Deterring the Spread of Misinformation on Social Network Sites: A Social Cognitive Theory-guided Intervention

Xinran Chen **Sei-Ching Joanna Sin** **Yin-Leng Theng** **Chei Sian Lee**
Wee Kim Wee School of Communication and Information, Nanyang Technological University
31 Nanyang Link, Singapore 637718
{chen0872, joanna.sin, tyltheng, leecs}@ntu.edu.sg

ABSTRACT
As more individuals turn to social network sites (SNSs) for information, the spread of misinformation in these sites is becoming a greater concern. Not only can misinformation cause individual users anxiety and harm, but it can also prevent SNSs from realizing their full potential as trustworthy sources of information. This study proposed and tested an intervention-based strategy that was designed to discourage behavior that promotes the spread of misinformation. Guided by the social cognitive theory (SCT), the intervention sought to modify users' *outcome expectations* by presenting them with a message that highlighted the negative consequences of misinformation. To investigate the effectiveness of this intervention message, a classical experiment was conducted on-line with 131 college-student participants. In the study's experimental group, the ANOVA results showed that the intervention effectively reduced the total number of "Likes" and "Shares" for postings that provided misinformation. Future development and testing of this SCT-guided, outcome-expectations-based intervention is promising.

Keywords
Information behavior, social media

INTRODUCTION
For many individuals, especially college students, social network sites (SNSs) are becoming significant channels for information (Kim, Sin, & Yoo-Lee, 2013). While SNSs lower the costs of communication and exchanging information (Ellison & boyd, 2013), they also reduce the cost of spreading misinformation (inaccurate information, regardless of the spreader's intent). Misinformation is prevalent within SNSs (Chen & Sin, 2013; World Economic Forum, 2014). It propagates to a large number of people with alarming speed, and is often aided by normal users who pass the misinformation unmindfully and without any nefarious intent (Starbird, Maddock, Orand, Achterman, & Mason, 2014). Widespread misinformation causes confusion and unnecessary anxiety within the public (Budak, Agrawal, & Abbadi, 2011; Tanaka, Sakamoto, & Matsuka, 2013). It prevents SNSs from growing into a powerful and trustworthy source of information. Proactive efforts are needed to deter SNS users' from spreading misinformation.

This study proposes an intervention message designed to modify users' expectations about the consequences of spreading misinformation. Social cognitive theory (SCT), and especially the concept of outcome expectations (Bandura, 1986), guided the design of the intervention. With the objective of testing the effectiveness of the proposed intervention, this first experiment focused on a popular SNS (Facebook) and a frequent user group (college students). The research question was: *Can exposure to the SCT-guided intervention message change college students' spreading of misinformation on Facebook?*

To capture activities that spread misinformation in a more comprehensive manner, the spread of misinformation on Facebook was operationalized as the number of times that participants "Liked" or "Shared" posts containing misinformation, without adding comments that questioned the veracity of the posts. Following the intervention, a significant drop in the number of misinformation spread by a participant would suggest a successful intervention.

This study can make several contributions. First, at a conceptual level, because the study proposes and examines a theory-guided original intervention to deter the spread of misinformation within SNSs, it differs from, and will complement, existing approaches (e.g., developing detection algorithms, improving users' cognitive skills, and increasing efficacy in assessments of credibility). Second, the study can add to literature about SCT by expanding the areas of behavioral change where SCT is applied. Third, and in terms of practical implications, the intervention message, if tested to be effective, can be incorporated into the design of social media and information literacy training to deter the spread of misinformation, contributing to the development of social media and reducing the negative consequences that misinformation can bring to individuals.

ASIST 2015,November 6-10, 2015, St. Louis, MO, USA.

LITERATURE REVIEW

Approaches to Controlling the Spread of Misinformation

Attempts to address the spread of misinformation can be classified into three main categories: diffusion detection, correction, and education. First, there are considerable efforts among computer and information scientists in detecting the diffusion of misinformation (Budak, Agrawal, & Abbadi, 2011; Castillo, Mendoza, & Poblete, 2011; Doerr, Fouz, & Friedrich, 2012). Most studies aim to understand the propagation of misinformation at a macroscopic level, personal factors are rarely investigated. Second, the correction approach uses the "self-correcting crowd" to report and correct misinformation (Mendoza, Poblete, & Castillo, 2010). It works by empowering users with correct information (Zhou & Zhang, 2007) and clarification tools (Ennals, Trushkowsky, & Agosta, 2010). However, this approach requires actions and initiatives from knowledgeable users, which are not always available. Finally, education is often recommended in the field of information literacy (Calvert, 2001; Floridi, 1996). Researchers have identified skill sets for identifying misinformation to teach to users (Fitzgerald, 1997; Walsh, 2010). Nevertheless, some users choose to share a message even when they are aware of its inaccuracy (Chen & Sin, 2013; Karlova & Fisher, 2013; Li & Sakamoto, 2014).

Overall, extant interventions have yet to focus on the social contexts and personal motivations and expectations that compel users to spread misinformation. SNSs are not only platforms for the distribution of information, but digital spaces that allow users to pursue leisure activities and to build and maintain social capital. For example, if a message is entertaining, and sharing it allows users to interact with friends and gain popularity within a SNS, then users may share a SNS post regardless of its perceived informational value. Given that individual behaviors in SNSs can extend beyond cognitive reasoning, beyond extant efforts to improve users' cognitive abilities, information literacy, and self-efficacy, we proposed a fresh angle of intervention that focuses on changing users' expectations about the negative outcomes of spreading misinformation.

Outcome Expectations in Social Cognitive Theory

Developed by Albert Bandura (1986), SCT advocates a model of emergent interactive agency, wherein behavior, personal factors, and environmental influences are conceptualized as interacting determinants. While SCT acknowledges that environmental factors can influence behaviors, it also recognizes the role of personal factors such as beliefs, expectations, and self-perceptions in shaping behaviors. Although SCT (specifically the construct of outcome expectations) has gone untested in deterring the spread of misinformation within SNSs, it has frequently and successfully been applied to behavioral change in fields such as health (Ashford, Edmunds, & French, 2010; Williams & French, 2011). SCT is promising for the current topic and for studies in information behavior (IB). To illustrate this point, Wilson's IB model incorporated social learning theory (the precursor of SCT) as a mechanism for activating IB (Wilson, 1999).

Outcome expectations refers to one's belief of the likely consequences of one's performance (Bandura, 1997). It is a primary construct of SCT that has been found to directly influence behavior (Bandura, 2004). SCT proposes that people will act in ways that they believe will bring positive outcomes and avoid behaviors that produce negative outcomes. Providing information about consequence is a well-tested behavioral change technique that is modeled after SCT (Abraham & Michie, 2008). Outcome expectations can contribute independently to behavioral change, especially when achieving an outcome requires minimal skill and performance (Bandura, 1997). This fits the common spread of misinformation within SNSs because sharing misinforming posts requires very little effort (clicking a button or two). We posit that changing users' outcome expectations can help deter them from spreading misinformation on SNSs.

METHOD

The study used a classical experimental design that is recommended due to its strong mechanisms in controlling internal validity threats. Participants were randomly assigned to a control group or an experimental group. Similar to many experimental studies, a cover story was used to mask the research's focus and to avoid the priming effects and social desirability effects. The study was introduced as an evaluation of a particular Facebook-post recommendation system.

All participants went through a pre-test where they viewed 10 posts that were designed to look like Facebook posts (five contained misinformation and five were accurate). After viewing each post, participants were given response options that mirrored actual Facebook settings ("Like," "Share," "Comment," or no response at all). The *pre-test spreading score* was calculated by counting the number of times that participants "Liked" or "Shared" the five posts that contained misinformation without questioning their veracity in the "Comment" area.

The next stage of the study involved the intervention. While the experimental group read an intervention message, the control group read a message that contained irrelevant information. The intervention message was written in the form of a magazine article (about 700 words) that focused on certain potential negative outcomes of misinformation such as monetary or physical harm. Based on SCT, it is likely that increasing participants' expectations that the spread of misinformation produces negative outcomes will deter them from sharing misinforming posts.

During the post-test, each participant used the response options listed above on an additional set of 10 posts (five with misinformation and five without). The *post-test spreading score* was calculated in the same way as the *pre-test spreading score*. The pre-test and post-test

misinformation posts were pilot-tested and designed to parallel one another. For example, the study contained two fake advertisements for cell phone software upgrades, where one claimed that the software upgrade could make a phone completely waterproof, and the other said that it could allow a phone to recharge inside of a microwave oven. These parallel posts helped to reduce the differences between the pre- and post-test items, thus decreasing the likelihood of introducing extra noise into the instrument. After the post-test, participants answered questions about their demographics and SNS usage frequency. Participants were debriefed immediately after the experiment, as required by human subject protection guidelines.

The experiment was delivered through an online survey platform and took approximately 30 minutes to complete. The participants were undergraduates attending a public university in Singapore, who were between the ages of 18- and 25-years old and visited Facebook at least once a week. The sample was drawn using stratified systematic sampling. Invitations for the online survey were emailed to the students. Since participation was voluntary, not everyone who was invited took part in the study. The data were analyzed using descriptive statistics and analysis of variance (ANOVA) in SPSS.

A total of 131 valid responses were received and analyzed. There were more female (n = 83) than male (n = 48) participants, and their average age was 21.3 years (SD = 1.75). All participants (N = 131) visited Facebook at least once a week, and over 60% (n = 81) visited more than once a day. Results of Chi-square tests for independence verified that the experimental and control groups showed no statistically significant differences in their SNS usage and demographics.

FINDINGS

Table 1 presents the pre-test and post-test spreading scores. Because some participants did not complete the entire experiment, the numbers of valid responses from the control group and the experimental group varied. To test the effect of the intervention message on the spread of misinformation, a 2 × 2 mixed-design ANOVA test was conducted with *time* (pre-test, post-test) as the within-subjects factor and *condition* (experimental group, control group) as the between-subjects factor.

Time showed a significant main effect, $F(1, 129) = 23.4$, $p < .001$ (Table 2). This revealed that the number of "Likes" and "Shares" of misinformation differed significantly, specifically, the number decreased between the pre-test and the post-test. *Condition* showed no significant main effect, $F(1, 129) = 0.169$, $p = .682$. There was a significant *Time* × *Condition* interaction effect, $F(1, 129) = 6.77$, $p = .010$. Figure 1 illustrates the interaction effects.

DISCUSSION

As hypothesized, the intervention message was found to reduce the spread of misinformation among the study's participants. The significant *Time* × *Condition* interaction showed that after receiving the intervention, participants in the experimental group spread significantly fewer posts than the control group. Specifically, out of the five posts containing misinformation, the experimental group spread nearly one post less (M = 0.91) after intervention, and the decrease in the control group was marginal (M = 0.27). This provides supporting evidence that the M = 0.91 reduction among the experimental group came in part from the intervention (and not solely from noises in the experimental setting or maturation effects), as the control group did not experience the same reduction level. While the intervention consisted of a single session that was relatively brief (about five minutes long), the study's findings suggest that it was able to induce in the experimental group a statistically significant drop in their spread of misinformation. These findings are promising. Even though users did not completely stop the spread of misinformation after the intervention, the feasibility and effectiveness of the intervention were preliminarily demonstrated. A limitation of the study is that it tested only the short term effect of the intervention; further testing with follow-up experiments will be needed to measure if (and when) the effects of intervention wear off. Another area of testing may involve the development of complementary studies that investigate

Condition	*N*	Pre-test		Post-test	
		M	*SD*	*M*	*SD*
Experimental	76	1.92	1.66	1.01	1.34
Control	55	1.51	1.46	1.24	1.36
Total	131	1.75	1.59	1.11	1.35

Table 1. Pre-test and post-test spreading scores

Variable	*SS*	*df*	*Error df*	*F*
Time (T)	22.2	1	129	23.4**
Condition (C)	0.568	1	129	0.169
T × C	6.44	1	129	6.77*

Table 2. ANOVA results of the effects of intervention

Note: ** $p < 0.01$; * $p < .05$

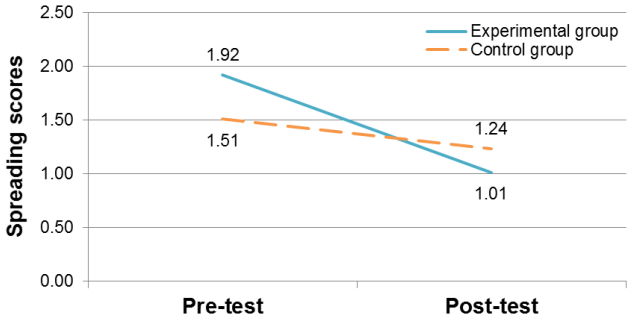

Figure 1. Estimated marginal means of the pre-test and post-test spreading scores

the effectiveness of intervention outside of experimental sessions. This will help verify the study's external validity and especially its naturalistic generalization.

The reduction in the spread of misinformation among individuals is of paramount importance. Findings from the current study suggest that the introduction of interventions that target users' outcome expectations can reduce the sharing of misinformation in the experimental setting. Further efforts to focus on the development of specific intervention strategies, via the incorporation of other SCT constructs and by testing these strategies in additional social media platforms, are strongly recommended.

REFERENCES

Abraham, C., & Michie, S. (2008). A taxonomy of behavior change techniques used in interventions. *Health Psychology, 27*(3), 379.

Ashford, S., Edmunds, J., & French, D. P. (2010). What is the best way to change self-efficacy to promote lifestyle and recreational physical activity? A systematic review with meta-analysis. *British Journal of Health Psychology, 15*(2), 265-288.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman.

Bandura, A. (2004). Health promotion by social cognitive means. *Health Education & Behavior, 31*(2), 143-164.

Budak, C., Agrawal, D., & Abbadi, A. E. (2011). Limiting the spread of misinformation in social networks. *Proceedings of the 20th International World Wide Web Conference*, 665-674.

Calvert, P. J. (2001). Scholarly misconduct and misinformation on the World Wide Web. *Electronic Library, 19*(4), 232-240.

Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on Twitter. *Proceedings of the 20th International World Wide Web Conference*, 675-684.

Chen, X., & Sin, S.-C. J. (2013). 'Misinformation? What of it?' Motivations and individual differences in misinformation sharing on social media. *Proceedings of the American Society for Information Science and Technology, 50*(1), 1-4.

Doerr, B., Fouz, M., & Friedrich, T. (2012). Why rumors spread so quickly in social networks. *Communications of the ACM, 55*(6), 70. doi: 10.1145/2184319.2184338

Ellison, N. B., & boyd, d. m. (2013). Sociality through social network sites. In W. H. Dutton (Ed.), *The Oxford Handbook of Internet Studies* (pp. 151-172). Oxford: Oxford University Press.

Ennals, R., Trushkowsky, B., & Agosta, J. M. (2010). Highlighting disputed claims on the web. *Proceedings of the 19th International Conference on World Wide Web*, 341-350.

Fitzgerald, M. A. (1997). Misinformation on the Internet: Applying evaluation skills to online information. *Emergency Librarian, 24*(3), 9-14.

Floridi, L. (1996). Brave.Net.World: The Internet as a disinformation superhighway? *Electronic Library, 14*(6), 509-514.

Karlova, N. A., & Fisher, K. E. (2013). A social diffusion model of misinformation and disinformation for understanding human information behaviour. *Information Research, 18* (1).

Kim, K.-S., Sin, S.-C. J., & Yoo-Lee, E. (2013). Undergraduates' use of social media as information sources. *College & Research Libraries, 75*(4), 442-457.

Li, H., & Sakamoto, Y. (2014). Social impacts in social media: An examination of perceived truthfulness and sharing of information. *Computers in Human Behavior, 41*, 278-287. doi: 10.1016/j.chb.2014.08.009

Mendoza, M., Poblete, B., & Castillo, C. (2010). Twitter under crisis: Can we trust what we RT? *Proceedings of the First Workshop on Social Media Analytics*, 71-79.

Starbird, K., Maddock, J., Orand, M., Achterman, P., & Mason, R. M. (2014). Rumors, false flags, and digital vigilantes: Misinformation on Twitter after the 2013 Boston marathon bombing. *Proceedings of the iConference 2014*.

Tanaka, Y., Sakamoto, Y., & Matsuka, T. (2013). Toward a social-technological system that inactivates false rumors through the critical thinking of crowds. *Proceedings of the 46th Hawaii International Conference on System Sciences*, 649-658.

Walsh, J. (2010). Librarians and controlling disinformation: Is multi-literacy instruction the answer? *Library Review, 59*(7), 498-511. doi: 10.1108/00242531011065091

Williams, S., & French, D. (2011). What are the most effective intervention techniques for changing physical activity self-efficacy and physical activity behaviour - and are they the same? *Health Education Research, 26*(2), 308-322.

Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation, 55*(3), 249-270.

World Economic Forum. (2014). Top 10 trends of 2014: 10. The rapid spread of misinformation online. from http://bit.ly/1edZQQF

Zhou, L., & Zhang, D. (2007). An ontology-supported misinformation model: Toward a digital misinformation library. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 37*(5), 804-813.