Lee, Chang Boon; Io, Hio Nam; Tang, Heng

Article

Sentiments and perceptions after a privacy breach incident

Cogent Business & Management

Provided in Cooperation with:
Taylor & Francis Group

Suggested Citation: Lee, Chang Boon; Io, Hio Nam; Tang, Heng (2022) : Sentiments and perceptions after a privacy breach incident, Cogent Business & Management, ISSN 2331-1975, Taylor & Francis, Abingdon, Vol. 9, Iss. 1, pp. 1-12, https://doi.org/10.1080/23311975.2022.2050018

This Version is available at:
https://hdl.handle.net/10419/288625

Cogent Business & Management

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/oabm20

Sentiments and perceptions after a privacy breach incident

Chang Boon Lee, Hio Nam Io & Heng Tang

To cite this article: Chang Boon Lee, Hio Nam Io & Heng Tang (2022) Sentiments and perceptions after a privacy breach incident, Cogent Business & Management, 9:1, 2050018, DOI: 10.1080/23311975.2022.2050018

To link to this article: https://doi.org/10.1080/23311975.2022.2050018

Lee et al., *Cogent Business & Management* (2022), 9: 2050018
https://doi.org/10.1080/23311975.2022.2050018

cogent business & management

OPERATIONS, INFORMATION & TECHNOLOGY | RESEARCH ARTICLE

Sentiments and perceptions after a privacy breach incident

Chang Boon Lee[1]*, Hio Nam Io[2] and Heng Tang[1]

Received: 13 September 2019
Accepted: 13 February 2022

*Corresponding author: Chang Boon Lee, Faculty of Business Administration, University of Macau, Taipa, Macau, China
E-mail: cblee@um.edu.mo

Reviewing editor:
Trevor Wilmshurst, Accounting and Corporate Goverance, University of Tasmania - Launceston Campus, Tasmania, Australia

Additional information is available at the end of the article

Abstract: This study used sentiment analysis, keyword extraction, and content reviews to analyze related posts and comments from the *Weibo* microblogging website to determine the sentiments and perceptions regarding online privacy following the *Facebook-Cambridge Analytica* privacy breach incident. The results provide insights on users' sentiments and perceptions regarding online privacy in the Chinese context. Many comments have negative sentiments and the underlying theme of the comments is that there is no online privacy. Many users are scared of companies collecting data to support targeted advertisements because these companies know a lot about the users and the users are scared that their data could be used for inappropriate purposes. The comments also show skepticism that data collectors would protect users' data privacy. This study discusses the implications of the results for research and practice.

Subjects: Legal, Ethical & Social Aspects of IT; Management of IT; Information & Communication Technology; ICT

Keywords: Privacy breach; sentiment analysis; big data

ABOUT THE AUTHOR

Chang Boon Lee is an Associate Professor of Business Information Systems at the Faculty of Business Administration, University of Macau. Besides his current research on online privacy, Professor Lee is also interested in areas related to management of information technology professionals, teaching of spreadsheets, and innovation of new technologies.
Dr. Hio Nam Io is an Assistant Professor in the School of Business of the Macau University of Science and Technology. He received his PhD degree in Information Systems from the University of Macau. His research interests include adoption of innovations, social media mining, simulation, and entrepreneurship.
Dr. Heng Tang is an Assistant Professor at the department of Accounting and Information Management, University of Macau. His research interests are in the areas of consumer behavior in digital markets, intelligent systems, and business intelligence. His research has appeared in journals such as Decision Support Systems, IEEE Transactions on Systems, Man and Cybernetics, Information Systems Frontiers, Knowledge-based Systems, etc.

PUBLIC INTEREST STATEMENT

Sentiments and perceptions after a privacy breach incident. This study used computational tools to crawl related posts and comments at the Weibo microblogging website and analyzed the sentiments and perceptions about online privacy after the Facebook-Cambridge Analytica privacy breach incident. The results show that Weibo users feel there is no online privacy and that they are scared of privacy invasion. They associate the Facebook-Cambridge Analytica incident with experiences they have encountered on the Internet, such as receiving targeted advertisements or recommendations that are based on their supposedly private conversations. Their comments show skepticism that online data collectors would protect users' data privacy. The results provide insights on users' sentiments and perceptions regarding online privacy in the Chinese context, which is seldom discussed in the literature. The study discusses the implications of the results for research and practice.

1. Introduction

This study examined the sentiments and perceptions among users of a Chinese microblogging website after the *Facebook-Cambridge Analytica* privacy breach incident. The privacy breach incident involved a political consulting firm called *Cambridge Analytica* collecting personal data inappropriately from 87 million *Facebook* users and then using the harvested data to predict voting patterns to influence the 2016 American presidential elections. As a result of this incident, there is growing conern about how organizations handle personal data as it shows that security and privacy issues are often not given serious considerations by companies that collect data (Schneble et al., 2014). This incident has attracted worldwide discussion. Even in China where *Facebook* cannot be officially accessed, there are many comments about this incident on the Chinese social media. There are many social networking sites (SNS) that work like *Facebook* in China (DeGennaro, 2020). They collect vast amount of data, such as user profiles and tracking information, and they provide a platform for users to broadcast their comments and messages. The *Facebook-Cambridge Analytica* incident evokes concerns about data privacy and ethics of data sharing. This incident enables Chinese SNS users to relate it to something that might also happen to them as a social media user in China.

The *Facebook-Cambridge Analytica* incident exposed protected information and this caused people's privacy to be affected. Data privacy involves the right of an individual to have control over how personal information is collected and used (Gellman & Dixon, 2011). It is invaded when there is unauthorised collection, disclosure, or other usage of personal information (Debatin et al., 2009). In the United States and European countries, people receive strong legislative protection for data privacy. There are frequent news about companies being penalized for online data breaches —for example, Equifax (Gressin, , Target (McCoy, 2017) and Home Depot (Stempel, 2020). However, there are limited reports about data breaches or data privacy invasion in China. In China, there is no defined concept of privacy in its Constitution (Kennedy & Zhang, 2017; Wu et al., 2011). Citizens have to negotiate their own presence on privacy issues (Wang et al., 2020). To date, there is limited research about how Chinese citizens feel about online data privacy.

In addition, a number of scholars have indicated that studies on SNS have mostly addressed problems dealing with storage, retrieval, and management of social network data, and that privacy concerns stemming from the use of social networks, or the dissemination of social network data have largely been ignored (Loukides & Gkoulalas-Divanis, 2009). It is important to address issues related to SNS privacy as many people are now into social media (Ortiz-Ospina, 2019). There is prior research that shows there exist differences in online privacy concerns among people across nations. These differences could be attributed to factors such as cultural values, Internet experiences, and institutional settings (Bellman et al., 2004; Labrie et al., 2018). The limited studies about SNS privacy are conducted mostly in western settings. Few are based on non-western developing nations (Chen & Cheung, 2018; Standaland & Lwin, 2013; Wang et al., 2011). To fill this research gap, this study gathered and analyzed comments about the *Facebook-Cambridge Analytica* privacy breach incident from the Chinese SNS users to understand their sentiments and perceptions about online privacy. The results of this study are beneficial to researchers, online service providers, and policy makers as they provide insights about online privacy sentiments from the Chinese perspectives. This study thus helps to contribute to the call for greater research on data privacy (Isaak & Hanna, 2018).

The structure of this paper is as follow. The next section provides a review of literature related to the current study. This is followed by the research method. The results are presented next and then followed by a discussion of the findings. This paper ends by highlighting the contributions of the research as well as the limitations of the study.

2. Literature review

As mentioned in the introduction, there are few studies about data privacy in the Chinese context. One of the earliest studies in China about online privacy was conducted by Kong (2016). The study

found that Chinese companies collected vast amount of personal data at their websites and that only about 50 percent of them have privacy policies or statements. For those sites with privacy policies, the study found that they were not in accordance with the generally accepted principles of information practices. Later, Wu et al. (2015) did a comparative study of online privacy regulations in the U.S. and China and found that American legislative initiatives were more comprehensive and far-reaching than those of their Chinese counterparts. The study noted that self-regulation was the primary modality to protect users' online privacy in China. The privacy context in China is evolving because with effect from 1 November 2021, China has implemented a new Personal Information Protection Law (PIPL) to protect individuals, society, and national security from harm due to abuse and mishandling of personal information (Burgess, 2021). The law provides authorities the power to impose huge fines and blacklist companies that violate the law. With the implementation of the PIPL, online privacy issues are likely to draw greater attention among various stakeholders in China.

Regarding SNS, it should be noted that SNS users are particularly vulnerable to privacy invasion. The business model for SNS is to encourage users to communicate and interact with one another through their platform, and when people use SNS, they may divulge details about themselves or others. SNS therefore have vast collection of data. Sometimes, these data may be sold to third parties to generate revenue. Alternatively, SNS may analyze the data to learn about users and the information can be used to provide marketing support for targeted advertising. SNS users can adjust their privacy settings to protect their privacy, but the settings to restrict access to users' data change frequently, and users must continually reveal information about themselves in order to interact with others. Given the way that SNS operate, the context for online privacy has evolved over the years. Online privacy is no longer confined to the plaftorm's settings or the users' published contents (Youyou et al., 2015). Nowadays, the users' digital footprints are used by computer algorithms to predict human behavior and people may not realise that their privacy can be violated through subtle and indirect ways via their own or others' data (Hinds et al., 2020). In the case of the *Facebook-Cambridge Analytica* incident, the violation of data privacy involved *Cambridge Analytica* collecting *Facebook* users' data inappropriately, and then analyzing the harvested data to influence users' voting behavior in the American Presidential elections.

Prior research on data privacy has shown that while people are concerned about data privacy, their behavior seems to show there is little relationship between privacy concerns and behavior. This phenomenon is known as the privacy paradox . Researchers have provided a number of explanations for the paradox, such as the behavioral mechanisms underlying self-control and instant gratifications (Acquisiti & Gross, 2006). Another explanation for the privacy paradox is privacy fatigue (Choi et al., 2018). Privacy fatigue can cause people to feel exhausted from hearing endless news about data breaches and therefore they feel it is futile to do anything to protect their data. Thus, while people may be concerned about the *Facebook-Cambridge Analytica* privacy breach, they may not do anything about it because they feel it is challenging to change their privacy settings. The users may also have a sense of helplessness as they have no idea how to control the way their data is being used by the SNS providers (Shklovski et al., 2014). People may also be willing to exchange their privacy exposure for some benefit in return. In this case, the privacy calculus is used to analyze people's behavior based on the outcome of the privacy trade-off (Dinev & Hart, 2006). For example, some people are willing to disclose personal information so that they can enjoy the convenience of online shopping.

There are three studies that are closely related to the current research to determine the reactions to the *Facebook-Cambridge Analytica* privacy breach. The first was conducted by Fiesler and Hallinan (2018). The study examined public reactions to news about two data sharing controversies that had raised media attention. The first controversy was about *WhatsApp's* announcement on changes to its privacy policy when it specified that it would be sharing data with *Facebook*. The other involves unroll.me, a service provider that helps users manage their mail box. It was reported that Uber had purchased data from unroll.me for users' email receipts from

Lyft, which is Uber's major competitor. The study analyzed public comments related to news articles about the two data sharing controversies and found that even though majority of the articles had a negative framing towards data sharing, the attitudes expressed by the commenters were more nuanced than simiply being negative towards the platforms. The two controversies instigated discussions on a number of issues, one of which is related to who is responsible for privacy. Some comments indicate that users should bear responsibility for data protection, while others said the platform should. If users were responsible, it would create pressure for them to read the privacy policies, accept the terms and conditions of usage, and stay up-to-date with their privacy settings. On the other hand, there would also be problems if organisations were made responsible for data protection. For example, users may not trust the organizations, they may lack control over their data, and organizations may share the data with third parties.

The second study that is closely related to the current research examined *Facebook* users' privacy concerns following the *Facebook-Cambridge Analytica* incident (Hinds et al., 2020). The study used semi-structured interviews to gather information from 30 participants based at a UK university. The interviewees discussed their understanding of online privacy in the aftermath of the privacy violations. Results of the study showed that contrary to reports about actions taken by many *Facebook* users after the privacy scandal (Timms & Heimans, 2018), the participants did not delete their accounts or frantically change their privacy settings. Many participants did not express much concern about the scandal as they considered themselves immune to psychologically tailored advertisements. Some participants felt "threatened" when organisations might be monitoring them or using their data in unknown ways. Some also described targeted advertisements as "*freaky*" or "*creepy*", particularly when they appear after looking at something once, or immediately after browsing another website. The other findings in this study also showed that participants lacked understanding on how automated approaches and algorithms work in relation to their own and their networks' personal data. For example, the participants were unaware that personal information could be inferred through their interactions or via their friends'/followers' interactions.

The third study that is closely related to the current reserach focused on privacy perception and protection on Chinese social media (Chen & Cheung, 2018). This is a qualitative study and it collected data from 40 young Chinese adults about how they perceived and behaved towards privacy in their everyday experience of using *WeChat*, which is a popular social media platform in China. The study was conducted using semi-structured interviews and the results showed that while participants have concerns about their personal information and privacy on *WeChat*, there exists a privacy paradox because the participants' privacy concerns were not translated into privacy-protection actions. The study found that once users have ingrained their social engagement within the *WeChat* system, the incentive for them to remain a part of the system outweighs their requirement to secure their privacy online. The study noted that the widespread use of *WeChat* as an integral part of modern social communication in China has bestowed on the platform a unique and highly influential form of power: it has the power to control users' privacy and construct new privacy norms. Thus, the study suggested that Chinese *WeChat* users had a declined sense of their right to privacy.

3. Research method

To determine the sentiments and perceptions among Chinese SNS users about their online privacy, this study extracted relevant posts and comments from the *Weibo* website after news of the *Facebook-Cambridge Analytica* privacy breach incident was made public. *Weibo* is the most popular microblogging website in China (Wang, 2011). It provides a free-text format for users to publish their opinions on its site. An analysis of the relevant posts published on the website can shed light on people's sentiments and perceptions about the privacy breach. The research method used in this study is different from most of the prior studies that used questionnaire surveys to collect quantitative data to conduct research related to social media (Snelson, 2016). The comments from *Weibo* reflect issues that users are truly concerned about as *Weibo* provides users the opportunity to express their comments freely without priming from the researchers.

Figure 1. Research method.

As there are many comments to analyze, this study used computational software to perform the analysis. This study used sentiment analysis to analyze the emotions embedded in the comments. The purpose of sentiment analysis is to mine textual comments to detect favorable or unfavorable opinions. Sentiment analysis is an automated process and it can be conducted very quickly. It is unlike manual coding which takes a longer time to process. Sentiment analysis (Vinodhini & Chandrasekaran, 2012) has been used in research areas such as information systems, tourism, and others (Edwards et al., 2017; Lee et al., 2016). It is used to extract sentiments from massive amount of data. However, sentiment analysis only provides a numerical number about the sentiments. When sentiment analysis is supplemented with content reviews of frequently-used keywords in the comments, they can help to facilitate better understanding of the data. As such, besides sentiment analysis, this study extracted frequently-used keywords in the comments. The comments with the frequently-used keywords were then reviewed to gather insights related to the privacy breach incident.

Data for this study were collected for a month, starting from 22 March 2018—the day when reports about the data breach first appeared on the newspapers. This study crawled and extracted the most frequently commented posts using the words "Facebook隐私 (privacy)" on *Weibo*. Note that after a user creates a post on *Weibo*, other users can reply to the post. The replies under the original post are defined as "comments" in this study. A web crawling tool called GooSeeker (https://gooseeker.com) was used to crawl the posts. In order to distil the most important topics that online users were concerned about, the posts that had more than 200 comments were selected. This research used a threshold of 200 comments to represent the posts that users were interested in. Note that if the study were to include all posts regardless of the number of comments, the posts would include many topics that were very diversified and diffused. The study had also used the words "脸书隐私" ("Facebook privacy") to search for related data, but the number of posts that used "脸书隐私" gathered very few comments. It could be that people were more comfortable to use the term "Facebook" rather than its Chinese equivalent ("脸书").

There were 25 posts that have more than 200 comments. These 25 posts were grouped into related topics. The study used the same crawling tool, i.e., GooSeeker, to crawl the comments in the 25 posts. After cleaning the data, a total of 10,424 comments were available for this study. In this study, the researchers used a "Weibo-friendly" natural language processing tool called BosonNLP (https://bosonnlp.com) (Min et al., 2019) to conduct sentiment analysis. After performing sentiment analysis, a Python script that embeds a Chinese text segmentation tool called Jieba (Sun, 2015) was used to extract frequently-used keywords from related comments. The contents of the comments containing the keywords were then reviewed to gain further insights and to determine the underlying theme. figure 1 summarises the method used in this research.

4. Results

4.1. Statistical summary

The 25 popular posts were grouped into five related topics. They are: topic#1: Issue started, topic#2: Perceptions of data privacy, topic#3: Zuckerberg apologized, topic#4: Reactions to *XYZ's* comments on data privacy, and topic#5: *Facebook* testimony.

Table 1 presents the statistical results for the sentiment analysis based on the five topics. The sentiment score for each comment can range from 0 to 1. Scores that are close to 0 are low in sentiments while those that are close to 1 have high sentiments. As shown in Table 1, topic#2 (perceptions on data privacy) has the lowest average sentiment score while topic#4 (reactions to *XYZ's* comments on data privacy) has the highest average sentiment score. In addition to the average

cogent business & management

Table 1. Summary of sentiment scores for the five topics

Topics	No. of comments	Average Sentiment score	No. (%) of negative sentiment [0, 0.3)	No. (%) of neutral sentiment [0.3, 0.7)	No. (%) of positive sentiment [0.7,1]
#1: Issue started (5 posts)	1,225	0.44	505 (41%)	403 (33%)	317 (26%)
#2: Perceptions on data privacy (3 posts)	2,948	0.37	1,460 (49%)	937 (32%)	551 (19%)
#3: Zuckerberg apologized (5 posts)	1,047	0.51	347 (33%)	358 (34%)	342 (33%)
#4: Reactions to XYZ's comments on data privacy (4 posts)	2,313	0.92	728 (32%)	493 (21%)	1,092 (47%)
#5: *Facebook* testimony (8 posts)	2,891	0.52	875 (30%)	994 (35%)	1,022 (35%)

sentiment scores, Table 1 also shows the distribution of negative, neutral, and positive sentiments for each topic. As presented in Table 1, when the sentiment score is less than 0.3, the sentiment is classified as negative. When the score is between 0.3 and less than 0.7, the sentiment is neutral. When the score is equal or larger than 0.7, the sentiment is positive. This classification method is more stringent than the one where a score from 0 to 0.5 is classified as negative and others as positive (Tao et al., 2019). The more stringent classification method provides a higher standard for evaluating sentiments. The distribution of sentiments in Table 1 shows that topic#2 (perceptions on data privacy) has the highest percentage of negative sentiments while topic#4 (reactions to *XYZ's*[1] comments on data privacy) has the highest percentage of positive sentiments. These distributions are consistent with the average sentiment scores for the five topics.

Topic#2 and topic#4 comprise posts that discuss people's perceptions on data privacy. They are also the topics that have the lowest and highest average sentiments respectively. The study further analysed the comments related to topic#2 and topic#4 using keyword extraction and content reviews.

4.2. Comments on topic#2 (perceptions of data privacy)

All the three posts for topic#2 were published on 22 March 2018. The gists of the three posts said that there is no longer any data privacy in this era of big data. The contents of the three translated posts are shown below:

> The Facebook privacy issue may just be the tip of the iceberg. In the era of big data, what you search for, whom you give a 'like', what you want to buy … there is no privacy for everyone.

> Big data seems to be pervasive. Your privacy is no longer private. The Facebook scandal just sounded the alarm for us …

> For the Facebook data breach scandal in these two days, if the whole Internet big data is used together to construct our profile, each of us may be transparent.

Among the comments that follow from these posts, about 49 percent of the comments have negative sentiments (sentiment score < 0.3). The comments with negative sentiments were further analysed based on the most frequently-used keywords. The keywords associated with negative sentiments include: (1) ABC,[1] which is a popular Chinese e-commence website, (2) scary ("可怕"), and (3) recommendation ("推送").

Table 2 shows sample comments with keywords related to the negative sentiments. The main sentiment in these comments shows that *Weibo* users were scared. There are a number of reasons

cogent business & management

Table 2. Comments associated with the words "*ABC*", "Scary", and "Recommendation"

Comments	Sentiment score
Scary ("可怕")	
So it is real they can listen to conversations. I thought I am imaging things. This is so **scary**!	0.02
Next time, I wouldn't simply like or comment, it's very **scary**.	0.03
It's ok if they just guide me to purchase, but it's too **scary** if they can monitor my chatting conversations.	0.04
It is **scary** and I am helpless.	0.04
Big data monitoring is really **scary**.	0.07
So how to avoid this? Feels **scary**.	0.11
ABC and Recommendation ("推送")	
I really want to know, how come ***ABC*** would know whatever that was on my mind?	0.01
Now I know why ***ABC*** recommended me the pink lipstick. Next time must be cautious when chatting.	0.03
After I said something or said what I wanted in *DEF*, these things would then be **recommended** in ***ABC***.	0.11
ABC made **recommendations** to me many times, even though I have not searched for the goods, I only talked about them to others …	0.13
My most profound experience is I searched in ***ABC*** in the last second, then I turned to another website in the next second. To my surprise, an advertisement about what I searched in ***ABC*** was **recommended** to me.	0.15

why users were scared. First, many associated the *Facebook-Cambridge Analytica* privacy breach incident with situations they encountered before. They mentioned that after they had said something about a product in *DEF*, the product appeared as a recommendation (targeted advertisement) in *ABC* and they received many messages regarding the product. Others also said that after they searched for something in *ABC*, the thing then appeared as an advertisement (recommendation) on another website. Many users recounted similar experiences in their comments. The following is a typical example: "Whenever I said something in my daily life, *ABC* would send related recommendations to me. It's so scary."

Many users attributed the recommendations to big data. They were amazed how big data is able to capture and analyze all kinds of data inputs—not just textual data they had typed on their devices but also whatever they said in Chinese or dialects when they were online as well as offline. A number of comments mentioned that big data also extracted information from photographs and videos in their phones. The users had a sense of uneasiness for the recommendations because they felt that big data knows more about them than they know about themselves. This uncomfortable feeling is reflected in similar comments that said: "I thought I was imagining how the recommendations know so much about me, but it seemed that everyone shared the same experience."

The users also felt scared because they sensed they were monitored. Some said the recommendations they received were based on their supposedly private conversations. They felt they were under surveillance, like big brother was watching them. They said: "It is okay if big data just entices me to spend money, but if it monitors all my activities, then that is scary." Also, many users said they were scared because big data can lead to improper usage. Some comments about improper usage point to big data swindlers engaging in price discrimination. One user said data swindlers kept increasing the price of air tickets when s/he kept searching for them on the Internet.

After analyzing the frequently-used keywords, the study reviewed the negative comments to determine the underlying theme. Many comments resonated with the posts, and their theme is "*There is no online privacy.*" The following are examples of comments related to the theme:

"The Facebook incident is not unusual. Nowadays, do we still have privacy?"

cogent business & management

Table 3. Selected positive comments containing "Haha", "Exchange" and "Convenience"

Post content	Sentiment score
Haha ("哈哈")	
Haha, the best joke of the year.	0.99
Haha, XYZ, are you serious?	0.85
Exchange ("交换") and Convenience ("便利性")	
Hahaha, last time, your boss just said, "Chinese tends to **exchange convenience** with privacy".	0.82
Haha, people like to **exchange convenience** with privacy.	0.88
Who said "Chinese tends to **exchange convenience** with privacy" before?	0.79
What a humorous guy, they just said "Chinese tends to **exchange convenience** with privacy" before.	0.88

"Since early days, there is already no privacy."
"It is all transparent, there is no privacy."
"Here there is no privacy."
"In today's society, there is no privacy unless you seclude in remote mountains and isolate."
"There is no privacy nowadays, everyone is under surveillance."
"Big data doesn't provide a feeling of safety. It is like being watched all the time."
"It is like living under the sky with eyes."
"Everyone is naked in the cyberworld."
"Privacy has already been leaked long time ago."
Some comments accompanying the theme offer suggestions to mitigate problems with online privacy. They include: (a) close the permission setting for the microphone, (b) use the phone less often, (c) speak in dialects so that big data will not understand, (d) erase cookies, and (e) lobby for data protection legisation. Other comments said they are resigned to not having privacy. They reasoned that: (a) they could not avoid the problem, (b) they did not know what to do even though they know the problem existed, and (c) they had erased the cookies but the cookies from *ABC* could not erased. Some comments said, however, they found the targeted advertisements/recommendations useful. Before the age of big data, it was difficult to search for the things they wanted. So big data makes life easier and convenient.

4.3. Comments on topic#4 (reactions to XYZ's comments on data privacy)

There are four posts related to topic#4, which is about reactions to *XYZ*'s comments on data privacy. *XYZ* is one of the largest Internet companies in China. It provides a variety of services, such as search engine, online maps, and cloud drive. *XYZ* owns enormous personal data about the Chinese Internet users. The four posts for topic#4 have similar contents. An example of a translated post is as follows:

> On the 9th, President of XYZ spoke at the 2018 Boao Forum for Asia. In response to the Facebook privacy data breach, he had responded: 'XYZ has a very clear position that security and privacy are prerequisites for everything we do.'

As background to the above post, the CEO of *XYZ* had said in a public speech prior to the *Facebook-Cambridge Analytica* incident that under many circumstances, Chinese users were willing to exchange convenience and efficiency with their privacy (Shen, 2018). The posts related to topic#4 triggered a lot of comments as people still remembered what *XYZ* had said before. Table 3 shows some examples of the comments and their sentiment scores. About 47 percent of the comments are positive (sentiment score ≥ 0.7). The most frequently used keywords among the positive comments are: Haha ("哈哈"), exchange ("交换") and convenience ("便利性"). Even though the sentiment scores indicate the comments are positive, a closer examination show that most

comments are ironic and sarcastic. For example, users laughed at the posts and said "Haha". These comments are not expressing happiness or satisfaction with the posts. Instead, they are ridiculing or mocking what *XYZ* had said earlier. These comments put *XYZ* in a bad light, and people are skeptical that the company would protect users' data privacy. People may be less likely to trust the company as the company's statements are conflicting.

5. Discussion and conclusion

With the popularity of SNS, there are growing concerns that users' personal data may be shared inappropriately, as in the case of the *Facebook-Cambridge Analytica* privacy breach incident. This study analyzed the sentiments and perceptions among *Weibo* users after news about the *Facebook-Cambridge Analytica* incident went public. The results of this study show that Chinese SNS users were scared of privacy leakages, and that privacy leakages can originate not only from SNS such as *DEF*, but also from e-commerce sites such as *ABC*, and Internet companies such as *XYZ*. Like *DEF, ABC* and *XYZ* have huge collection of users' data. Companies that collect and track users' data have good knowledge about users and users are afraid that their privacy may be lost when their data are used inappropriately.

This study's results show that even though the *Facebook-Cambridge Analytica* incident was about using data to influence political outcomes, the focus of the comments on *Weibo* is not about politics. Instead, the comments centered on companies collecting data to influence people's purchasing decisions. The difference in focus could be because people in China feel uncomfortable discussing politics in public. It could also be because there has been a tremendous increase in the number of online shoppers in China in the last decade (Ma, 2015), and many people could relate very well to the experience of receiving targeted advertisements or recommendations. Users are scared of the recommendations because they wondered how big data knows so much about them. They feared they were under surveillance and that their data could be used inappropriately. They felt helpless as they could not avoid privacy issues when using the Internet. On the whole, users felt there was no online privacy.

Even though users were afraid of privacy invasion, there were few indications on what they would do to protect their online privacy. They did not say they would stop using the Internet. Some users reasoned that they were reliant or addicted to the Internet and so they cannot stop using it. These results show that users have a privacy paradox. A few users did said, however, they were willing to exchange their privacy for the benefits of convenience such as receiving useful recommendations.

According to the authors' knowledge, there are few studies about sentiments and perceptions of data privacy among Chinese SNS users. The results of this study show that many *Weibo* users believe there is no online privacy and that they are scared of online privacy invasion. These results have implications for research and practice. In terms of research, the results point to a few potential areas for future studies. First, many users were scared of the recommendations as they did not know how the algorithms for targeted advertisements work and how data were obtained about them. Future studies can determine if understanding how the recommendations were derived could help users become more receptive to the recommendations. Related to this area of research include investigating whether sentiments like apprehension about privacy invasion could affect users' trust and purchasing behavior. Second, this study found that "convenience" is a frequently-used keyword among the comments. "Convenience" is seldom discussed in the privacy literature. As "convenience" and "privacy" are important experiences for online users, future research can pay attention on how to provide "convenience" and "privacy" to online users. Future research can also use intelligent methods to predict users' privacy preferences (Nakamura et al., 2017) based on the premise that online users may trade privacy concerns with factors such as convenience and efficiency. Convenience and efficiency may help to resolve issues surrounding the privacy paradox.

In terms of practice, the results of this study imply that if users want to have better privacy, they need to look beyond the Internet for better alternatives. This could spur businesses to develop better privacy-protected services for users. Also, based on the current sentiments about online privacy among Chinese SNS users, it would be useful for companies that collect online data to clarify their privacy policy and assure users about online privacy protection. With the implementation of the new Personal Information Protection Law in China (Burgess, 2021), policy makers should also carry out enforcement to ensure that personal data collected by companies are not used inappropriately.

There are a number of limitations for this study. First, this study analyzed the comments from *Weibo* users and care must be exercised in generalizing the sentiments and perceptions to non-*Weibo* SNS users. Second, the data analysed in this study were based on a convenience sample using comments about a privacy breach incident. They may not be representative of the opinions about online privacy in general. Nevertheless, as an initial study about data privacy among Chinese SNS users, the results are useful and they provide implications for future research and practice. Third, the accuracy of the tool used for sentiment analysis in this study is around 80 to 85 percent (Ying et al., 2017). Even though the performance and accuracy of sentiment analysis have improved over the years (Dhaoui et al., 2017), it may be difficult to achieve absolute accuracy as sometimes the comments may be ironic or ambiguous and the software may have difficulty interpreting the comments. To overcome issues regarding the accuracy of sentiment analysis, this study reviewed the negative and positive comments to determine whether the sentiment scores are appropriate. The review found that the word "Haha" has a negative connotation rather than a positive sentiment when used in the context when users reacted to *XYZ's* comments about data privacy.

To conclude, this study has examined the sentiments and perceptions of online privacy among Chinese SNS users after news about the *Facebook-Cambridge Analytica* privacy breach incident was made public. The results show that people felt scared because they relate the *Facebook* data breach incident with personal experiences they encountered online, such as when e-commerce companies make purchase recommendations based on their supposedly private chats. The results also indicate that people were skeptical about companies protecting the users' data privacy, given that the CEO of *XYZ* had previously said that users were willing to exchange convenience and efficiency with privacy. As there are few studies that analyzed sentiments and perceptions of online privacy from the perspective of the Chinese SNS users, the results of this study provided insights and contributed to the data privacy literature.

Funding
The authors gratefully acknowledge the research funding provided by the University of Macau, grant reference MYRG2018-00118-FBA

Author details
Chang Boon Lee[1]
E-mail: cblee@um.edu.mo
Hio Nam Io[2]
E-mail: hnio@must.edu.mo
Heng Tang[1]
E-mail: hengtang@um.edu.mo
Trevor Wilmshurst
[1] Faculty of Business Administration, University of Macau, Taipa, Macau, China.
[2] School of Business, Macau University of Science and Technology, Taipa, Macau, China.

Citation information
Cite this article as: Sentiments and perceptions after a privacy breach incident, Chang Boon Lee, Hio Nam Io & Heng Tang, *Cogent Business & Management* (2022), 9: 2050018.

Note
1. ABC, DEF, and XYZ are used to replace the actual names of the companies.

Disclosure statement
No potential conflict of interest was reported by the author(s).

References
Acquisiti, A., & Gross, R. (2006). Imagined communities: Awareness, information sharing, and privacy on the Facebook. *Proceedings of the International Workshop on Privacy Enhancing Technologies* Germany, 36–58.
Bellman, S., Johnson, E., Kobrin, S. K., & Lohse, G. L. (2004). Internationla differences in information privacy concerns: A global survey of consumers. *The Information Society*, *20*(5), 313–324. https://doi.org/10.1080/01972240490507956
Burgess, M. (2021). *Ignore China's new privacy law at your peril*, WIRED, Retrieved May 11, 2021, from https://www.wired.com/story/china-personal-data-law-pipl/
Chen, Z. T., & Cheung, M. (2018). Privacy perception and protection on Chinese social media: A case study of WeChat. *Ethics Information Technology*, *20*(4), 279–289. https://doi.org/10.1007/s10676-018-9480-6
Choi, H., Park, J., & Jung, Y. (2018). The role of privacy fatigue in online privacy behavior. *Computers in Human Behavior*, *81*(1), 42–51. https://doi.org/10.1016/j.chb.2017.12.001
Debatin, B., Lovejoy, J. P., Horn, A., & Hughes, B. N. (2009). Facebook and online privacy: Attitudes, behaviors,

and unintended consequences. *Journal of Computer-Mediated Communication*, *15*(1), 83–108. https://doi.org/10.1111/j.1083-6101.2009.01494.x

DeGennaro, T. (2020). *The top ten most popular social media websites in China*, *Dragon Social*, https://www.dragonsocial.net/blog/social-media-in-china/

Dhaoui, C., Webster, C., Tan, L. P., Norberg, P., & Fortin, D. (2017). Social media sentiment analysis: Lexicon versus machine learning. *Journal of Consumer Marketing*, *34*(6), 480–488. https://doi.org/10.1108/JCM-03-2017-2141

Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research*, *17*(1), 61–80. https://doi.org/10.1287/isre.1060.0080

Edwards, D., Cheng, M., Wong, I. A., Zhang, J., & Wu, Q. (2017). Ambassadors of knowledge sharing: Co-produced travel information through tourist-local social media exchange. *International Journal of Contemporary Hospitality Management*, *29*(2), 690–708. https://doi.org/10.1108/IJCHM-10-2015-0607

Fiesler, C., & Hallinan, B. (2018). "We are the product": Public reactions to online data sharing and privacy controversies in the Media, *Proceedings of the 2018 Human Factors Conference in Computing Systems*, https://dl.acm.org/doi/10.1145/3173574.3173627

Gellman, R., & Dixon, P. (2011). *Online privacy: A reference handbook*. ABC Clio.

Hinds, J., Williams, E. J., & Joinson, A. N. (2020). It wouldn't happen to me": Privacy concerns and perspectives following Cambridge Analytica scandal. *International Journal of Human-Computer Studies*, *143*(4), 1–13. https://doi.org/10.1016/j.ijhcs.2020.102498

Isaak, J., & Hanna, M. J. (2018). User data privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer*, *51*(8), 56–59. https://doi.org/10.1109/MC.2018.3191268

Kennedy, G., & Zhang, X. (2017). China passes cybersecurity law. *Intellectual Property & Technology Law Journal*, *29*(3), 20–21.

Kong, L. (2007). Online privacy in China: A survey on information practices of Chinese websites. *Chinese Journal of International Law*, *6*(1), 157–183. https://doi.org/10.1093/chinesejil/jml061

Labrie, R., Steinke, G. H., Li, X., & Cazier, J. A. (2018). Big data analytics sentiment: US-China reaction to data collection by business and government. *Technological Forecasting and Social Change*, *130*(1), 45–55. https://doi.org/10.1016/j.techfore.2017.06.029

Lee, T. H., Sung, W. K., & Kim, H. W. (2016). A text mining approach to the analysis of information security awareness: Korea, United States, and China. *Proceedings of the 20th Pacific Asia Conference on Information Systems, PACIS 2016* Taiwan (Association of Information Systems).

Loukides, G., & Gkoulalas-Divanis, A. (2009). Privacy challenges and solutions in the social web. *XRDS: Crossroads, the ACM Magazine for Students*, *16*(2), 14–18. https://doi.org/10.1145/1665997.1666002

Ma, Y. (2021). *Number of online shoppers in China 2009–2020*, Statista Retrieved September 21, 2021, Statista https://www.statista.com/statistics/277391/number-of-online-buyers-in-china/

McCoy, K. (2017). Target to pay $18.5M for 2013 data breach that affected 41 million consumer, USA Today, 23 May. https://www.usatoday.com/story/money/2017/05/23/target-pay-185m-2013-data-breach-affected-consumers/102063932/

Min, K., Ma, C., Zhao, T., & Li, H. (2015). BosonNLP: An ensemble approach for word segmentation and POS tagging Li, J., Ji, H., Zhao, D., Feng, J. In *Natural language processing and chinese computing* (pp. 520–526). Springer.

Nakamura, T., Tesfay, W. B., Kiyomoto, S., & Serna, J. (2017). Default privacy setting prediction by grouping user's attributes and settings preferences Garcia-Alfaro, J., et al., . In *Data privacy management, cryptocurrencies and blockchain technology* (pp. 107–123). Springer, Cham.

Ortiz-Ospina, E. (2019). *The rise of social media*, *Our World in Data*, Sep 2019: https://ourworldindata.org/rise-of-social-media

Schneble, C. O., Elger, B. S., & Shaw, D. (2018). *The Cambridge Analytica affair and Internet-mediated research. EMBO Reports*. http://embor.embopress.org/content/early/2018/07/02/embr.201846579.

Shen, X. (2018). *Chinese internet users criticize Baidu CEO for saying people in China are willing to give up data privacy for convenience*, *South China Morning Post*, https://www.scmp.com/abacus/tech/article/3028402/chinese-internet-users-criticize-baidu-ceo-saying-people-china-are

Shklovski, I., Mainwaring, S. D., Skuladottir, H. H., & Borgthorsson, H. (2014). Leakiness and creepiness in app space: Perceptions of privacy and mobile app use. *Proceedings of the ACM Conference on Human Factors in Computing Systems* Toronto, Canada (New York: Association for Computing Machinery), 2347–2356.

Snelson, C. L. (2016). Qualitative and mixed methods social media research: A review of the literature. *International Journal of Qualitative Methods*, *15*(1), 1–15. https://doi.org/10.1177/1609406915624574

Standaland, A. J. S., & Lwin, M. O. (2013). Online privacy practices: Advances in China. *Journal of International Business Research*, *12*(2), 33–46.

Stempel, J. (2020). *Home Depot reaches $17.5million settlement over 2014 data breach*, Reuters, Nov 25: https://www.reuters.com/article/us-home-depot-cyber-settlement-idUSKBN2842W5

Sun, J. (2015). *Jieba Chinese word segmentation tool*, https://github.com/fxsjy/jieba

Tao, Y., Zhang, F., Shi, C., & Che, Y. (2019). Social media data-based sentiment analysis of tourists' air quality perception. *Sustainability*, *11*(1), 1–23. https://doi.org/10.3390/su11185070

Timms, H., & Heimans, J. (2018). *Commentary: #DeleteFacebook is just the beginning. Here's the movement we could see next*. *Fortune*, April 16. https://fortune.com/2018/04/16/delete-facebook-data-privacy-movement/

Vinodhini, G., & Chandrasekaran, R. M. (2012). Sentiment analysis and opinion mining: A survey. *International Journal of Advanced Research in Computer Science and Software Engineering*, *2*(6), 282–292.

Wang, Y. (2017). *The rise of Weibo: Lessons Twitter can learn from Chinese upstart*. Forbes. https://www.forbes.com/sites/ywang/2017/06/06/the-rise-of-weibo-lessonstwitter-can-learn-from-chinese-upstart/#4bf1d3ba20b0

Wang, Y., Balnaves, M., & Sander, J. (2020). *Shameful secrets and self-presentation: Negotiatign privacy practices among youth and rural women in China*, *SAGE Open*, Jan-Mar, 1–12.

Wang, Y., Norcie, G., & Cranor, L. R. (2011). Who is concerned about what? A study of American, Chinese and Indian users' privacy concerns on social network sites. In J. M. McCune, B. Balacheff, A. Perrig, A. R. Sadeghi, A. Sasse, & Y. Beres Eds., *Trust and trustworthy computing. Trust 2011. Lecture notes in computer science* (Vol. 6740). Springer, Berlin, Heidelberg 146–153 . https://doi.org/10.1007/978-3-642-21599-5_11

Wu, Y., Lau, T., Atkin, D. J., & Lin, C. A. (2011). A comparative study of online privacy regulations in the U.S. and China. *Telecommunication*, *35*(7), 603–616. https://doi.org/10.1016/j.telpol.2011.05.002

Ying, K., Pan, J., & Wu, M. (2017). Research on sentiment analysis of micro-blog's topic based on textrank's abstract. In *Proceedings of the 2017 International Conference on Information Technology*, Singapore, 27–29 December 2017, ACM, NY, 86–90.

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgements are more accurate than those made by human. *Proceedings of the National Academy of Science* 112 4 . https://doi.org/10.1073/pnas.1418680112

Cogent Business & Management **(ISSN: 2331-1975) is published by Cogent OA, part of Taylor & Francis Group.**