Claimed vs Observed Information Disclosure on Social Networking Sites

Phumezo Ntlatywa, Reinhardt A. Botha, Bertram Haskins
Institute for ICT Advancement and School of ICT
Nelson Mandela Metropolitan University
Port Elizabeth, South Africa
phumezon@gmail.com, reinhardta.botha@nmmu.ac.za, bertram.haskins@nmmu.ac.za

***Abstract*—Research on internet users reports a gap between reported privacy concerns and observed privacy behavior. This gap has also been reported on social networking sites like Facebook and MySpace. This current study explores the extent of this gap by examining individual Facebook components. Unlike the previous studies that have explored reported privacy concerns versus observed behavior, this study explores claimed information disclosure versus observed information disclosure. A questionnaire was used to examine the reported information disclosure and observations on public Facebook profiles were conducted to examine the actual behavior. The study shows that there is a gap between the reported information disclosure and observed disclosure, like in the case of reported privacy concern versus observed behavior. However, this study shows that the users act more securely than they claim to act.**

Keywords-reported behavior; observed behavior; Facebook; social networking sites; privacy; gap

I. INTRODUCTION

Social networking sites like Facebook, Twitter, LinkedIn and the like have given people the ability to communicate and share information with almost anyone, anywhere and at any time. An important aspect of social networking sites is the fact that users can publish status updates, upload photos, audio files and video files for their friends to view. However, users often publish information about themselves that is extremely personal or sensitive [1]. Although privacy settings can be used to protect private information, it is often found that users do not use these privacy settings [1], [2].

For the purpose of this study focus is put on Facebook. Facebook is currently the most popular social networking site and has the largest number of users. As of January 2011, Facebook had more than 600 million active users [3]. This number increased in the same year to 800 million users [4]. This makes Facebook the most popular social networking site and more popular than Google as the most visited site on the Internet [5]. Facebook is the biggest photo site, even bigger than Flickr or Photobucket [6].

Facebook started as a social networking site of Harvard University in 2004 created by Mark Zuckerberg. As with all social networking sites, participation requires a user to create a profile that lists the demographics of the user, a profile photo and requires that a user be above 13 years old. In addition, after users are signed on they can share different information types that can be updated at any time. Contributing to the success of Facebook is the developer platform which was launched in 2007. The developer platform enabled the development of applications that can be used directly or indirectly to interact with the application servers through Facebook [7]. These applications have been labeled to increase the risk of data harvesting [8].

The ease of information storage on the Internet makes it possible for user information to be harvested later. This information may then be used in situations where it can be harmful to the owner [9]. Scholars, the media and many other privacy stakeholders have raised concerns regarding the risks associated with the disclosure of personal information on social networking sites [1], [10–12]. Research on internet users has shown that users claim that privacy is important to them, but do not act likewise [13], [14]. This has also been debated to be the same case on social networking sites users [15], [16], [17] [18]. All these studies focused on the privacy concerns versus observed behavior. This gap between privacy concerns and observed behavior is termed as "privacy paradox" [19]. Other well-known studies in this field include [1], [10], [11], these studies focused on the amount and type of information users disclosed and on the reported privacy concerns or preferences.

This current study puts focus on the claimed information disclosure versus observed information disclosure. According to our knowledge this is the first study to focus on the users' claimed behavior, rather than claimed privacy preferences as previous studies have. This research will add to the existing body of knowledge within privacy studies, predominantly within the rapidly growing area of studying human behavior on social networking sites. This study formulated the following research question:

Is there a gap between claimed information disclosure and observed information disclosure on Facebook?

The following sections will discuss the methods employed in arriving at the results of this study. The results will be discussed, followed by the discussion of future work and the conclusion. First we start by discussing the participants of the study.

II. RESEARCH METHODS

The research methods chosen for this study were questionnaires and observations. Questionnaires were used to understand the claimed information disclosure on Facebook. Requests to respond to the questionnaire were made via e-mail to

all the recruited participants. Two participants failed to receive the invitation because they provided invalid e-mail addresses. Of the 167 requests made, 100 participants responded to the questionnaire, giving a 61% response rate. To examine the observed information disclosure, observations were made on the users' public Facebook profiles. A total of 131 participants were observed at a Facebook component-level. We approached this study on a component base. Participants were asked to indicate all the Facebook components they had disclosed on their profiles. During observations all those components were observed to identify the gap for each component.

All data collection took place during the month of September 2011. During data collection we adopted the classifications of Facebook components that Nosko, Wood and Molema described in 2010 [20]. They categorized Facebook components into the following three categories:

1. **Personal identifiable information:** This refers to hometown, gender, birthday, birth year, e-mail address, address, profile photo, photos, and personal website. This is any information about an individual that can be used to distinguish or trace an individual's identity or linkable to an individual [21].
2. **Sensitive personal information:** This refers to employer information, secondary school, university, current location, mobile number, friends' list and relationship status. This is any personally revealing or sensitive information that could be used to locate an individual and could be used to threaten or harm the owner.
3. **Potentially stigmatizing information:** This refers to the wall, religious status, political views, people who inspire me, favorite quotations, music, books, movies, television show, games, sport, activities, interests, gender interested in dating, language and personal description. This is any information that could result in stigmatization within the society.

Our methodology differs from the previous studies in the sense that we tested our research question for each Facebook component.

III. PARTICIPANTS

Participants in this study were freshmen enrolled for IT at Nelson Mandela Metropolitan University for the academic year of 2011. During recruitment 167 participants indicated willingness to participate in the study by signing a participation form. Ethical clearance was obtained before data collection and participation in the study was voluntary. Fig. 1 shows the classification of participants at different stages of the study. The participants that took part in each stage of the study are explained on the relevant sections below. The gray area depicts the probability of participants that filled the questionnaire and were also observed. Possibility is that some participants that completed the questionnaire were not observed and some that did not complete the questionnaire were observed. Table 2 gives the groupings of the participants during observations. The table explains all the information disclosed by each group and the number of participants that were in that group.

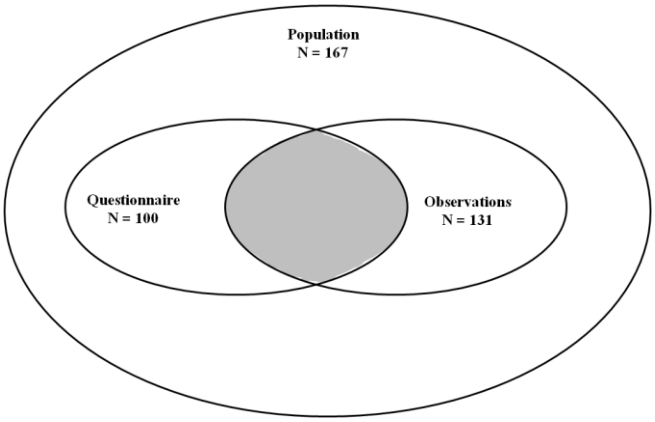

Figure 1. Participants

IV. RESULTS

A. Questionnaire

Males dominated the sample, with 75% while females registered 25%. A variety of age groups were represented in the sample, ages 18 – 21 dominated the sample with 91%, with ages 22 – 25 and 26 – 30 on 7% and 2% respectively. The majority of the participants (63%) have used Facebook for more than a year and 37% have used Facebook for less than a year. The full results of the claimed information disclosure are given in Table 1.

B. Observations

Of the 167 participants recruited 35 participants could not be found on Facebook and one participant was deemed invalid as she was already a Facebook friend of the researcher. As we could not determine the behavior regarding specific Facebook components of the 21% that could not be found, they were disregarded for the rest of the observation. However, if they were using a pseudonym they could still be exposing the detail components to relative strangers who might still identify them.

Table 1 gives the full results of the claimed information disclosure versus observed information disclosure. The table provides the list of Facebook items examined on the first column. The second and the third columns show the claimed information disclosure and the observed information disclosure respectively. The last column gives the difference between the claimed information disclosure and the observed information disclosure. As an example, for hometown, 92% claimed that they have disclosed it on their profiles, while 61% were observed to have disclosed it on their profiles. Therefore the difference between claimed disclosure and observed disclosure is 31%. The information in Table 1 is represented in three figures according to the difference between the claimed information disclosure and observed information disclosure.

TABLE I. CLAIMED BEHAVIOR VERSUS OBSERVED BEHAVIOR

Personal identifiable information	*Claimed disclosure (N = 100)*	*Observed disclosure (N = 131)*	*Difference between behaviors*
Hometown	92%	61%	31%
Gender	99%	91%	8%
Birthday	90%	19%	71%
Birth year	68%	11%	57%
E-mail address	74%	5%	69%
Address	26%	2%	24%
Profile photo	100%	97%	3%
Personal website	13%	2%	11%
Sensitive personal information			
Employer	18%	13%	5%
Secondary school	77%	68%	9%
University	94%	77%	17%
Current location	71%	57%	14%
Mobile number	60%	7%	53%
Friends list	71%	87%	-16%
Relationship status	59%	21%	38%
Potential stigmatizing information			
Religious status	61%	23%	38%
Political views	42%	10%	32%
People who inspire you	31%	15%	16%
Favourite quotations	51%	16%	35%
Favourite music	76%	59%	17%
Favourite books	55%	39%	16%
Favourite TV show	74%	56%	18%
Favourite games	52%	35%	17%
Favourite sport	53%	19%	34%
Activities	58%	41%	17%
Interests	68%	37%	31%
Gender interested in dating	55%	39%	16%
Language	72%	25%	47%
Personal description	29%	27%	2%

TABLE II. CLASSIFICATION OF PARTICIPANTS

Status	*Personal identifiable information*	*Sensitive personal information*	*Potential stigmatizing information*	*Number of participants*
Invalid & not found	N/A	N/A	N/A	21%
Completely open	All	All	All	12%
Somewhat open	Hometown, gender, profile photo	Secondary school, university, friends list	All	47%
Somewhat closed	Gender, profile photo	Secondary school, university, friend's list	N/A	13%
Completely closed	N/A	N/A	N/A	7%

Fig. 2 shows the difference between claimed information disclosure and observed information disclosure for components that had more than 50% difference. All the components in this figure show that users acted more securely than they claimed. The observed information disclosure is less than claimed information disclosure. This is counterintuitive and contrary to popular studies. In the case of reported privacy behavior versus observed privacy behavior, most studies revealed that users act less securely than their privacy preferences. However, the result of this group suggests otherwise. A discussion on the reasons that might cause this is discussed with the results of the study on the discussion section below.

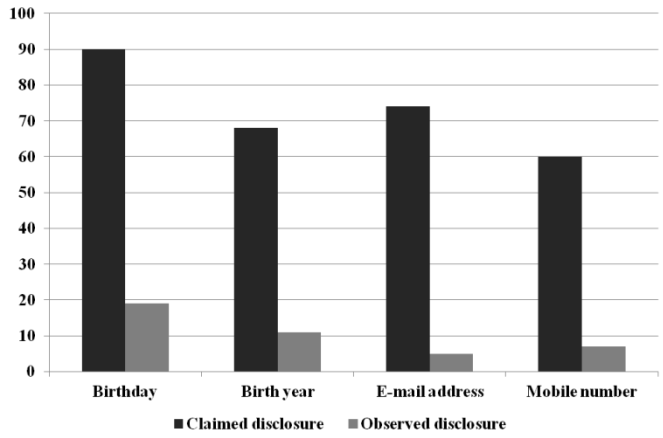

Figure 1.Components with more than 50% difference between claimed and observed disclosure

Fig. 3 shows the difference between claimed information disclosure and observed information disclosure for components that had a difference between 31% - 50%. All the components in this group show that users acted more securely than they claimed. Again, the result of this group gives a different finding than previous studies.

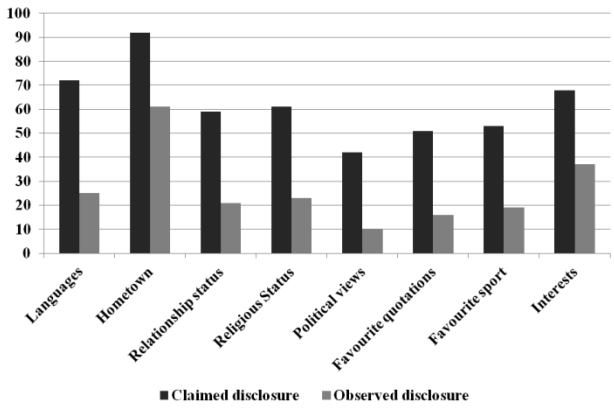

Figure 2. Components with 31% - 50% difference between claimed and observed disclosure

Fig. 4 shows the difference between claimed information disclosure and observed information disclosure for components that had a difference that is less than 30%. The results on this group show that users acted more securely than they claimed with the exception of friends list. The friends list was the only Facebook component where the observed information disclosure scored more than claimed information disclosure. The following section will discuss the results of the study.

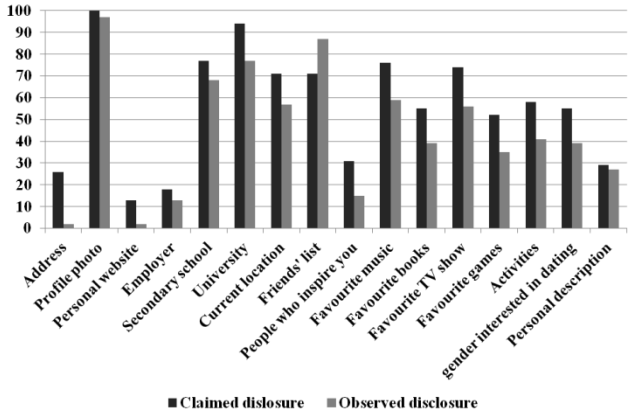

Figure 3. Components with less than 31% difference between claimed and observed disclosure

V. DISCUSSION

This study conducted an exploration in which we contrasted self-reported/claimed information disclosure against observed information disclosure of Facebook users. Our preliminary research question was to examine if there is a gap between claimed information disclosure and observed information disclosure on Facebook. This study discovered that, as it is in the case of reported privacy concern and observed privacy behavior [15–17], there is a gap between claimed information disclosure and observed information disclosure. The gap was found for all Facebook components except for the friends list.

All the previous studies have argued that users of social networking sites report that privacy is important to them but do not act according to their privacy preferences. However, in the case of claimed behavior this study shows that users act more securely than what they claim. This is counterintuitive to popular views brought about the forerunner studies in this field. Keep in mind that this study examined a different variable than the previous studies. This counterintuitive finding can be attributed to a number of reasons and these are discussed below.

The first reason might be that this study is examining a different variable than the previous studies. The methodology used in this study is not the same as the previous studies in the sense that this study examined the gap for each Facebook component. This study extends that the gap is not only on privacy concerns versus observed privacy, but also found on claimed information disclosure versus observed information disclosure. Research in other fields like organizational behavior research has found that users are bias in self-report. They tend to under-report behaviors deemed inappropriate by researchers or other observers, and they tend to over-report behaviors viewed as appropriate [22]. We do not rule out the possibility of this theory in this study as well.

Another reason would be that the users are not really aware of who currently can view what information on their profiles. This maybe because users have not revisited their privacy settings after all the changes Facebook has done on the site's privacy settings. Therefore, users might be unaware of their disclosure. Research over the years have argued that most users rarely alter default settings [1], [23–25]. This might also be the case also for the participants of this study.

Another cause might be that the users might be sharing their information with friends and not with strangers. This study was approached from a stranger point of view. Therefore we could only observe what a normal stranger would see. We would have wanted to examine the gap from a Facebook friend point of view, but the ethical clearance we got from the University only limited us to stand on a stranger point of view. Table 2 shows that only 12% of the participants we were able to observe all possible Facebook components on their profiles. 47% of the participants had somewhat open profiles while the rest yielded little to no information. From all the participants it is interesting to note that 21% of participants were not disclosing information that could easily be associated with by a stranger. It is assumed that the participants maybe using pseudonym names and that maybe another form of privacy protection mechanism.

Additionally, both the number of the participants that returned the questionnaire and the participants that were observed are less than the number of participants recruited. This makes it a

possibility that some participants that were observed did not fill the questionnaire and some that filled the questionnaire were not observed. Again, our ethical clearance did not allow us to collect names during the questionnaire stage and we could not verify that the participants we observed are exactly the same as the ones that administered the questionnaire. Again, not all recruited participants filled the questionnaire. Both the reasons for filling the questionnaire and not filling the questionnaire are unknown to the researchers.

Lastly, we did not specify to the participants the perspective we were questioning their information disclosure from. The participants could have misunderstood the question to think that we were asking from a Facebook friend's point of view. The following section concludes the study and also discusses future work.

VI. Conclusion

Social networking sites provide users with a platform to create maintain online social networks and share information. Along with the benefits of social networking sites there are threats, risks and privacy concerns with sharing information with large amounts of people. The privacy implications on these sites are the cause scholars put effort to carefully study these sites. The results of this study contribute to the body of knowledge within the rapidly growing area of social networking sites. There are many reasons for the discrepancy between claimed information disclosure and observed information disclosure. This unpredictable behavior adds to the difficulty that developers of social networking sites have in catering to the privacy needs and concerns of all their users

Not only does the study present a counterintuitive result, but it examines a different variable using a different methodology than previous studies. Again, future study will look to compare the individual's questionnaire results and the observations. We could not do this because the questionnaire was on anonymous basis

In evaluating the results of this study, certain limitations should be taken into consideration. Firstly, this only observed the participants from a stranger point of view. Future work will look to observe participants from a Facebook friend point of view to determine if the gap will still be present or decreased.

This study could not determine if the participants that filled the questionnaire were the same as the participants observed. Future study will look to observe the same participants as the ones that filled the questionnaire.

References

[1] R. Gross and A. Acquisti, "Information revelation and privacy in online social networks," in *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society*, New York, USA, 2005, pp. 71–80.

[2] R. Goettke and J. Christiana, "Privacy and social networking sites," 2007, http://www.eecs.harvard.edu/ cs199r/fp/RichJoe.pdf [Accessed: 06-Jun-2011], unpublished.

[3] A. Ahmad, "Social network sites and its popularity," *International Journal of Research and Reviews in Computer Science*, vol. 2, no. 2, pp. 522 – 526, Apr. 2011.

[4] A. Ostrow, "Facebook now has 800 million users," *Mashable Social Media*, 22-Sep-2011. [Online]. Available: http://mashable.com/2011/09/22/facebook-800-million-users/. [Accessed: 19-Mar-2012].

[5] J. Jiang, C. Wilson, X. Wang, P. Huang, W. Sha, Y. Dai, and B. Y. Zhao, "Understanding latent interactions in online social networks," in *Proceedings of the 10th Annual Conference on Internet Measurement*, New York, USA, 2010, pp. 369–382.

[6] C. H. Bentley, "Citizen journalism: back to the future?," in *Geopolitics, History, and International Relations*, vol. 3, New York, USA: Addleton Academic Publishers, 2008, pp. 103 – 118.

[7] D. M. Boyd and N. B. Ellison, "Social network sites: definition, history, and scholarship," *Journal of Computer-Mediated Communication*, vol. 13, no. 1, pp. 210–230, 2007.

[8] D. Evans and A. Felt, "Privacy protection for social networking platforms," in *2008 IEEE Symposium on Security and Privacy*, Oakland, 2008, vol. 2, pp. 1–8.

[9] A. Azimi and A. A. Ghomi, "Social networks privacy issues that affect young societies," in *Planetary Scientific Research Center Proceeding*, Bangkok, 2011, pp. 35 – 39.

[10] T. Govani and H. Pashley, "Student awareness of the privacy implications when using Facebook," 2005, http://lorrie.cranor.org/courses/fa05/tubzhlp.pdf [Accessed: 16-Nov-2011].

[11] T. Taraszow, E. Aristodemou, G. Shitta, Y. Laouris, and A. Arsoy, "Disclosure of personal and contact information by young people in social networking sites: An analysis using Facebook profiles as an example," *International Journal of Media & Cultural Politics*, vol. 6, no. 1, pp. 81–101, 2010.

[12] S. Trepte and L. Reinecke, Privacy online "perspectives on privacy and self-disclosure in the social web, 1st ed. Heidelberg: Springer, 2011.

[13] S. Spiekermann, J. Grossklags, and B. Berendt, "E-privacy in 2nd generation e-commerce: privacy preferences versus actual behavior," in *Proceedings of the 3rd ACM conference on Electronic Commerce*, New York, USA, 2001, pp. 38–47.

[14] M. Teltzrow and A. Kobsa, "Impacts of user privacy preferences on personalized systems," in *Designing Personalized User Experiences in eCommerce*, vol. 5, C.-M. Karat, J. O. Blom, and J. Karat, Eds. Dordrecht: Kluwer Academic Publishers, 2004, pp. 315–332.

[15] A. L. Young and A. Quan-Haase, "Information revelation and internet privacy concerns on social network sites: a case study of Facebook," in *Proceedings of the Fourth International Conference on Communities and Technologies*, New York, NY, USA, 2009, pp. 265–274.

[16] J. Fogel and E. Nehmad, "Internet social network communities: risk taking, trust, and privacy concerns," *Computers in Human Behavior*, vol. 25, no. 1, pp. 153–160, Jan. 2009.

[17] Z. Tufekci, "Can you see me now? audience and disclosure regulation in online social network sites," *Bulletin of Science, Technology & Society*, vol. 28, no. 1, pp. 20 –36, Feb. 2008.

[18] B. Reynolds, J. Venkatanathan, J. Gonçalves, and V. Kostakos, "Sharing ephemeral information in online social networks: privacy perceptions and behaviours," in *Proceedings of the 13th IFIP TC 13 international conference on Human-computer interaction - Volume Part III*, Berlin, Heidelberg, 2011, pp. 204–215.

[19] S. Barnes, "A privacy paradox: social networking in the United States," *First Monday*, vol. 11, no. 9, 2006.

[20] A. Nosko, E. Wood, and S. Molema, "All about me: disclosure in online social networking profiles: the case of Facebook," *Computers in Human Behavior*, vol. 26, no. 3, pp. 406–418, May 2010.

[21] E. McCallister, T. Grance, and K. Scarfone, "Guide to protecting the confidentiality of personally identifiable information (PII)." National Institute of Standards and Technology, Apr-2010.

[22] S. I. Donaldson and E. J. Grant-Vallone, "Understanding self-report bias in organizational behavior research," *Journal of Business and Psychology*, vol. 17, no. 2, pp. 245 – 260, 2002.

[23] W. E. Mackay, "Triggers and barriers to customizing software," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Reaching Through Technology*, New Orleans, Louisiana, United States, 1991, pp. 153–160.

[24] R. C. Shah and C. Sandvig, "Software defaults as de facto regulation the case of the wireless Internet," *Information Communication Society*, vol. 11, no. 1, pp. 25–46, 2008.

[25] R. Shah and C. Sandvig, "Software defaults as de facto regulation: the case of wireless aps," in *Proceedings of The 33RD Research Conference on Communication, Information and Internet Policy*, Virginia, USA, 2005, vol. 11, pp. 25 – 46.