5-2-2023

USING TRANSACTION COST ECONOMICS SAFEGUARDING TO REDUCE THE DIFFUSION OF DISINFORMATION ON SOCIAL MEDIA

David Eccles
The University of Melbourne, ecclesd@student.unimelb.edu.au

Sherah Kurnia
The University of Melbourne, sherahk@unimelb.edu.au

Nicholas Geard
The University of Melbourne, nicholas.geard@unimelb.edu.au

Tilman Dingler
The University of Melbourne, tilman.dingler@unimelb.edu.au

Follow this and additional works at: https://aisel.aisnet.org/ecis2023_rip

USING TRANSACTION COST ECONOMICS SAFEGUARDING TO REDUCE THE DIFFUSION OF DISINFORMATION ON SOCIAL MEDIA

Research in Progress

Eccles, David, A., The University of Melbourne, Parkville, Australia, ecclesd@student.unimelb.edu.au

Kurnia, Sherah, The University of Melbourne, Parkville, Australia, sherahk@unimelb.edu.au

Geard, Nicholas, The University of Melbourne, Parkville, Australia, nicholas.geard@unimelb.edu.au

Dingler, Tilman, The University of Melbourne, Parkville, Australia, tilman.dingler@unimelb.edu.au

Abstract

Human users contribute to the spread of disinformation on Social Media. To reduce the spread, we apply Transaction Cost Economics (TCE) Safeguarding, which penalises the sharing of disinformation. Using the economic theory TCE positions Social Media platforms as free markets, in which actors are motivated to protect their assets and peer reputation. We conducted a study exploring TCE Safeguarding as a market correction mechanism to change the disinformation diffusion behaviour of users. Our findings show that users will be less likely to post a comment and more likely to correct their previous disinformation diffusion actions when TCE Safeguarding is applied. Focusing on Social Media as a market rather than its individual components may provide a mechanism to address the "fake news" phenomenon.

Keywords: Transaction Cost Economics, disinformation, Fake news, economic models, Social Media.

1 Introduction

False and harmful messages (disinformation) on Social Media have repeatedly been shown to negatively affect individuals, groups, organizations, and society (Schuetz, Sykes, and Venkatesh, 2021). In 2016, two events, the Brexit referendum (Cadwalladr, 2017) and the outcome of the U.S. Presidential election (Allcott and Gentzkow, 2017), brought attention to the real-world impact of disinformation on Social Media, and its negative influence on democratic processes and institutions (Howard, 2018).

Most of the theoretical and applied responses to Social Media's disinformation phenomenon have focused on post hoc (after exposure) remediation such as fact-checking content (Vinhas and Bastos, 2022), tagging of potential disinformation (Figl et al., 2019), and understanding Social Media user behaviour (Moravec, Minas, and Dennis, 2019), including cognitive biases (Moravec, Kim, and Dennis, 2020) and cognitive weaknesses (Horner et al., 2021). Much of the extant research focuses on the separate components of a Social Media platform such as users (Nelson and Taneja, 2018), content (George, Gerhart, and Torres, 2021), algorithms and bots (Delgado et al., 2021), disinformation producers & creators (Kragh and Äsberg, 2017), points of origination (Chadwick, Vaccari, and O'Loughlin, 2018), and social network contagion (Vosoughi, Roy, and Aral, 2018), with limited focus on the holistic operation of Social Media as a market. Social Media promotes itself as the free market of ideas, where ideas and opinions are exchanged with minimal oversight. This language evokes the idea of mercantile free markets where buyers and sellers come to exchange goods and services. When mercantile markets fail or behave abnormally market participants or governments intervene (e.g., EU's Digital Services Act). In Social Media, the free market

of ideas has seen failure or abnormality in the form of the disinformation phenomenon (NATO, 2017). Despite attempts from governments (House of Commons Digital Media Sport Committee, 2019), Social Media platforms (Ross et al., 2018; S. Wang, Pang, and Pavlou, 2021) and organizations (e.g. BBC Reality Check), disinformation's negative consequences disrupt the desired operation of Social Media.

In this paper, we investigate if the free market economic theory of Transaction Cost Economics (Williamson, 1987), particularly the Network Governance Safeguarding mechanism (Jones, Hesterly, and Borgatti, 1997), could protect Social Media from disinformation and its diffusion. The research question addressed is: *How does TCE Safeguarding affect Social Media user's willingness to spread disinformation?*

Our experiment tested the effect of reputation ranking based on user's previous behaviours including user's prior diffusion of disinformation. In this online study with 35 participants, we discovered that the introduction of TCE Safeguarding in the form of reputation ranking did impact Social Media user behaviour. Participants treated with TCE Safeguarding were less likely to post positive comments about disinformation posts, and would undo prior actions which previously diffused disinformation. Treating Social Media as an economic market provides new approaches to addressing the disinformation diffusion problem.

2 Background

In this section we provide an overview of Social Media, online social networks, the "fake news" phenomenon, the research to date on empowering users, and, an overview of TCE Safeguarding. We identify humans as users with agency on Social Media, and actors in an economic market - a novel view of disinformation diffusion on Social Media.

2.1 Social Media as an Online Social Network

Social Media is a specialised form of online social network (OSN) (Kane et al., 2014). Researchers use social network analysis (SNA) to study online social networks (Borgatti and Foster, 2003). In SNA a network is a set of actors (nodes) connected by a set of ties. Ties connect pairs of nodes and can be directed, undirected, dichotomous, or valued. Nodes can be individuals, groups, or concepts. Researchers focus on a 2x2 matrix typology (Table 1)(Borgatti and Foster, 2003). Our interest is in the diffusion axis through *contagion* and *environmental shaping*.

	Social Capital	**Diffusion (Social Homogenity)**
Structure	Structural Capital	Environmental Shaping
Content	Resource Access	Contagion

Table 1.   Typology of SNA Studies (Borgatti & Foster, 2003)

On Social Media, users are individual nodes in an online network. The ties they connect to other nodes are largely defined and constrained by the Social Media platform. Users are impacted by the structure of the network, network position, and shared content. Both factual and false information diffuses through the network as a contagion process (Vosoughi, Roy, and Aral, 2018). The network structure results in strengthening social homogenity (environmental shaping) because nodes are tightly coupled with many ties (Kitchens, Johnson, and Gray, 2020). Environmental shaping of nodes has been labelled as "echo chambers" and "filter bubbles", but these labels are inaccurate descriptors for environmental shaping (see Bruns, 2021, Shore, Baek, and Dellarocas, 2018, and Törnberg, 2018 for a nuanced overview of the labelling problem). Human users, whether active or passive, play a role in the diffusion of disinformation. Some users because of their network position within Social Media (so-called "influencers") have greater social capital than others (Kim and Dennis, 2019). However, all users have agency as a member of Social Media.

2.2   Fake News and Disinformation

The broader community uses the phrase "fake news" to refer to the phenomenon of false and harmful information created, and diffused on Social Media platforms like Instagram, Facebook, and Twitter. Researchers have various definitions of "fake news" although two highly cited articles suggest news in the journalistic sense of news media and news form. Allcott and Gentzkow (2017) define "fake news" as "news articles that are intentionally and verifiably false, and could mislead readers" (Allcott and Gentzkow, 2017 p.213) whereas Lazer et al. (2018) extends the definition to "fabricated information that mimics news media content in form but not in organisational process or intent" (Lazer et al., 2018 pp.1094). We argue that specific focus on news media omits other forms of false and/or harmful information such as rumor (Simon et al., 2016), lying, deception, and knowledge hiding (Khan, Brohman, and Addas, 2021) and fails to distinguish from comedy, parody, and satire (Garrett and Poulsen, 2019). Disinformation created and diffused on Social Media has negative effects on individuals, groups, organizations, and nations (Lazer et al., 2018), and have demonstrated successes in subverting and disrupting democracies (Howard, 2018). Disinformation on Social Media contributes to polarisation (Schmidt et al., 2018), creating attitudes that will not actively listen to opposing attitudes on a topic, and seek to remove opposing topic attitudes (Lewandowsky, Ecker, and Cook, 2017). Social Media platforms attract, engage, and keep users by learning what they like and don't through extensive data collection (Matsakis, 2018), and provide Social Media feeds that appeal to user's biases, values, and existing topic attitudes (Kitchens, Johnson, and Gray, 2020). Algorithms promote user accounts to follow that have a high degree of homogeneity on thousands of issues of interest to users resulting in "in-group" attitude reinforcement, and lower tolerance for opposing attitudes and beliefs (Bruns, 2021). For these reasons we will refer to the "fake news" phenomenon as disinformation.

2.3   Social Media User Agency

Human users continue to create and diffuse disinformation on Social Media (Allcott and Gentzkow, 2017). Research has called for empowering individuals to improve individual evaluation of posts (Lazer et al., 2018). However, much research focus is on user's cognitive biases (Moravec, Kim, Dennis, and Minas, 2022) and cognitive weaknesses (Moravec, Minas, and Dennis, 2019), which sees users as victims with limited accountability and responsibility. A number of post-exposure tools such as fact checking (Figl et al., 2019) and AI identification of message veracity (Delgado et al., 2021) focus on content rather than user empowerment. Correcting attitude change caused by disinformation is much harder than preventing attitude change (Lewandowsky, Cook, et al., 2020). Social Media users have agency and a responsibility to comply with Social Media Platforms' terms of service. Currently users contribute consciously and sometimes inadvertently to disinformation diffusion with minimal accountability and responsibility. For example, users spread false information on a topic further, deeper and longer than factual information on the same topic (Vosoughi, Roy, and Aral, 2018), and form online vigilante groups with negative consequences (Simon et al., 2016) including the 'Pizzagate' conspiracy theory (Hsu, 2017). All users have agency in what they create, share, like, and comment on (Patel and Constantiou, 2020) with limited mechanisms that reward good, and penalise poor behaviour. Attempts by Sine Weibo including "social score" and forwarding restrictions (K. C. Ng, Tang, and Lee, 2021) rely on users and their biases and weaknesses reporting other user's incidents of harmful information. The same problem of user vigilante and mob rule responses remain. Our research addresses two gaps in extant research: Social Media user accountability and responsibility beyond the disinformation producer, and prevention strategies that support user empowerment.

2.4 Transaction Cost Economics Safeguarding

Transaction Cost Economics explains when a firm should take its asset to the market; enters into a hybrid relationship (market and firm); or keeps asset production within the firm (Williamson, 1987). Originally perceived as a bilateral exchange between the a vendor from the market and the firm, the theory has been expanded to include Network Governance where a group of vendors from the market work with the firm in the production of the asset (Jones, Hesterly, and Borgatti, 1997). Network governance was first observed in the movie industry where various vendors (catering, set construction, editors, production design, directors) work together for the production of a movie (the asset). Today in a connected world, Network Governance is commonplace in supply chain asset production where a client takes asset production to the market and works with a number of vendors and participates as both client and vendor (Wever et al., 2012).

Sometimes, assets in the free market need protection from opportunistic behaviour exploiting an advantage not open to all. This protection is known as Safeguarding, and is used to restore the market to normal operation (Williamson, 1987). In TCE Network Governance, if one vendor under performs or violates the informal contract, social mechanisms can be employed to protect the asset. The types of social mechanism safeguards include: temporary restriction of access to the market; collective sanctions or boycotts of the offending vendor; and, exclusion from current and future asset production based on past performance (Jones, Hesterly, and Borgatti, 1997). Reputation as a social mechanism safeguard in TCE Network Governance depends on the user's own past behaviour. The user's interactions with other nodes in the network, their collaboration, good citizenship, and cooperation is likely to establish a good reputation and a high chance of participating in future networked market exchanges (Jones, Hesterly, and Borgatti, 1997). If users are found to be uncooperative, poor citizens, and hostile, then their own past actions and behaviour limit opportunities in the current and future marketplace. We decide to test reputation ranking Safeguarding to see if it impacts user behaviour surrounding disinformation and its diffusion on Social Media. TCE is a good fit for Social Media. TCE theory explicitly assumes irrational and opportunistic behaviour of human actors and assumes that humans have bounded rationality—that is, not all decisions and actions are rational—and they will behave opportunistically when possible to do so (Richter, Riemer, and Vom Brocke, 2010). TCE's two core assumptions (bounded rationality and opportunistic behaviour) and the three dimensions (asset specificity, uncertainty and transaction frequency) have previously been identified in Social Media in the context of corporate fan support (C. S.-P. Ng and E. T. Wang, 2019). The novelty of our approach to the Social Media disinformation problem sees Social Media users as participants in an economic market. Guided by TCE Safeguarding, we propose the following hypotheses:

H1a TCE Safeguarding reduces users sharing posts containing disinformation on Social Media.

H1b TCE Safeguarding reduces users 'Liking' posts containing disinformation on Social Media.

H1c TCE Safeguarding reduces users commenting positively on posts containing disinformation on Social Media.

We also want to know if users would undo their actions if they knew that their past behaviour had negatively impacted their current reputation.

H2: TCE Safeguarding encourage users to undo actions that shared Social Media posts containing disinformation.

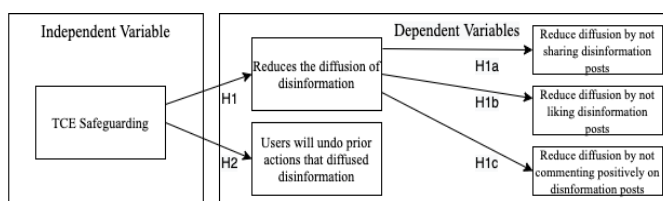We summarise our research model shown in Figure 1.

Figure 1. Research Model.

3 Method

To investigate the effect of TCE on user behaviour, we conducted an online experiment in which we manipulated Social Media posts with reputation rankings and assessed participants' commenting and sharing behaviour.

Study Materials. We chose six topics of moderate to strong topic interest to our participants: Indigenous affairs; Refugee Immigration; Climate Change; Housing Affordability; Welfare Payments; and Decriminalising personal Drug Use. We researched factual and false information on each of these topics using government statutory websites, peer-reviewed information, and reputable fact-checking organizations such as RMIT/ABC fact check; Reuters fact check; BBC Reality check; and CNN fact check. For each topic, we manufactured one factual and three false posts containing disinformation. All posts were between 18-22 words in length to ensure the same cognitive effort was involved in reading the post (Figure 2).
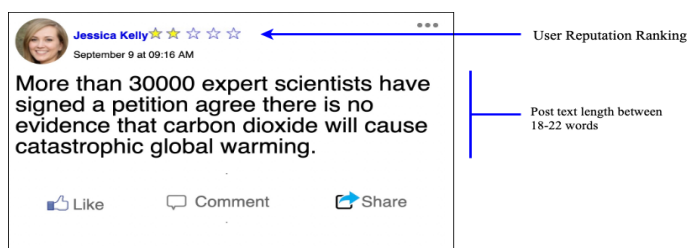

Figure 2.   Social Media post example used in our study.

Using https://thispersondoesnotexist.com we generated four Social Media avatars: two female and two male. We modified our Facebook-like post to contain user reputation rankings based on a 0-10 point scale: 0-5 stars in half-star increments (Figure 2). We gave one female and one male a good reputation ranking (4 stars or higher) and one female and one male a poor reputation ranking (2 stars or less) to neutralise the effect of gender on participants' perception. We then created a scripted recorded presentation to explain the TCE Safeguarding using reputation ranking treatment. We used the movie industry example used in Jones, Hesterly and Borgatti (Jones, Hesterly, and Borgatti, 1997) to explain Network Governance TCE Safeguarding. This example establishes reputation ranking based on past performance. The presentation then explains how the reputation ranking will be applied to users on Social Media where prior diffusion actions (share, like, comment) are evaluated (Figure 3) to derive a current reputation ranking.

Figure 3.   Explaining TCE Safeguarding Reputation

Explicitly if that user had contributed to the creation, production, or diffusion of false, harmful, or false and harmful information, their current reputation ranking would be lower; if they had not participated in any harmful or false information, their reputation ranking would remain as it is or improve (Figure 4). The control presentation was a four-minute promotional video about our University.
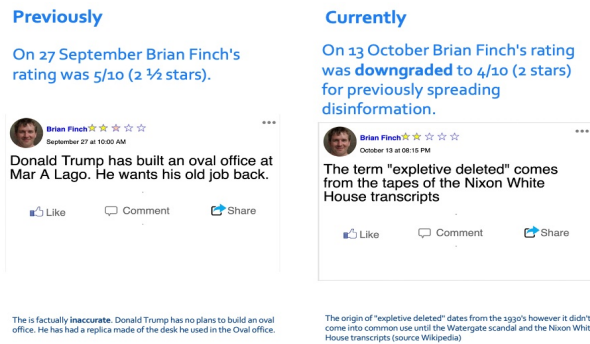
Previously

On 27 September Brian Finch's rating was 5/10 (2 ½ stars).

Brian Finch ⭐ ☆ ☆ ☆ ☆
September 27 at 10:00 AM

Donald Trump has built an oval office at Mar A Lago. He wants his old job back.

👍 Like   💬 Comment   ↪ Share

This is factually **inaccurate**. Donald Trump has no plans to build an oval office. He has had a replica made of the desk he used in the Oval office.

Currently

On 13 October Brian Finch's rating was **downgraded** to 4/10 (2 stars) for previously spreading disinformation.

Brian Finch ⭐ ☆ ☆ ☆ ☆
October 13 at 08:15 PM

The term "expletive deleted" comes from the tapes of the Nixon White House transcripts

👍 Like   💬 Comment   ↪ Share

The origin of "expletive deleted" dates from the 1930's however it didn't come into common use until the Watergate scandal and the Nixon White House transcripts (source Wikipedia)

Figure 4.   Explaining downgrade in reputation ranking. Original (left) recalculated downgraded score (right).

Study Design. Registered participants completed a pre-study survey measuring their interest in a range of topics. We built the survey using the Qualtrics platform, hosted by the University of Melbourne. Our pre-treatment posts were chosen from two topics for a total of eight posts (two factual, six disinformation). We randomly divided participants into treatment and control groups. After the treatment/control the remaining 4 topics (16 posts: 4 factual and 12 disinformation) were viewed. Both pre- and post-treatment posts were randomised to eliminate order effect. Each post was in its own Qualtrics block which included a timing block and asked four questions: I) How interested are you in this topic?; II) How likely are you to share this Social Media post?; III) How likely are you to like this Social Media post?; and IV) How likely are you to post a favourable comment? Participants could respond on a 7-point scale (1 Extremely Unlikely - 7 Extremely Likely). Lastly, they were asked if they would undo their actions if they were informed that they had previously shared false or misleading information and if their undo action was determined by topic interest. All factual posts served as pre-/post-control, as our interest is disinformation diffusion. After completing the experiment, participants viewed the debriefing pack which provided information on all posts to confirm facts and debunk disinformation. We had a distress protocol which included counselling and support information, including the contact details of the lead researchers. Participants who completed the study went into a draw to win one of five gift vouchers for $100.

Participants. We invited 118 participants through University of Melbourne's electronic newsletters, and recruitment posters placed on community noticeboards at supermarkets near the University campus. Participants had to be over 18 years of age and actively use Social Media. Of the 118 participants invited to participate in the study, 43 started and 35 completed the survey. The majority of participants were women (60%), and one participant identified as non-binary. 11% of participants were below 25 years of age, 83% were between 25 and 64, and 6% were over the age of 65. Of the participants who completed the survey, 31% held a bachelor's degree and 54% held a master's or doctorate. 20% of participants used Social Media hourly, and 74% used Social Media daily. When asked about what purposes they used Social Media, 74% nominated entertainment, 74% social contact, 69% information seeking and 23% excitement seeking, 4% nominated other reasons such as self-promotion, work, 'laughs', and boredom.

4   Results

In this section, we report our results with regard to different hypotheses derived from the research question. After removing incomplete study responses we conducted an ANOVA of results using the R statistical package. Hypothesis 1 tested if TCE Safeguarding had any effect on diffusion behaviour. Hypothesis 1A tested if TCE Safeguarding reduced the likelihood of a user sharing a post containing disinformation. There was a minor change in diffusion behaviour (-0.15 in likelihood to share) (Table 2) but not statistically significant. H1A was not supported (Figure 5a).
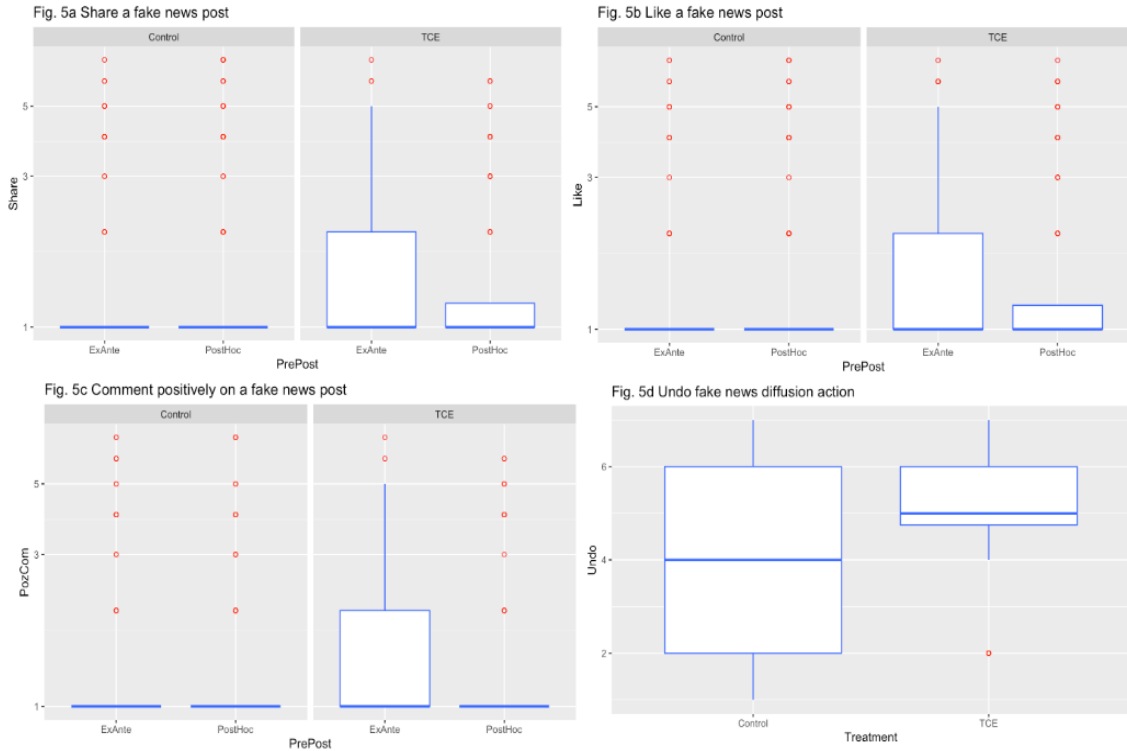
Figure 5.   Summary of results. Top (left to right) Likelihood to Share (Fig. 5a), Like (Fig. 5b), Bottom (left to right) Comment Positively (Fig. 5c) and Undo previous disinformation (Fig. 5d)

Hypothesis 1B tested if TCE Safeguarding would reduce the likelihood of liking a post containing disinformation. There was a minor change in diffusion behaviour (-0.11 in likelihood to 'Like') (Table 2), but it was not statistically significant. H1B was not supported (Figure 5b). Hypothesis 1C tested if TCE Safeguarding reduced the likelihood users would comment favourably on a post containing disinformation. Again, there was a minor yet statistically insignificant change in behaviour (-0.33 in likelihood to share, p = 0.1033 using Wilcoxon rank sum test) (Table 2). Hypothesis 1C was not supported (Figure 5c).

	Control Pre	**Control Post**	**TCE Pre**	**TCE Post**
Share	M=1.66 (SD=1.39)	M=1.62 (SD=1.44)	M=1.82 (SD=1.43)	M=1.67 (SD=1.30)
Like	M=1.70 (SD=1.46)	M=1.61 (SD=1.41)	M=1.92 (SD=1.59)	M=1.81 (SD=1.54)
Pos Comment	M=1.69 (SD=1.45)	M=1.52 (SD=1.30)	M=1.76 (SD=1.41)	M=1.43 (SD=1.12)

Table 2.   Median change in participant diffusion behaviour (Likert scale 1 Unlikely - 7 LIkely)

H2 - the second hypothesis tested if users would undo a diffusion action (share, like, comment) on a post that was identified as disinformation and would negatively impact their reputation. The TCE safeguard group of participants would undo their diffusion behaviour to a greater degree than those in the control group. The TCE median was "slightly likely" mean = 5.06 (SD = 1.44) , the Control mean = 4 (SD = 2.08) with p<0.001 using the Wilcoxon rank sum test. Hypothesis 2 was supported (Figure 5d).

5   Discussion

This study used TCE Safeguarding, an economic theory, to test whether we could alter users' disinformation diffusion behaviour. We found that the use of TCE Safeguarding, using reputation ranking based

on past performance, affected some diffusion behaviour but not all. Our research question was partially answered in the positive and partially answered in the negative: Participants who received the TCE Safeguarding treatment were less likely to comment positively on disinformation posts. However, other diffusion actions, including sharing (network contagion), did not change. TCE Safeguarding participants would undo their previous actions that diffused disinformation to protect their reputation ranking. Prior research has encountered reluctance to engage with posts (Figl et al., 2019), and a strong affinity with reputation and status affecting diffusion (Havakhor, Soror, and Sabherwal, 2018) demonstrating principles of social validation and consistency (Cialdini, 2001). We believe our participants demonstrate "lurking" behaviour (Hong, Hu, and Zhao, 2023) preferring to observe but not engage despite moderate to strong topic interest. Early research into "lurking" behaviour suggests users will only engage with Social Media accounts they are familiar with and trust. This further supports the idea of Social Media as a market, where market participants are correcting behaviours to market malfunctions. Our study contributes to research into both disinformation diffusion on Social Media and Social Media user behaviour by demonstrating that economic models like TCE and TCE's market correction mechanism—Safeguarding—can be used to address the "fake news" phenomenon. Our study's novelty is to take a holistic view of Social Media as an economic model and not focus on individual components: users, content, origination, diffusion, and disinformation identification. Thus, we position Social Media's disinformation problem as an economic market malfunction. The contribution to practice is that user accountability and responsibility mechanisms in the design of Social Media can be done in ways that do not restrict freedom of expression nor rely on community reporting. The introduction of accountability measures for behaviours such as disinformation should be explored to address the self acknowledged problem with Social Media. Our study provides first insights into the effects of TCE Safeguarding using reputation star ratings on sharing behaviour. Future work will look into algorithms to calculate and update users' reputations, including different weights according to factors, such as recency and spread of individual posts. Despite a limited sample size, we saw effects in changing user behaviour in commenting, and users electing to undo previous actions that could affect current and future reputations. A larger sample size and a field trial of an actual Social Media service should be conducted to further assess the generalizability of our findings.

5.1 Conclusion

This study addresses the disinformation problem by focusing on Social Media platforms as an economic market and not by focusing on the individual components of that market. There are a number of limitations to this study. Despite the findings, the sample size of participants (N=35) limits the gerneralizability and a repetition of the study with a larger sample size (n=350 - 500 participants) could confirm the findings. The introduction of post survey qualitative approaches such as a structured interview may shed knowledge on "lurking" motivations. Second, the proposed algorithmic model to derive the reputation ranking can be gamed, and users are wary of algorithms (Kießling, Figl, and Remus, 2021). Third, marionette, and bot disinformation accounts continually evolve and respond and could achieve a reputation ranking. Future research could also investigate other Safeguarding mechanisms (temporary suspension, collective sanctions) within Transaction Cost Economics and other market economic theories, such as Incentive Alignment and Agency Theory as models to address disinformation on Social Media. Seeing the problem as a market malfunction with bad faith actors allows us to apply TCE—an economic market theory—to correct the market malfunction and protect the asset—factual information on Social Media—from bad faith actors.

References

Allcott, H. and M. Gentzkow (2017). "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31 (2), 211–236. ISSN: 0895-3309. DOI: 10.1257/jep.31.2.211.

Borgatti, S. and P. Foster (2003). "The Network Paradigm in Organizational Research: A Review and Typology." *Journal of Management* 29 (6), 991–1013. ISSN: 0149-2063. DOI: 10.1016/S0149-2063(03)00087-4.

Bruns, A. (2021). *Echo chambers? Filter bubbles? The misleading metaphors that obscure the real problem*. Hate Speech and Polarization in Participatory Society. Taylor and Francis, pp. 33–48. ISBN: 9781003109891. DOI: 10.4324/9781003109891-4.

Cadwalladr, C. (May 2017). *The great British Brexit robbery: how our democracy was hijacked*. Newspaper Article. URL: https://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy.

Chadwick, A., C. Vaccari, and B. O'Loughlin (2018). "Do tabloids poison the well of social media? Explaining democratically dysfunctional news sharing." *New Media & Society* 20 (11), 4255–4274. ISSN: 1461-4448. DOI: 10.1177/1461444818769689.

Cialdini, R. B. (2001). "The Science of Persuasion." *Scientific American* 284, 76–81.

Delgado, A. A. C., W. B. Glisson, N. Shashidhar, J. T. McDonald, R. Benton, and G. Grispos (2021). "Detecting deception using machine learning." In: *Proceedings of the Annual Hawaii International Conference on System Sciences*. Vol. 2020-January. IEEE Computer Society, pp. 7122–7131. ISBN: 9780998133140.

Figl, K., S. Kießling, C. Rank, and S. Vakulenko (2019). "Fake news flags, cognitive dissonance, and the believability of social media posts." In: *International Conference on Information Systems*. Vol. 27. Munich, Germany: Association for Information Systems.

Garrett, R. K. and S. Poulsen (2019). "Flagging Facebook Falsehoods: Self-Identified Humor Warnings Outperform Fact Checker and Peer Warnings." *Journal of Computer-Mediated Communication* 24 (5), 240–258. ISSN: 1083-6101. DOI: 10.1093/ccc/zmz012. URL: %3CGo%20to%20ISI%3E://WOS:000493101200003.

George, J., N. Gerhart, and R. Torres (2021). "Uncovering the Truth about Fake News: A Research Model Grounded in Multi-Disciplinary Literature." *Journal of Management Information Systems* 38 (4), 1067–1094. ISSN: 07421222 1557928X.

Havakhor, T., A. A. Soror, and R. Sabherwal (2018). "Diffusion of knowledge in social media networks: effects of reputation mechanisms and distribution of knowledge roles." *Information Systems Journal* 28 (1), 104–141. ISSN: 1350-1917. DOI: 10.1111/isj.12127.

Hong, Y., J. Hu, and Y. Zhao (2023). "Would you go invisible on social media? An empirical study on the antecedents of users' lurking behavior." *Technological Forecasting & Social Change* 187. ISSN: 0040-1625. DOI: 10.1016/j.techfore.2022.122237.

Horner, C. G., D. Galletta, J. Crawford, and A. Shirsat (2021). "Emotions: The Unexplored Fuel of Fake News on Social Media." *Journal of Management Information Systems* 38, 1039–1066. ISSN: 07421222.

House of Commons Digital Media Sport Committee, U. (2019). *Disinformation and 'fake news': Final Report*. Government Document. URL: https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf.

Howard, P. (2018). "Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration." *Journal of Information Technology & Politics* 15 (2), 81. ISSN: 1933-1681. URL: https://www.oii.ox.ac.uk/blog/algorithms-bots-and-political-communication-in-the-us-2016-election.html.

Hsu, S. (2017). *'Pizzagate' gunman sentenced to four years in prison, as prosecutors urged judge to deter vigilante justice*. Newspaper Article. URL: https://www.washingtonpost.com/local/public-safety/pizzagate-gunman-sentenced-to-four-years-in-prison-as-prosecutors-

urged-judge-to-deter-vigilante-justice/2017/06/22/a10db598-550b-11e7-ba90-f5875b7d1876_story.html.

Jones, C., W. Hesterly, and S. Borgatti (1997). "A general theory of network governance: exchange conditions and social mechanisms." *Academy of Management Review* 22 (4), 911–945. ISSN: 03637425. DOI: 10.5465/AMR.1997.9711022109.

Kane, G., M. Alavi, G. Labianca, and S. Borgatti (2014). "What's different about social media networks? A framework and research agenda." *MIS Quarterly* 38 (1), 275–304. ISSN: 02767783.

Khan, A., K. Brohman, and S. Addas (2021). "The anatomy of 'fake news': Studying false messages as digital objects." *Journal of Information Technology*. ISSN: 02683962 14664437.

Kießling, S., K. Figl, and U. Remus (2021). "Human Exports or Artificial Intelligence? Algorithm Aversion in Fake News Detection." *Proceedings of the European Conference on Information Systems (ECIS)*, 1–17.

Kim, A. and A. Dennis (2019). "Says Who? The Effects of Presentation Format and Source Rating on Fake News in Social Media." *MIS Quarterly* 43 (3), 1025–1039. ISSN: 02767783. DOI: 10.25300/MISQ/2019/15188.

Kitchens, B., S. L. Johnson, and P. Gray (2020). "Understanding Echo Chambers and Filter Bubbles: The Impact of Social Media on Diversification and Partisan Shifts in News Consumption." *MIS Quarterly* 44 (4), 1619–1649. ISSN: 02767783. DOI: 10.25300/MISQ/2020/16371.

Kragh, M. and S. Äsberg (2017). "Russia's strategy for influence through public diplomacy and active measures: the Swedish case." *Journal of Strategic Studies* 40 (6), 773–816. ISSN: 0140-2390. DOI: 10.1080/01402390.2016.1273830.

Lazer, D. M., Y. Benkler, A. Berinsky, K. Greenhill, F. Menczer, M. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, M. Schudson, S. Sloman, C. Sunstein, E. Thorson, D. Watts, and J. Zittrain (2018). "The science of fake news." *Science* 359 (6380), 5. ISSN: 1837-7009. URL: https://science.sciencemag.org/content/359/6380/1094.

Lewandowsky, S., J. Cook, U. Ecker, D. Albarraćin, M. Amazeen, P. Kendeou, D. Lombardi, E. Newmand, G. Pennycook, D. Rand, D. Rapp, J. Reifler, J. Roozenbeek, P. Schmid, C. Seifert, G. Sinartra, B. Swire-Thompson, S. van der Linden, E. Vraga, T. Wood, and M. Zaragova (2020). *The Debunking Handbook 2020*. Electronic Book. DOI: DOI:10.17910/b7.1182. URL: https://sks.to/db2020.

Lewandowsky, S., U. Ecker, and J. Cook (2017). "Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era." *Journal of Applied Research in Memory and Cognition* 6 (4), 353–369. ISSN: 2211-3681. DOI: 10.1016/j.jarmac.2017.07.008.

Matsakis, L. (Apr. 2018). *Facebook's Targeted Ads Are More Complex Than It Lets On*. Web Page. URL: https://www.wired.com/story/facebooks-targeted-ads-are-more-complex-than-it-lets-on/.

Moravec, P., A. Kim, and A. Dennis (2020). "Appealing to Sense and Sensibility: System 1 and System 2 Interventions for Fake News on Social Media." *Information Systems Research* 31 (3), 987–1006. DOI: https://doi.org/10.1287/isre.2020.0927.

Moravec, P., A. Kim, A. Dennis, and R. Minas (2022). "Do You Really Know if It's True? How Asking Users to Rate Stories Affects Belief in Fake News on Social Media." *Information Systems Research* 33 (3), 887–907. ISSN: 10477047 15265536.

Moravec, P., R. Minas, and A. Dennis (2019). "Fake News on Social Media: People believe what they want to believe when it makes no sense at all." *MIS Quarterly* 43 (4), 1343–1360. ISSN: 02767783. DOI: 10.25300/MISQ/2019/15505.

NATO, S. C. (Nov. 2017). *Digital Hydra: security implications of false information online*. Report. Riga, Latvia: NATO Strategic Command Centre of Excellence.

Nelson, J. L. and H. Taneja (2018). "The small, disloyal fake news audience: The role of audience availability in fake news consumption." *New Media & Society* 20 (10), 3720–3737. ISSN: 1461-4448. DOI: 10.1177/1461444818758715.

Ng, C. S.-P. and E. T. Wang (2019). "Impact of social media management styles on willingness to be a fan: A transaction cost economics perspective." *Pacific Asia Journal of the Association for Information Systems* 11 (2), 2.

Ng, K. C., J. Tang, and D. Lee (2021). "The Effect of Platform Intervention Policies on Fake News Dissemination and Survival: An Empirical Examination." *Journal of Management Information Systems* 38 (4), 898–930. ISSN: 0742-1222.

Patel, S. and I. Constantiou (2020). "Human Agency in the Propagation of False Information - A Conceptual Framework." In: *European Conference on Information Systems*. Vol. 2020-Research In Progress. Association for Information Systems.

Richter, D., K. Riemer, and J. Vom Brocke (2010). "Social transactions on social network sites: Can transaction cost theory contribute to a better understanding of internet social networking?" In: *23rd Bled eConference eTrust: Implications for the Individual, Enterprises and Society - Proceedings*. Bled eCommerce Conference, pp. 33–49.

Ross, B., A. Jung, J. Heisel, and S. Stieglitz (Dec. 2018). "Fake News on Social Media: The (In)Effectiveness of Warning Messages." In: *International Conference on Information Systems*. San Francisco, USA: Association for Information Systems, 16.

Schmidt, A. L., F. Zollo, A. Scala, C. Betsch, and W. Quattrociocchi (2018). "Polarization of the vaccination debate on Facebook." *Vaccine* 36 (25), 3606–3612. ISSN: 0264-410X. DOI: 10.1016/j.vaccine.2018.05.040.

Schuetz, S. W., T. A. Sykes, and V. Venkatesh (2021). "Combating COVID-19 fake news on social media through fact checking: antecedents and consequences." *European Journal of Information Systems*. ISSN: 0960085X-14769344. DOI: 10.1080/0960085X.2021.1895682.

Shore, J., J. Baek, and C. Dellarocas (2018). "Twitter Is Not the Echo Chamber We Think It Is." *MIT Sloan Management Review* 60 (1), 1–5. ISSN: 15329194.

Simon, T., A. Goldberg, D. Leykin, and B. Adini (2016). "Kidnapping WhatsApp - Rumors during the search and rescue operation of three kidnapped youth." *Computers in Human Behavior* 64, 183–190. ISSN: 0747-5632. DOI: 10.1016/j.chb.2016.06.058. URL: %3CGo%20to%20ISI%3E://WOS:000384775200019.

Törnberg, P. (2018). "Echo chambers and viral misinformation: Modeling fake news as complex contagion." *PLoS ONE* 13 (9), 1–21. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0203958.

Vinhas, O. and M. Bastos (2022). "Fact-Checking Misinformation: Eight Notes on Consensus Reality." *Journalism Studies* 23 (4), 448–468. ISSN: 1461-670X. DOI: 10.1080/1461670X.2022.2031259. URL: https://doi.org/10.1080/1461670X.2022.2031259.

Vosoughi, S., D. Roy, and S. Aral (2018). "The spread of true and false news online." *Science* (6380), 1146. ISSN: 0036-8075. DOI: 10.1126/science.aap9559.

Wang, S., M.-S. Pang, and P. A. Pavlou (2021). "Cure or Poison? Identity Verification and the Posting of Fake News on Social Media." *Journal of Management Information Systems* 38 (4), 1011–1038. ISSN: 0742-1222. DOI: 10.1080/07421222.2021.1990615.

Wever, M., P. M. Wognum, J. H. Trienekens, and S. W. F. Omta (2012). "Supply chain-wide consequences of transaction risks and their contractual solutions: Towards an extended transaction cost economics framework." *Journal of Supply Chain Management* 48 (1), 73–91. ISSN: 15232409 1745493X. DOI: 10.1111/j.1745-493X.2011.03253.x.

Williamson, O. (1987). *The economic institutions of capitalism : firms, markets, relational contracting*. 1st Free Press pbk. ed. Free Press. ISBN: 0029348218.