 32.1% of the rumor-first group to 49.3% of the criticism-first group. However, even in the criticism-first group, 50.7% was rumor-spreader. In order to examine the difference between rumor-stopper and rumor-spreader in the criticism-first group, we focused on the three psychological factors that were related to the order effect on rumor-tweets: self-accuracy, anxiety, and importance.

A one-way ANOVA of self-accuracy, with the type of spread (rumor-stopper vs. weak rumor-spreader vs. strong rumor-spreader) revealed a significant effect of the type of spread, F (1, 457) = 60.4, $p < 0.001$, (see the left figure in Figure 4). Post hoc Tukey's tests indicated that self-accuracy of rumor-stopper (M = 2.4, SD = 1.4) was lower than weak rumor-spreader (M = 3.3, SD = 1.4) ($p < 0.001$), and that self-accuracy of weak rumor-spreader was lower than strong rumor-spreader (M = 4.1, SD = 1.4) ($p < 0.001$).

As for anxiety, a one-way ANOVA with the type of spread (rumor-stopper vs. weak rumor-spreader vs. strong rumor-spreader) showed a significant effect of the type of spread, F (1, 457) = 100.4, $p < 0.001$, (see the center figure in Figure 4). Post hoc Tukey's tests indicated that anxiety of rumor-stopper (M = 2.3, SD = 1.6) was lower than weak rumor-spreader (M = 4.0, SD = 1.8) ($p < 0.001$), and that self-accuracy of weak rumor-spreader was lower than strong rumor-spreader (M = 4.8, SD = 1.5) ($p < 0.005$).

We also performed a one-way ANOVA of importance, with the type of spread (rumor-stopper vs. weak rumor-spreader vs. strong rumor-spreader). As a result, the effect of the type of spread was significant, (1, 457) = 122.34, $p < 0.001$, (see the right figure in Figure 4). Post hoc Tukey's tests showed that importance of rumor-stopper (M = 2.3, SD = 1.4) was lower than weak rumor-spreader (M = 3.8, SD = 1.6) ($p < 0.001$), and that self-accuracy of weak rumor-spreader was lower than strong rumor-spreader (M = 4.8, SD = 1.6) ($p < 0.001$).

5. Discussion

The present study investigated the effect of exposure to criticisms on the spread of rumor-tweets. We compared the two groups, the rumor-first group and the criticism-first group, in their intent to spread rumor-tweets. This investigation was part of a larger endeavor to design a social-technological system that

tries to stop the spread of false rumors by promoting the critical thinking of crowds. Answering *RQ1*, we have shown that exposing people to criticisms can reduce their intent to spread rumors that are associated with the criticisms, providing support for the system. We also analyzed what psychological factors are associated with the lowered intent to spread false rumors after experiencing criticisms. Answering *RQ2*, we found that self-accuracy, anxiety, and importance were related to the effect of exposure to criticisms. In this section, we discuss our results and design recommendations for a social-technological system.

5.1. Exposure to criticisms

In the experiment, our manipulation was the order of rumor and criticism exposures. In the rumor-first group, participants were exposed to rumor-tweets followed by criticism-tweets. Participants in the criticism-first group were exposed to criticism-tweets followed by rumor-tweets. The quantity and the quality of information given by the tweets in the two groups were controlled.

Chi-square analyses revealed that the proportion of responses to stop the spread of rumor-tweets in the criticism-first group increased to 49.3% from 32.1% in the rumor-first group. Exposure to criticisms increased the proportion of people who stop the spread of rumor-tweets approximately 1.5 times. This result indicates that whether a receiver is exposed to rumor or criticism first makes a difference in her decision to spread the rumor. Another interpretation of the result is that, even if a receiver is exposed to a number of criticisms, she will benefit less from this exposure when she sees rumors first than when she sees criticisms before rumors.

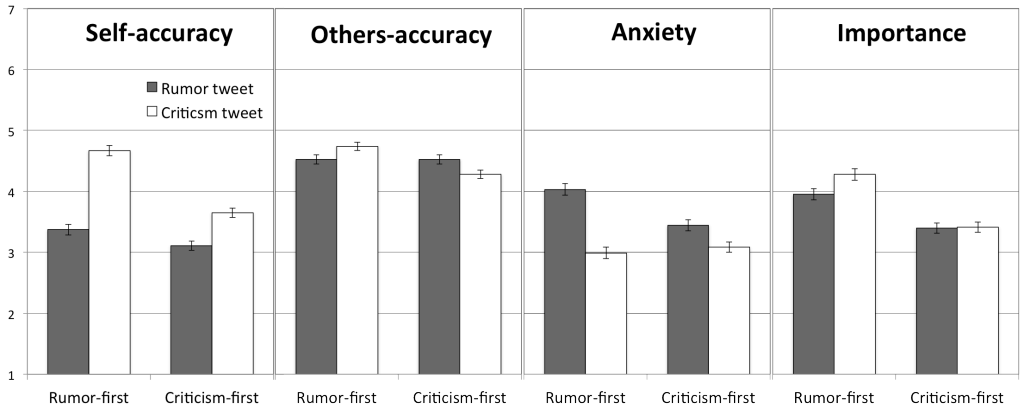

Figure 3. Psychological factors in rumor-tweets and criticism-tweets in the rumor-first group and criticism-first group

The each scale was in the range of 1 (Not at all) to 7 (Highly). Error bars indicate the standard error of the mean.

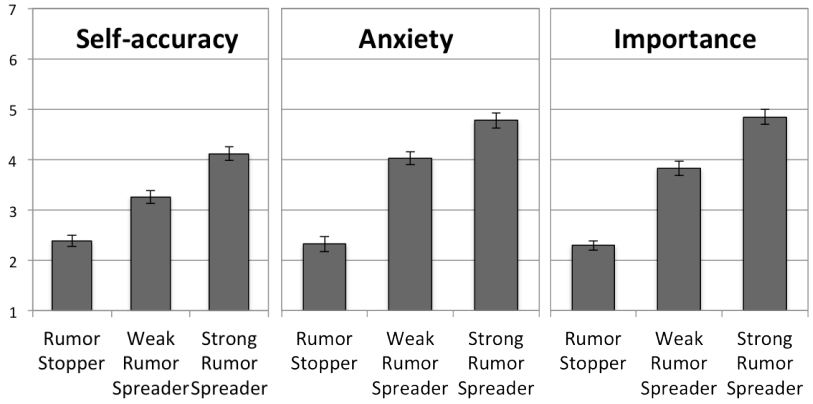

Figure 4. Self-accuracy, anxiety, and importance about rumor-tweets in the criticism-first group

The each scale was in the range of 1 (Not at all) to 7 (Highly). Error bars indicate the standard error of the mean.

Chi-square tests further showed that the population of the strong rumor spreader shrank in the criticism-first group relative to the rumor-first group. Taken together, exposure to criticisms before rumors can increase the number of rumor-stopping responses and decrease the number of far-reaching rumormongering responses.

5.2. Psychological Factors

We analyzed the psychological differences in responses to rumor-tweets in order to examine why exposure to criticisms increased the proportion of rumor-stopping responses. The results revealed three psychological factors that were related to the differences in the spread of rumor-tweets between the two groups: self-accuracy, anxiety, and importance

Participants in the criticism-first group perceived rumor-tweets to be less accurate than those in the rumor-first group. This result indicates that exposure to criticisms reduces the perceived accuracy of the succeeding rumor-tweets, paralleling the findings by previous research that refutations or denials decrease the degree of belief in rumor [3, 4, 15]. In addition, the perceived accuracy of criticism-tweets in the rumor-first group was significantly higher than the criticism-first group.

We found a similar pattern of results for perceived anxiety. Seeing criticisms before rumors reduced anxiety associated with rumor-tweets relative to seeing rumors first. This result is also consistent with previous research findings that denial messages reduce anxiety about rumors [3, 4].

Participants in the criticism-first group also perceived rumor-tweets to be less important than those in the rumor-first group. As Rosnow et al. (1988) showed that importance was related to rumor spread [22], the difference in perceived importance of rumor-tweets can be an explanation for the difference in the proportions of rumor-spreading behavior between the rumor-first group and the criticism-first group.

Although exposure to criticisms increased the rumor-stopping responses, about half of the responses still indicated that the rumor-tweets should be shared. One reason for this is that self-accuracy, anxiety, and importance interact with the effect of exposure to criticisms on rumor-spreading behavior: When the rumor-tweets are perceived as more accurate, the intent to spread the rumor-tweets are stronger; when rumor-tweets cause more anxiety, the intent to spread the rumor-tweets is stronger; when the rumor-tweets are perceived as more importance, the intent to spread the rumor-tweets is also stronger.

5.3. Practical Implications

What implications can we draw from these findings for enhancing the critical thinking of crowds and for designing a social-technological system?

We envision a social-technological system, in which, once a criticism about a rumor is detected, the rumor appears together with the corresponding criticism. By this design, information system itself can enhance the critical thinking of the crowds. On the other hand, if information is shown in the system in the order they are posted, as in some existing social media technologies, the system would not be able to make good use of the critical thinking of the crowds. In this way, the design of the social-technological system can play an important role in determining whether people who receive information from the environment behave like critical thinkers or not.

A social-technological system that promotes the critical thinking of crowds would become important especially in a disaster situation. As previous research showed, a disaster creates an uncertain condition that encourages the generation and the transmission of rumors [2, 23, 31]. It would be inevitable that rumor spread to some extent during a disaster situation. However, as our results show, we can reduce anxiety, perceived accuracy, and importance of rumors by changing the design of the system that transmits information so that criticisms appear together or before rumors. Existing social media should be open to the possibility to have new functions in the future, such as a “disaster mode” or a “critical thinking mode” that detects questionable comments or unreliable information as well as criticisms, and then displays information in an effective order.

6. Limitations

Most of the participants in the current study were the victims who experienced the Great Japan Earthquake. However, the study was conducted about seven months after the disaster and the data were collected in an experimental environment. Thus, there was a gap between what people went through in real disaster scenarios and in laboratory experiments.

Another limitation of the current work was that we measured only the intent to spread rumors before exposure to criticisms and the intent to spread rumors after exposure to criticisms. It is unclear whether or not the intent to spread rumors can be lowered solely by the effect of exposure to criticisms. For example, the intent to spread rumors may be lowered just by exposure to others’ opinions whether they are critical or not. Further experiments are needed to clarify the

effects of criticism exposure on the rumor-spreading behavior.

Third, we only examined 10 rumor-tweets and 10 associated criticism-tweets that a human analyzed one by one. In a social-technological system, this approach is not ideal. Instead, it would be useful to have a technique to automatically process tweets. In future work, we plan to formulate an algorithm to detect criticisms and rumors based on words used in messages as well as a method to determine which criticisms, if many exist, would be presented with the associated rumors. We are also testing other aspects of a social-technological system, such as how social information (e.g., how many people write criticisms) and credibility (e.g., whether the source of a message is reliable) influences people's decisions to trust and spread messages.

7. Concluding remarks

False rumors during disasters would never go away. It is human nature to transmit rumors under uncertainty. In addition to trying to educate everyone on critical thinking, we might be able to design a social-technological system that can enhance the critical thinking of crowds. In this system, not everyone needs to be critical thinker all the time. Instead, people can benefit from each other's critical thinking. The system is an indirect collaboration system in that users work together through their output, in this case criticisms, to solve a social problem, here to inactivate false rumors. The present study shows that such a system is promising.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant No. IIS-1138658. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors thank three anonymous reviewers for their thoughtful comments.

References

[1] Allport, G.W. and Postman, L. *The Psychology of Rumor*. Henry Holt and Company, New York, New York, 1947.

[2] Anthony, S. Anxiety and Rumor. *The Journal of Social Psychology,* 89(1), 1973, 91–98.

[3] Bordia, P., DiFonzo, N., Haines, R., and Chaseling, E. Rumors Denials as Persuasive Messages: Effects of Personal Relevance, Source, and Message Characteristics. *Journal of Applied Social Psychology,* 35(6), 2005, 1301–1331.

[4] Bordia, P., DiFonzo, N., and Schulz, C.A. Source Characteristics in Denying Rumors of Organizational Closure: Honesty is the Best Policy. *Journal of Applied Social Psychology,* 30(11), 2000, 2309–2321.

[5] Chorus, A. The Basic Law of Rumor. *Journal of Abnormal Psychology*, 48(2), 1953, 313–314.

[6] Dam, G. ten, Volman, M., and Tendam, G. Critical Thinking as a Citizenship Competence: Teaching Strategies. *Learning and Instruction,* 14(4), 2004, 359–379.

[7] DiFonzo, N. and Bordia, P. *Rumor Psychology: Social and Organizational Approaches*. American Psychological Association, Washington, DC, 2007.

[8] Difonzo, N. and Bordia, P. How top PR Professionals Handle Hearsay: Corporate Rumors, Their Effects, and Strategies to Manage Them. *Public Relations Review,* 26(2), 2000, 173–190.

[9] Ennis, R.H. Critical Thinking Dispositions: Their Nature and Assessability. *Informal Logic,* 18(2), 1996, 165-182.

[10] Feuerstein, M. Media Literacy in Support of Critical Thinking. *Journal of Educational Media,* 24(1), 1999, 43–54.

[11] Frijters, S., Dam, G. ten, and Rijlaarsdam, G. Effects of Dialogic Learning on Value-Loaded Critical Thinking. *Learning and Instruction,* 18(1), 2008, 66–82.

[12] Giroux, A. Promoting Critical Thinking: Rediscovery in Teaching. *Canadian Journal of Education,* 16(4), 1991, 397–419.

[13] Halpern, D.F. *Thought and Knowledge: An Introduction to Critical Thinking,* Lawrence Erlbaum Associates, Inc. Mahwah, New Jersey, 2002.

[14] Klaczynski, P. a and Lavallee, K.L. Domain-Specific Identity, Epistemic Regulation, and Intellectual Ability as Predictors of Belief-Biased Reasoning: A Dual-Process Perspective. *Journal of Experimental Child Psychology,* 92(1), 2005, 1–24.

[15] Koller, M. Rebutting Accusations: When Does It Work, When Does It Fail? *European Journal of Social Psychology,* 23(4), 1993, 373–389.

[16] Mendoza, M. and Poblete, B. Twitter Under Crisis: Can We Trust What We RT? *In Proceedings of the First Workshop on Social Media Analytics*, 2010, 71–79.

[17] Ogiue, C. *An Examination of Rumors and False Rumors during the Great East Japan Earthquake (検証 東日本大震災の流言・デマ)*. Kobunsha, Tokyo, 2011.

[18] Oh, O., Kwon, K.H., and Rao, H.R. An Exploration of Social Media in Extreme Events: Rumor Theory and Twitter during the Haiti Earthquake 2010. *In Proceedings of the International Conference on Information Systems (ICIS)*, 2010, 231.

[19] Penningroth, S. and Despain, L. A Course Designed to Improve Psychological Critical Thinking. *Teaching of Psychology,* 34(3), 2007, 153–158.

[20] Prasad, J. The Psychology of Rumor: A Study relating to the Great Indian Earthquake of 1934. *British Journal of Psychology: General Section,* 26(1), 1935, 1–15.

[21] Prasad, J. A Comparative Study of Rumours and Reports in Earthquakes. *British Journal of Psychology: General Section,* 41(3-4), 1950, 129–144.
[22] Rosnow, R.L., Esposito, J.L., and Gibney, L. Factors Influencing Rumor Spreading: Replication and Extension. *Language Communication,* 8(1), 1988, 29–42.
[23] Rosnow, R.L. Psychology of Rumor Reconsidered. *Psychological Bulletin,* 87(3), 1980, 578–591.
[24] Rosnow, R.L. Inside Rumor: A Personal Journey. *American Psychologist,* 46(5), 1991, 484–496.
[25] Sa, W.C., West, R.F., and Stanovich, K.E. The Domain Specificity and Generality of Belief Bias: Searching for A Generalizable Critical Thinking Skill. *Journal of Educational Psychology,* 91(3), 1999, 497–510.
[26] Semiocast. Brazil Becomes 2nd Country on Twitter, Japan 3rd, Netherlands Most Active Country. 2012. http://semiocast.com/publications/2012_01_31_Brazil_becomes_2nd_country_on_Twitter_superseds_Japan.
[27] Shibutani, T. *Improvised News: A Sociological Study of Rumor*. The Bobbs-Merrill Company, Inc., Indianapolis, Indiana, 1966.
[28] Sinha, D. Behaviour in A Catastrophic Situation: A Psychological Study of Reports and Rumors. *British Journal of Psychology: General Section,* 43(3), 1952, 200–209.
[29] Sun, J., Jin, S., and Cao, S. A Qualitative Psychological Analysis of the Rumors Appeared after A Disaster: The Evidence from the Rumors Appeared after "5·12 Wenchuan Earthquake". *Advances in Psychological Science,* 17(3), 2009, 602–609.
[30] Tachiiri, K. *An Examination of the Great East Japan Earthquake: What Did Social Media Report during the Disaster? (検証 東日本大震災 そのときソーシャルメディアは何を伝えたか?)*. Discover Twenty One, Tokyo, 2011.
[31] Walker, C.J. and Beckerle, C.A. The Effect of State Anxiety on Rumor Transmission. *Journal of Social Behavior and Personality,* 2(3), 1987, 353–360.
[32] Yang, S. and Chung, T. Experimental Study of Teaching Critical Thinking in Civic Education in Taiwanese Junior High School. *British Journal of Educational Psychology,* 79(1) 2009, 29–55.

Appendix. Lists of rumor-tweets and corresponding criticism-tweets

Tweet type		Summary of tweet
1	**Rumor**	My friend at an insurance company said "Cancer insurance commercials stopped after the nuclear accidents."
	Criticism	If you put it that way, sure. But that's a false rumor.
2	**Rumor**	According to my friend, a radioactive material was detected from urine after he ate sushi.
	Criticism	It's unclear the radioactive material was caused by the fish.
3	**Rumor**	Toxic substance will drop with rain due to an explosion at Cosmo oil company.
	Criticism	That's a definitely false rumor. NHK denied it.
4	**Rumor**	Medical license is deprived by MEXT if a doctor gives a certificate of being exposed to radiation.
	Criticism	The license cannot be deprived easily by MEXT or MHLW.
5	**Rumor**	Robberies and rapes occurred during the Kobe earthquake.
	Criticism	Few robberies and rapes occurred. Victims helped each other orderly. Why do you spread lies and false rumors? Stop it.
6	**Rumor**	It was denied strongly, but after all, the meltdown occurred.
	Criticism	Has the possibility of a meltdown been pointed out, hasn't it?
7	**Rumor**	Airdrop of supplies is not allowed in Japan!
	Criticism	Airdrop is not prohibited by the law.
8	**Rumor**	Tokyo Electric Power Co.'s workers run and left. They were drinking in other city.
	Criticism	Tokyo Electric Power Co. "The workers were found dead."
9	**Rumor**	Did anyone watch "Senior vice transport minister Tsujimoto protested against the rescue operation by US army" on NHK?
	Criticism	There's no source but the tweet, so it would be a rumor.
10	**Rumor**	Chubu, Kansai, and Kyusyu Electric Power companies are beginning to transfer electricity to Kanto. Please cooperate!
	Criticism	Transfer is impossible because of the difference in frequency.