If you see something, say something (to others): A spread of misinformation and

socially-contingent corrections within social networks

Hyunjin Song

Department of Communication, University of Vienna, Austria

Draft in progress. Please do not cite without permission. Please direct any questions and

inquiries to hyunjin.song@univie.ac.at

Author Note

Abstract

As the citizens' news consumption is increasingly driven by online sources, the propagation of misinformation and so-called "fake news" on those platforms become an increasing concern for the public and policy makers. Our goal in this contribution is to offer a more systematic assessment of underlying mechanisms of misinformation spreading and its correction, combining a macro social contextual factor and individuals' cognitive basis of adopting misinformation into a more integrated, dynamic system model perspective. We first review existing evidence concerning individuals' cognitive basis of adopting such misinformation, and social context of which exposure to misinformation and its corrections are received. Next, adopting a well-known class of an epidemic model of virus infection and recovery, we combine this micro and macro dynamics into comprehensive, integrated model of misinformation diffusion on social networks. We do so by focusing on the distinction between simple contagion of misinformation vs. complex contagion of adopting corrective messages. Relying on Agent-based simulations, we further explore various boundary conditions of such dynamics, aiming to uncover how and when such misinformation propagates into the public, as well as what factors facilitate or hinder such diffusion process.

If you see something, say something (to others): A spread of misinformation and

socially-contingent corrections within social networks

Citizens across the worlds are experiencing major changes in their news environments with the development of digital media. One of the most dramatic changes in the news environment in recent decades involves the role social networking sites (SNS) such as Facebook and Twitter play as a primary source of news outlets. Not only citizens' news consumptions are increasingly driven by such online sources (Shearer & Gottfried, 2017), but it also appears that citizens themselves are actively participating in news dissemination on those platforms by sharing news contents with their peers (e.g., Lee & Song, 2017; Weeks & Holbert, 2013).

An effective deliberation among public is regarded as a keystone of thriving democracies, and modern political systems squarely depend on informed decisions of citizens in that regard (Delli Carpini & Keeter, 1996; Strömbäck, 2005). Yet, a propagation of rumors, misinformation, and so-called "Fake news" on those platforms becomes an increasing concern for the public and policy makers alike (Allcott & Gentzkow, 2017). There are abundant, yet still fragmentary, evidence of viral spreads of unsubstantiated, often factually dubious (mis)information that potentially affecting millions of citizens across the globe – as evidenced in recent 2016 U.S. presidential election (Allcott & Gentzkow, 2017; Giglietto, Iannelli, Rossi, & Valeriani, 2016; Guess, Nyhan, & Reifler, 2018) and in Brexit votes (The New York Times, 2017). While a wide circulation of factually dubious information is not entirely new to politics, a growing trend of digitally disseminated rumors and misinformations – often termed as a "Fake news" phenomenon – is increasingly recognized as a serious threat to liberal democratic societies (Allcott & Gentzkow, 2017; Lewandowsky, Ecker, & Cook, 2017). Either based on unsubstantiated rumors or based on factually wrong beliefs, many of the misinformed behave differently than those who are accurately informed (Garrett & Weeks, 2013; Kuklinski, Quirk, Jerit, Schwieder, & Rich, 2000). They often disagree about basic facts about many political and public issues (e.g., Nisbet, Cooper, & Garrett, 2015), and continue to believe and rely on such false information when evaluating political matters (Nyhan & Reifler, 2010; Thorson, 2016).

Along with these trends, there has been an growing interest among scholars on how people process and maintain factually false (or at least factually dubious) information from the

perspectives of an individual's cognitive and affective mechanisms (Garrett, Weeks, & Neo, 2016; Kuklinski et al., 2000; Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012; Lewandowsky et al., 2017; Weeks, 2015). These studies have generated a valuable insights of how individuals maintain false beliefs, and how corrections to such false beliefs are received and processed under various scenarios (Lewandowsky et al., 2017; Thorson, 2016). However, despite growing interest and continued research effort to better understand the nature and its exact mechanism, what we know about the spread of misinformation and fake news on online social networks is largely based on limited evidence due to its complex nature of the problem.

Against this backdrop, our goal in this contribution is to offer a more systematic assessment of underlying mechanisms of misinformation spreading and its correction, focusing on one's *social contexts* in which such (mis)information and corrective messages are received and processed. We argue that while *exposure* to (mis)information is likely to follow a simple contagion process, *changes* in one's beliefs regarding such (mis)information – which ultimately *the* goal of corrective messages – is likely to be, in Centola and Macy's (2007) term, a "complex contagion" where such changes require multiple sources of affirmation and reinforcement compared to simple contagion process. As a result, the effects of fact-checking and corrective messages are likely to be highly *socially* contingent, yet the vast majority of prior studies have not considered this possibility to date (Margolin, Hannak, & Weber, 2017).

In what follows, we first review existing evidence regarding political misperceptions and the effect of fact-checking (i.e., corrective) messages. We advance our dynamic system perspective by combining an individual-level cognitive and affective basis of adopting such misinformation with a social context of which an exposure to misinformation and corrections are received. Based on a well-known class of an epidemic model of virus infection and recovery, we propose an integrated model of misinformation diffusion and socially-contingent corrections on social networks, with a special focus on the differences between simple contagion of misinformation and complex contagion of corrections. Relying on Agent-based simulations, we further explore boundary conditions of such dynamics, aiming to uncover how and when such misinformation propagates into the public, as well as what factors facilitate or hinder such diffusion process.

A Psychology of Fake News, Misperceptions, and Corrections

Following Allcott and Gentzkow's (2017) definition, we define *Fake news* as "distorted signals uncorrelated with the truth" (p. 212). This encompasses several related concepts, such as misinformation, rumors, and disinformation. Literature on this topic generally maintain loosely defined, but at the same time highly interrelated, definitions of those related terms. For instance, (political) rumors are often defined as "unsubstantiated claims about candidates and issues that are often false" (Weeks & Garrett, 2014, p. 401). Similarly, misinformation (or misperceptions) are defined as factual information (or beliefs) "that are false or contradict the best available evidence in the public domain" (Flynn, Nyhan, & Reifler, 2017, p. 128). In relation to this, *dis*information campaigns often denote organized, strategic efforts that trying to sway public opinion using rumors and misinformation (Garrett, 2017; Lewandowsky et al., 2012). Understood in this way, fake news often exclude unintentional reporting mistakes, parodies and satires, or unverifiable conspiracy theories (Allcott & Gentzkow, 2017). While term *fake news* often than not additionally entail specific pseudo-journalistic styles that mimic legitimate news sources to intentionally deceiving audiences (Egelhofer & Lecheler, 2017), we use term "fake news" somewhat loosely, denoting any type of misinformation – information that is not supported by best-available evidence – that is deliberatively circulated among publics.[1]

Literature on misinformation and its persistence often converges to the observation that publics' exposure to and acceptance of misinformation are largely driven by one's motivated consistency needs. That is, people are more often than not likely to be disproportionately drawn by information that conforms to their partisan priors (Guess et al., 2018; Weeks & Garrett, 2014), and also more likely to accept and endorse such messages (Kunda, 1990; Nyhan & Reifler, 2010). A mounting evidence – largely based on Kunda's (1990) or on Taber and Lodge's (2006) motivated reasoning framework – suggests that citizens tend to evaluate attitudinally congruent information as more convincing and valid *regardless of its truth-value*, while inconsistent information is likely to be perceived as weak and therefore likely to be rejected (Taber & Lodge, 2006; Weeks, 2015). Therefore, it is perhaps not surprising to find that most of the prior studies based on motivated reasoning framework document that fact-checking messages (sometimes denoted as "corrective" or "debunking" messages in the literature) have

only limited effects due to inherent tendency of humans to directionally process (politically)

relevant information (Flynn et al., 2017; Taber & Lodge, 2006; Thorson, 2016). Or worse, often

such corrective messages may backfire (i.e., more likely to increase than lower the endorsement

of false beliefs: Nyhan & Reifler, 2010).

Another line of studies focusing on an individual-level cognitive processing and memory

suggests that while attitudinally-congruent misinformation creates automatic and strong

affective responses (therefore more easily remembered and readily available in their memory),

corrections rarely produce such responses, and therefore requires more resources to recall and

incorporate into relevant judgments (Lewandowsky et al., 2012; Thorson, 2016). Also, since

misinformations tend to form a coherent mental model of an relevant event based on their

partisan schema or based on stereotypes (e.g., Garrett, Nisbet, & Lynch, 2013), people tend to

fill any gaps caused by corrections (that invalidates some parts of the mental model) with flawed

but attitudinally congruent (mis)information available in memory (Lewandowsky et al., 2012),

especially when correction messages does not update the initial mental model that justifies

misinformation (Chan, Jones, Jamieson, & Albarracín, 2017). Studies also find that this effect is

much likely when initial misinformation is more readily accessible due to fluency bias in one's

cognitive processing (Lazer et al., 2017; Lewandowsky et al., 2012), when individuals can

generate (counter-arguing) reasons in support for initial misperceptions (Chan et al., 2017;

Garrett et al., 2013), and most importantly, *even when they remember such information is

factually incorrect* (Lewandowsky et al., 2012). Due to asymmetrical nature in one's cognitive

processing fluency, people have to rely on more deliberative process to recall attitudinally

inconsistent correction messages (which requires significant resources), whereas

misinformation is more easily processed and automatically activated. Therefore, even in the

face of corrections, the effect of misperceptions linger and continue to exert influence on

subsequent judgments via spontaneous (i.e., automatic) affective process (e.g., Thorson, 2016).

However, other studies also find that, under certain situations, citizens *indeed* can adhere

factual information based on correction messages despite of their perpetual partisan bias (e.g.,

Garrett & Weeks, 2013; Weeks, 2015; Wood & Porter, 2018). Notably, Wood and Porter's

(2018) series of experiments document that

References

Allcott, H. & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*, 211–36.

Centola, D. & Macy, M. (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology*, *113*, 702–734.

Chan, M.-p. S., Jones, C. R., Jamieson, K. H., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, *28*, 1531–1546. doi:10.1177/0956797617714579

Delli Carpini, M. X. & Keeter, S. (1996). *What americans know about politics and why it matters*. Yale University Press.

Egelhofer, J. L. & Lecheler, S. (2017, September). Conceptualizing "Fake News" for political communication research: A framework and research agenda. Paper presented at The Third Annual IJPP Conference, Oxford, UK.

Flynn, D., Nyhan, B., & Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Political Psychology*, *38*(S1), 127–150.

Garrett, R. K. (2017). The "Echo Chamber" distraction: Disinformation campaigns are the problem, not audience fragmentation. *Journal of Applied Research in Memory and Cognition*, *6*, 370–376. doi:10.1016/j.jarmac.2017.09.011

Garrett, R. K., Nisbet, E. C., & Lynch, E. K. (2013). Undermining the corrective effects of media-based political fact checking? The role of contextual cues and naïve theory. *Journal of Communication*, *63*, 617–637.

Garrett, R. K. & Weeks, B. E. (2013). The promise and peril of real-time corrections to political misperceptions. In *Proceedings of the 2013 conference on computer supported cooperative work* (pp. 1047–1058). CSCW '13. San Antonio, Texas, USA: ACM. doi:10.1145/2441776.2441895

Garrett, R. K., Weeks, B. E., & Neo, R. L. (2016). Driving a wedge between evidence and beliefs: How online ideological news exposure promotes political misperceptions. *Journal of Computer-Mediated Communication*, *21*, 331–348.

Giglietto, F., Iannelli, L., Rossi, L., & Valeriani, A. (2016). Fakes, news and the election: A new taxonomy for the study of misleading information within the hybrid media system. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2878774

Guess, A., Nyhan, B., & Reifler, J. (2018). Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 us presidential campaign. Retrieved from https://www.dartmouth.edu/~nyhan/fake-news-2016.pdf

Kuklinski, J. H., Quirk, P. J., Jerit, J., Schwieder, D., & Rich, R. F. (2000). Misinformation and the currency of democratic citizenship. *The Journal of Politics*, *62*(3), 790–816.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480–498.

Lazer, D., Baum, M., Grinberg, N., Friedland, L., Joseph, K., Hobbs, W., & Mattsson, C. (2017). Combating fake news: an agenda for research and action. *Harvard Kennedy School, Shorenstein Center on Media, Politics and Public Policy*, *2*.

Lee, J. & Song, H. (2017). Why people post news on social networking sites: A focus on technology adoption, media bias, and partisanship strength. *Electronic News*, *11*, 59–79.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*, 106–131. doi:10.1177/1529100612451018

Lewandowsky, S., Ecker, U. K., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "Post-Truth" era. *Journal of Applied Research in Memory and Cognition*, *6*, 353–369. doi:10.1016/j.jarmac.2017.07.008

Margolin, D. B., Hannak, A., & Weber, I. (2017). Political fact-checking on twitter: When do corrections have an effect? *Political Communication*, 1–24.

Nisbet, E. C., Cooper, K. E., & Garrett, R. K. (2015). The partisan brain: how dissonant science messages lead conservatives and liberals to (dis) trust science. *The ANNALS of the American Academy of Political and Social Science*, *658*(1), 36–66.

Nyhan, B. & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*, 303–330.

Shearer, E. & Gottfried, J. (2017). News use across social media platforms 2017. *Pew Res Cent*.

Strömbäck, J. (2005). In search of a standard: Four models of democracy and their normative implications for journalism. *Journalism Studies*, *6*, 331–345.

Taber, C. S. & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*, 755–769.

The New York Times. (2017, November). Signs of russian meddling in brexit referendum. Retrieved from https://www.nytimes.com/2017/11/15/world/europe/russia-brexit-twitter-facebook.html?smid=tw-share

Thorson, E. (2016). Belief echoes: The persistent effects of corrected misinformation. *Political Communication*, *33*, 460–480. doi:10.1080/10584609.2015.1102187

Weeks, B. E. (2015). Emotions, partisanship, and misperceptions: How anger and anxiety moderate the effect of partisan bias on susceptibility to political misinformation. *Journal of Communication*, *65*, 699–719.

Weeks, B. E. & Garrett, R. K. (2014). Electoral consequences of political rumors: Motivated reasoning, candidate rumors, and vote choice during the 2008 us presidential election. *International Journal of Public Opinion Research*, *26*, 401–422.

Weeks, B. E. & Holbert, R. L. (2013). Predicting dissemination of news content in social media: A focus on reception, friending, and partisanship. *Journalism & Mass Communication Quarterly*, *90*, 212–232.

Wood, T. & Porter, E. (2018). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*, 1–29. doi:10.1007/s11109-018-9443-y

Footnotes

[1] Often, the term *fake news* is used as derogatory, rhetorical label to attack political opponents. While such use of the term as a *label* is an important conceptual dimension to consider, this aspects of *fake news* is beyond the scope of this manuscript. See Egelhofer and Lecheler (2017) instead for a detailed conceptualization involving this distinction.