Empirical Analysis of Online Anonymity and User Behaviors: the Impact of Real Name Policy

Daegon Cho
H. John Heinz III College
Carnegie Mellon University
daegonc@andrew.cmu.edu

Soodong Kim
Google, Inc.
soodongandrew@gmail.com

Abstract

Advanced information technologies have enabled the emergence of new communication practices. Since online communications do not rely on physical interaction, individuals can communicate with others without disclosing their real identities - by either staying anonymous or using pseudonyms. In this paper, we examine the relationship between online anonymity and users' communication behavior taking advantage of a natural experiment, which occurred in South Korea in 2007. In July 2007, South Korea enacted a Real Name Verification Law, requiring Internet users to verify their real names before posting messages at popular websites, including major portals and newspaper websites. This unique policy initiative allows us to investigate how South Korean citizens' behaviors changed as function of different levels of online anonymity before and after the law's enactment. Our finding suggests that the enhanced identification process reduced the number of messages containing swear words and slanderous comments. The change was limited to a certain group of users and to specific types of behaviors: swear words and anti-normative expressions significantly decreased among more-frequently participating users, whereas no substantial similar changes were found for less-frequently participating users. This study contributes to online anonymity and privacy literature and *can shed light on useful implications to policy makers.*

1. Introduction

With the growth in sociality and interaction around online discussion forums, these mediums are increasingly becoming places for communities to discuss and to address common issues [1], [2]. That is, Internet-based information technologies have enabled the emergence of new types of communicative practices [3]. The unique characteristic of the Internet communication is *anonymity (or pseudonymity)*. People surf the web, talk, and post messages without exposing their identities, interests, and activities to others [4]. This salient feature provides some merits in online discussion by which participants are under an equal condition, no matter what their backgrounds are. For instance, minorities in the society can have an equal opportunity to express their thoughts. However, at the same time, anonymity is likely to counter social norms and lead to undesirable phenomena. Malicious profanities and groundless online rumors can proliferate quickly under this circumstance, and those online slanders may lead to serious privacy invasion and defamation [1], [5], [6].

Not surprisingly, there have been ongoing debates surrounding freedom of speech and anonymity by researchers and policy makers. In this respect, South Korea offers a compelling case. According to a recent study,[1] South Korea's broadband penetration reached 95%, which was the highest rate among those of all 57 surveyed countries. This fact indicates that cyberspace has become a more important part of daily life for South Koreans than people in any other country. In spite of several positive impacts of this advanced network, the radical change has created several negative spillover effects in South Korea's society. For instance, online harassment in anonymous cyberspace has frequently occurred, such as the aforementioned 'Dog Shit Girl'[2] case. This study examines the impact of the policy that was firstly taken into effect in 2007 by which Internet users are required to verify their real identity when

[1] http://www.strategyanalytics.com/default.aspx?mod=PressReleaseViewer&a0=4748

[2] When a girl riding a South Korean subway refused to clean up her dog's excrement, a witness took pictures of her and posted them on a popular website, along with an account of her misbehavior. Within hours, she and her dog were recognized everywhere they went. "Within hours, she was labeled 'dog-shit-girl' and her pictures and parodies were everywhere on the cyberspace. Within days, her identity and her past were revealed. Request for information about her parents and relatives started popping up and people started to recognize her by the dog and the bag she was carrying as well as her watch, clearly visible in the original picture. All mentions of privacy invasion were shouted down with accusations of being related to the girl. (Last sentence unclear – did the relatives speak of privacy invation?)
(Source: http://boingboing.net/2005/06/29/woman_doesnt_clean_u.html)

978-0-7695-4525-7/12 $26.00 © 2012 IEEE
DOI 10.1109/HICSS.2012.241

they write at frequently visited websites.[3] By collecting unique data from the two most popular online discussion forums subject to the law, we first examine how the law has affected the use of swear words and anti-normative expressions that lead to privacy invasion. Our findings suggest that the proportion of aggressive postings has decreased on the pseudonym-based forum after the law was enacted, whereas the law was not influential on the website in which real name policy has been used regardless of the law. Second, even though the main policy goal to reduce insulting words was achieved, the enforcement can have a spillover effect if the law discourages user participation. But, our results state that the law is not associated with the participation. Individual-level analysis was also conducted to see whether particular users actually changed their behaviors, and the finding suggests that there is a limitation in forcing users to change their behaviors by law.

This paper contributes to the growing literature on online privacy and anonymity. Whereas the previous research mostly relied on either experimental design or survey approach due to the lack of real world data, Real Name Verification Law provides an abundant source of data for empirical analysis. Also, the paper can shed light on the effectiveness of Internet regulation in terms of privacy and anonymity to policy makers. As cyberbullies, online fraud, and privacy invasion has become growing issues in most countries, appropriate controls or sanctions may be necessary in some cases. To our knowledge, this paper is the first study to examine the policy impact relevant to online anonymity with a large-scale real-world dataset. Furthermore, empirical analysis regarding online user behavior under different anonymity levels strengthens the previous findings that used experimental design.

The paper proceeds as follows: Section 2 discusses the description of legislation and related literature. Section 3 lays out the research design and method. Section 4 describes the data, and main results regarding policy impacts are reported in Section 5. Finally, we conclude sin Section 6.

2. Backgrounds and Relevant Works

2.1. Real Name Verification Law

The South Korea presidential election in 2002 was a turning point from the traditional campaign to an Internet-oriented campaign [7]. Citizens' online-based political support clubs and online media was a critical factor for the winning campaign by mobilizing young voters [7]. In this process, however, the election was inundated with a variety of rumors and defamation; therefore, the National Assembly enacted the Real Name Verification Law that was applied to election-related online discussion forums for the first time in 2005 to protect the privacy of candidates and to nurture more reliable Internet politics. Under this rule, only verified users with their real identity were able to express their opinions on those websites.[4] Due to the constant social issues caused by online slanders, the extended version of Real Name Verification Law was launched in July 2007. This law was immediately effective across all websites with an average daily viewership of over 300,000.[5] In 2009, the law was extended to websites with an average daily viewership of over 100,000. Consequently, 37 and 153 websites were subject to the law in 2007 and in 2009, respectively. According to the law, a person involved in online libel can be sentenced up to 3 years of imprisonment or imposed a fine of up to 20 million Korean won (about $20,000). A number of prosecutions have been reported by the National Policy Agency to date, and the Constitutional Court is contemplating the constitutionality of the law. However, this does not imply users' real name would be disclosed on discussion forums. Most websites adopted a pseudonym-based discussion system rather than using real name. That is, this policy did not mandate users' real name to appear on web pages, once the user's verification process was completed.

The main policy goal of the Real Name Verification Law is to reduce aggressive postings that may cause defamation and slanders. However, at the same time, if this strict legal enforcement leads to discouragement of users' willingness-to-express related to freedom of speech, the law is dysfunctional and the results may yield undesirable spillover effects [8], [9]. Researchers in South Korea tried to examine the impact of Real Name Verification Law in South Korea, but most argued the legal aspects and appropriateness of the policy [10-12]. More recently, an empirical study was pioneered to explore the effects of the law with a real world dataset [13]. However, the study only observed a very short period

[3] To the best of our knowledge, South Korea is the only country in which the identity verification regulation is enacted. China and the Middle East have strict Internet regulations, but these can be regarded as Internet censorship, which are distinctive.

[4] In order to verify a user's identity, Resident Registration Number (RRN) is used at the associated websites.

[5] Since most large websites in Korea had adopted real identity verification in their sign-up process, there was no grace period regarding this law. Existing subscribers (already verified with their RRN) did not have to do any further action.

(10 days before and after the law) with a small size of sample.

2.2. Anonymity and Pseudonymity

First of all, the relevant terminology should be clearly defined. *Anonymity* is the state of being unidentifiable. *Unlinkability* and *unobservability* are accompanied with anonymity [14]. The distinction between anonymity and *pseudonymity* is another important issue in this context. Pseudonymity indicates that the user maintains one or more persistent pseudonyms that are not connected to the user's physical identity [15], and the term comprises all degrees of linkability to a subject. For example, third parties (website operators) may have the possibility to reveal the identity of the holder (users) in order to provide means for investigation or prosecution. In online communications, pseudonyms contain various degrees of anonymity [16], [17]. Highly linkable *public pseudonyms* indicate that the link between a pseudonym (or nickname) and a real identity is publicly known or easy to discover. *Unlinkable pseudonyms* mean that system operators or third parties cannot detect a certain pseudonym's real identity [14]. In this respect, Real Name policy in South Korea causes a switch from an unlinkable pseudonymous condition to a publicly pseudonymous condition in Internet space.

In addition, the notion that identity shapes behavior has been affirmed in diverse fields, such as psychology, economics, organizational behavior, marketing, and information systems [18-20]. More recently, trust and privacy concerns in social networking sites and electronic commerce sites are being acknowledged in information systems field [18], [21-24]. In addition to the academic literature, disclosure of a user's real identity has become more controversial on the Internet itself. Though some users prefer to use pseudonyms on social network sites, Google's SNS (Google+) nudges users to use their real name, and according to Facebook's policy, accounts can be deleted if their owner is found using fake names or another person's real identity.

2.3. Deindividuation and CMC

The paper is directly associated with a large body of literature on the social value of anonymous communication and deindividuation research. Two main streams are existent with regard to anonymous communication: positive and negative aspects. On the positive side, anonymous communications enable minorities to express their own opinions and to maintain privacy protection [16]. An anonymous environment is helpful in promoting more active involvements without revealing personal identity [25]. Also, anonymous speech helps to settle the imbalance of information [26]. According to SIDE (Social Identity model of Deindividuation Effects) model, anonymity should accentuate the effects of the salient social identity and the dominant normative response associated with it [27]. Anonymity leads to reduction in behavioral constraints and enables individuals to engage in behavior they would not engage in when identified [28].

On the other hand, opponent groups argue that it is more likely that defamation, threat, insulting words, and slander can occur under an anonymous communication environment [29]. According to classical deindividuation theory, anonymity in the group can lead to reduced self-awareness and influence of social norms that ultimately correspond to anti-normative behaviors [30]. Finally, according to recent literature on impacts and influence of computer-mediated communication (CMC), reduction of identifiable information reduces self-awareness and stimulates anti-normative behavior [31]. However, the aforementioned SIDE model proposes that certain features of CMC can intensify social influence and enhance normative behavior [30], [32].

The main limitation of the previous studies is that they are based on either experiments or surveys. To the best of our knowledge, there is little empirical studies in investigating online communication behaviors by using real-world data. We use field data and take advantage of a large-scale natural experiment to provide empirical evidence that exogenous shocks to anonymity levels indeed changes the users' behaviors.

3. Research Design and Method

3.1. Research Design

Three research questions are explored in this study.

RQ1. Was Real Name Verification Law positively associated with privacy protection by reducing anti-normative expressions?
RQ2. Did Real Name Verification Law affect users' participation and posting activities?
RQ3. How did the legal enforcement lead to individual's behavioral shift?

To assess the impact of the law on user behaviors, we first need to identify key variables. *Posting* is a

message that is written by an individual (nickname), and each posting is represented with time, date, writing number by time order, and nickname set by each user.

Previous studies suggested that anonymity may lead to greater use of abusive words and slander by being free from criticism and evaluation, because others cannot observe and trace the commenter [5], [6]. Along these lines, we conducted content analysis by counting aggressive messages, including swear words and abusive expressions. To count postings including profane language, we used a set of 34 swears words and a set of 52 slanderous expressions. On the one hand, we suppose that postings with frequently-used swear words are regarded as aggressive expressions in the online forum. On the other hand, we define anti-normative expressions as postings that would be associated with defamation of others, such as politicians and discussion participants. The 34 swear words are selected according to a suggestion by Nielsen Korea, and the 52 slanderous expressions include a vulgar nickname for the President and other famous politicians and rude phrases applied to discussion opponents. In this context, Table 1 displays detailed research design and specification.

Table 1. Research Design

	Site 1	Site 2
Level of anonymity	Nickname (Pseudonym)	Real name
Subject to the law	Yes	Yes
Type of user (User Group by the number of monthly postings)	Light User: (1) 1 posting (2) 2 postings Middle User: (3) 3 postings (4) 4-10 postings Heavy User: (5) 11-15 postings (6) Over 15 postings	

	Periods of study	Descriptions
Treatment Group	Short-term	60 days before & after the law
	Long-term	6 months before & after the law
Control Group	Robustness check	Same periods in the previous year on a reference date (July, 31st 2006)

Site 1 and Site 2[6] subject to the policy are selected for the analysis, and they are distinctive in terms of the level of anonymity. In order to examine the behavioral shift across the nature of users, six categories and three groups of users are defined on the basis of the number of postings per month as shown in Table 1. We also checked the robustness of the results by conducting an identical analysis with a data from the same period in the previous year when the law was not in effect.

6 More details about Site 1 and Site 2 are explained in the following section.

3.2. Research Methods

We completed two sets of analysis – one on the aggregated data, and one on individual-level data. In order to evaluate the effect of the law on user behavior, we measured differences in communication patterns between the periods (short-term and long-term) before and after the law. Two-sample tests of proportion were first applied: the standard hypothesis test is H_0: $\boldsymbol{p}_{j,before\ law} = \boldsymbol{p}_{j,after\ law}$ against the alternative (one-sided) H_1: $\boldsymbol{p}_{j,before\ law} > \boldsymbol{p}_{j,after\ law}$ where $\boldsymbol{p}$ is a proportion of postings including aggressive or anti-normative expressions out of total postings in each category j. We then conducted paired t-tests to compare posting behaviors before and after the law.
This aggregation approach gauges the effectiveness of the law, but it may bear a drawback of being unable to see individual behavioral change. For example, users who had posted messages before the law may leave the discussion board after the law, and a new user may join the discussion after the law. Therefore, we also examined change in posting behaviors at the individual level using regressions analysis. More specifically, we focused on two components: (1) whether or not individuals who used swear words before the law continue the behavior after the law and (2) whether or not individuals who participated in the discussion before the law continue the participation after the law. In order to do this, we applied a difference-in-difference method that is useful for examining a policy impact. Samples and variables are defined in Table 2, and a specification is as follows:

$$y_i = \alpha + \beta_1 dTreat_i + \sum_{j=2}^{6} \beta_j dTreat_i dG_{ij} + \varepsilon_i \qquad (1)$$

Table 2. Description of Variables

	(1) Aggressive expressions	(2) Participants
Sample (Treatment)	Users who used swear words within a month before the law	Users who participated within a month before the law
Sample (Control)	Users who used swear words in July 2006	Users who participated in July 2006
y_i	1, if user i expressed swear words again within a month after a reference point. 0, otherwise	1, if user i still do postings within a month after a reference point. 0, otherwise.
$dTreat_i$ & $dGroup_{ij}$	1, if the user i is in Treatment group. 0, otherwise. & 1, if the user i is in Group j. 0, otherwise	

Linear Probability Model is used to estimate this specification while allowing heteroskedasticity-robust standard error. The baseline group is Group 1 (1 postng) in the Light User Category. Coefficient estimates of *dTreat* and five interaction terms (*dTreat* * *dG*) by each group dummy are key variables of interest. A negative sign of these coefficient estimates would indicate that the policy indeed affected the reduction of postings involving swearing and of participants, respectively.

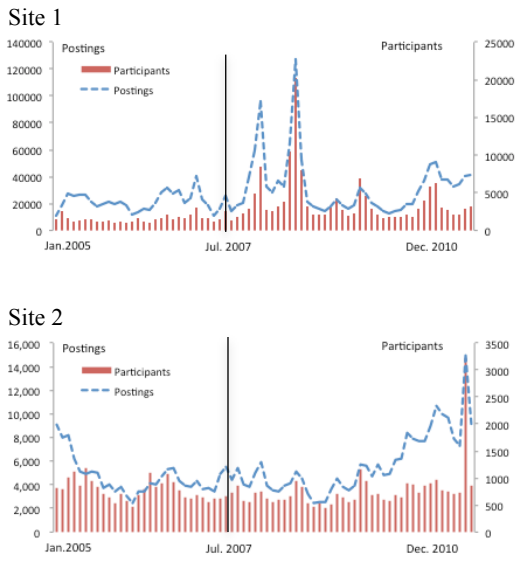
4. Data

By utilizing a web-crawling method[7], data is collected from two sources. Site 1is one of the top portal websites in terms of daily visits and the length of operation. Site 1 is particularly famous for its discussion forum section, and the board in which political issues are discussed is selected for the study. Political and diplomatic issues, including topics relevant to North Korea, public policy, and elections are more controversial in the online discussion. Site 2 is the discussion forum of a leading newspaper site.

Site 1 uses a nickname (pseudonym) policy; by contrast, a discussion forum in Site 2 uses real names on the website itself rather than using a pseudonym. That is, Site 1 shows higher level of anonymity with pseudonyms regardless of the law. In this context, one potential concern is whether or not a user can change their pseudonym at low cost or for free. This is unlikely for users who have to change their user ID (pseudonym), because they would have to cancel and re-subscribe, which is costly. Therefore, we can assume that each pseudonym likely represents a separate participant throughout the period of study.

During the period from January 2005 to December 2010, 2,024,274 postings are collected from Site 1, and the mean of postings per month is 28,114. Mean of participants (nicknames) per month are 2,858 for postings. In Site 2, means of participant (real names) and postings are 774 and 5,304, respectively. Figure 1 displays the trends in the number of monthly postings and comments throughout 2005-2010 in Site 1 and Site 2. A vertical solid line in the middle of the graph is the point of Real Name Verification Law enactment, and the oval around the line is the period of interest in this study. In the long term point-of-view, it is conjectured that the number of postings and comments were more heavily influenced by certain political controversies rather than by the law. The trend of Site 2 is relatively stable over the period, compared to Site 1.

Figure 1. Overall Trend of at Site 1 and Site 2 (2005-2010)



For the individual-level study, user data was selected according to the pre-defined method, and summary statistics is reported in Table 3. There does not seem to be significant differences in Row (1) between treatment and control groups. In Row (2), while the consistency of participation in Site 1 has decreased from .3744 to .3115, that in Site 2 has increased from .4263 to .5388.

Table 3. Descriptive Statistics

(1) Aggressive expression		Obs.	Mean	Std. dev.
y (Site 1)	Treatment	199	.3165	.4663
	Control	378	.3015	.4595
(2) Participation				
y (Site 1)	Treatment	2,545	.3115	.4632
	Control	1,509	.3744	.4841
y (Site 2)	Treatment	670	.5388	.4988
	Control	849	.4263	.4948

5. Results[8]

5.1. The change in using bad postings

[7] The following technical approach is implemented. To run and display content on the browsers, web data such as HTML and JavaScript are downloaded to the client sides. This data stream can be read and its target content can be parsed, a concept known as "crawling and parsing". A web crawler is designed to reconstruct all URLs, also known as the seeds of the target sites, and retrieve content from those pages using Java and MS-SQL. Based on extracted data from the target sites, data mining and filtering methods can be implemented to retrieve postings with specific keywords (selected swear words and anti-normative expressions).

[8] Tables in this section are reported in Appendix.

The main goal in this section is to see how user behavior regarding swear words and anti-normative expressions have shifted after the law. The results from Site 1 are reported in Table 4. First, when it comes to results from the treatment group, the proportion of aggressive postings decreased after the law enforcement in both the short-term and the long-term compared to the control group, and they are statistically significant according to the chi-square test. By contrast, even though changes in aggressive postings in the control group are significant, it appears that swear words and anti-normative expressions surprisingly increased in Period 2. Second, while the proportion of swear words and anti-normative expressions in the Light User group are mostly greater than those in the Middle and Heavy User groups, the actual numbers of aggressive postings in the Heavy User group is greater than those in the Light and Middle User groups across columns. There seems to be two reasons for this finding. On the one hand, a few heavy users post far greater number of postings; therefore, the distribution is highly skewed to this group. On the other hand, participants in the Light User group usually post one or two postings according to their interests, and these postings are more likely to contain aggressive words. Third, behavioral shifts are more saliently observed in the Heavy User group, and this mainly leads to the difference of total proportions between two periods. Finally, results from paired t-test indicate whether the difference in mean of two samples is significantly different or not, and the sign of t-statistics tells which sample's mean is greater. It is observed that all t-statistics from the treatment group are positive and statistically significant, which means that the means of proportion in aggressive postings after the law are smaller than those before the law across all cases.

Table 5 reports the result from Site 2 in which real name postings are allowed regardless of the law. One can clearly see that proportions of aggressive postings are smaller than those in Table 4. Furthermore, there is a slight reduction of aggressive postings in most cases of the treatment group, but these are not salient, compared to the results from the control group. This can be interpreted that the law did not influence users on the real-name board, and they did not change their behaviors on the online forum. Namely, the Real Name Verification Law did not lead to a substantial change in Site 2. To some extent, the findings of this study suggest that the governmental intervention was effective to reduce swear words and anti-normative expressions in an aggregate manner. This result is also consistent to the findings from experimental designs in the previous research that argued people would behave more politely and conformingly under the identifiable condition rather than anonymous condition.

5.2. The change in the number of postings and participants

The result in the previous section might relate to the reduction of postings in itself, so-called willingness-to-express in the online forum, caused by the law. Thus, it is necessary to verify if user participation has been reduced at the same time. To accomplish this task, the number of postings of participants is also measured.[9] Short term (60 days) and long term (6 months) results are reported in Table 7 and Table 8. First, in Site 1, the decrease in participants is not observed after the law, whereas the number of postings is significantly decreased. It might indicate that users still participated in the discussion, but they posted fewer amounts of writings after the law enforcement. However, the number of postings and participants were dramatically increased after the law in the long-term. That is, participation more heavily relies on the importance of issues as shown in Figure 1. This fact is consistent with the mixed results in control group.

Site 2 in Table 8 also shows a similar result. With regards to participants, there is no significant difference after the law. Moreover, total postings even increased after the law; therefore, it appears that users on the real-name board were not concerned with the implementation of the law. The findings suggest that the Real Name Verification Law did not directly lead to discouraging users' willingness-to-

[9] Before showing the results of the short-term and long-term in Site 1 and Site 2, along with the previous section, the change in each 3 weeks (short-term) before and after the Real Name Verification Law is provided in Table 6. While a decrease in average number of daily postings in Site 1 is observed, there is no impact by the law in other cases. The number of postings and participants even increased in Site 2 after the law. As a result, it is difficult to conclude that users' willingness-to-express was discouraged due to the law, and this very short-term result might be biased in some sense.

Table 6. 3-Week Participation

Site 1	Group	Mean		p-value
		Before	After	
Postings	Treatment	462	388	0.017**
	Control	976	1006	0.363
Participants	Treatment	122	117	0.225
	Control	145	165	0.001***
Site 2		Before	After	
Postings	Treatment	139	165	0.004***
	Control	156	158	0.369
Participants	Treatment	87	106	0.001***
	Control	95	95	0.483

Note: * significant at 10%, ** significant at 5%, *** significant at 1%
1. p-value is based on paired t-test.
2. Paired t-test is based on each individual's posting before and after the reference point. (e.g. Real Name Verification Law(July, 27th, 2007 in Treatment group, July 31st, 2006 in Control group)

express, but more concrete exploration would be required.

5.3. The behavioral changes of particular individuals

Aggregation approach is not sufficient, because it is not certain whether a particular individual has actually changed their behaviors or not with regards to participation and uninhibited behaviors. Thus, as supplemental analysis, an individual-based study is conducted and the results are reported in Table 9.

In Column (1), Group 1 (light users) shows a significant reduction of aggressive expressions by 30%, whereas similar postings by users in Group 5 & 6 (heavy users) actually increase after the law. Consistent with this finding, Column (2) shows that heavy users exhibited greater participations after the law, whereas light users who posted one message before the law did not participate in the discussion significantly compared to the control group.

6. Conclusion

This paper analyzed privacy and anonymity issues surrounding the Real Name Verification Law in South Korea. Due to the law, the country's cyberspace has made user identity more traceable and freedom of speech more vulnerable.. The results suggest that identification of postings has significant effects on reducing uninhibited behaviors, suggesting that the Real Name Verification Law encouraged user behavioral changes in the positive direction to some extent and the law was effective only on the pseudonym-based site. Also, the legal enforcement had a dampening effect on overall participation in the short-term at the site with a nickname policy; however, overall, it has become irrelevant to participation. The impact is greater for the Heavy User group than others. Discussion participants with their real names showed more discreet behaviors regardless of the enforcement of the law. It seems that users recognized that the level of anonymity was shifted by the law, from complete dissociation of real and online identities to only visual anonymity by pseudonyms in which their real identity can be detectable.

This study contains some limitations. Only a few websites have been considered for analysis and comparison. Another limitation is the *self-selection* problem. Presumably, some users who are reluctant to show their real identity might use pseudonym-based online discussion forums, and they are more likely to be assertive and abusive. Notwithstanding these limitations, users' behaviors on online boards may be affected by anonymous conditions and topics rather than loosely proclaimed legal enforcement. There are several promising future research agenda. First, future research can be conducted by collecting data from more discussion boards in the given period. People usually have unique frequent-visit websites and participate in the discussion in those spaces. Thus, when more discussion forums are included in the study and current results are still maintained, the argument in this paper becomes highly strengthened and reasonable. Second, this study can be extended to research regarding the impact of real identity policy on social networking websites. In addition to this, a variety of news and media websites have adopted so-called social commenting systems by which the commenter's real identity is automatically disclosed simultaneously. It would be interesting to observe whether user behavior shifts when users post comments with their real personalities and how this result differs from the findings in this study.

References

[1] A. Wood and M. Smith, *Online communication: linking technology, identity, and culture*. Erlbaum Asoociates, 2005.

[2] N. Diakopoulos and M. Naaman, "Towards quality discourse in online news comments," *Proceedings of of the ACM 2011 conference on Computer Supported Cooperative Work*, 2011.

[3] S. Kokolakis, *Privacy perceptions among members of online communities*. Digital Privacy: Theory, Technologies and Practices, 2008, pp. 253–266.

[4] A. Acquisti, R. Dingledine, and P. Syverson, "On the Economics of Anonymity," *Springer-Verlag*, 2003.

[5] L. Sproull, "Reducing Social Context Cues: Electronical Mail in Organizational Communication," *Management Science*, vol. 32, pp. 1492–1512, 1986.

[6] P. Danielson, *Pseudonyms, MailBots, and Virtual Letterheads: The Evolution of Computer-Mediated Ethics*. SUNY Press, 1996.

[7] E. Shin, "Presidential Elections, Internet Politics, and Citizen's Organizations in South Korea." *Development and society*, vol. 34, pp. 25–47, 2005.

[8] D. G. Myers and H. Lamm, "The group polarization phenomenon.," *Psychological Bulletin*, 83, no. 4, pp. 602–627, 1976.

[9] J. B. WALTHER, "Interpersonal Effects in Computer-Mediated Interaction: A Relational Perspective," *Communication Research*, vol. 19, no. 1, pp. 52–90, Feb. 1992.

[10] Y. Hwang, "Critical approach to the implementation of real-name system on bulletin board of the internet: exploring right of anonymous communication," *Media & Society (in Korean)*, vol. 15, pp. 97–130, 2007.

[11] K. Kim, "The technical, legal and business issues of the real-name-verification requirement," *Human Rights and Justice (in Korean)*, 2000.

[12] J. Myung, "Trace of the implemetation of real-name system on the internet bulletin board in a public

institution," *Center for LAw & Information Society (in Korean)*, pp. 2–13, 2003.
[13] J. Woo, H. Nah, and J. Choi, "An Empirical Analysis of the Effect of Real Name System on Internet Bulletin Boards," *South Korea Journal of Public Administration (in Korean)*, vol. 48, pp. 78–96, 2010.
[14] A. Pfitzmann and M. Koehntopp, "Anonymity, unobservability, and pseudonymity—a proposal for terminology," *Springer*, 2001.
[15] I. Goldberg, "A pseudonymous communications infrastructure for the internet," *UC Berkeley Working Paper*, 2000.
[16] A. Froomkin, "Flood Control on the Information Ocean: Living With Anonymity, Digital Cash, and Distributed Databases," *Journal of Law and Commerce*, 15:396–507, 1995.
[17] A. Froomkin, "Anonymity and Its Enmities," *Journal of Online Law*, vol. 4, 1996.
[18] K. Schweizer and J. Leimeister, "The role of virtual communities for the social network of cancer patients," in *Proceedings of the Americas Conference on Information Systems*, 2006.
[19] G. A. Akerlof and R. E. Kranton, "Economics and Identity*," *Quarterly Journal of Economics*, vol. 115, no. 3, pp. 715–753, 2000.
[20] M. Ma and R. Agarwal, "Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities," *Information Systems Research*, vol. 18, pp. 42–67, 2007.
[21] C. Dwyer, "Digital Relationships in the 'MySpace' Generation: Results From a Qualitative Study," in *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, 2007, pp. 19–19.
[22] C. Dwyer and S. Hiltz, "Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace," in *Proceedings of AMCIS*, 2007.
[23] L. Schaupp and L. Carter, "Ethics in Social Networking: A Framework for Evaluating Online Information Disclosure," *HICSS*, 2011.
[24] A. Acquisti, "Information revelation and privacy in online social networks," in *Proceedings of the 2005 ACM workshop on Privacy Enhancing Technologies*, 2005.
[25] J. Siegel, "Group processes in computer-mediated communication*1," *Organizational Behavior and Human Decision Processes*, vol. 37, no. 2, pp. 157–187, Apr. 1986.
[26] T. Zarsky, "Thinking Outside the Box: Considering Transparency, Anonymity, and Pseudonymity as Overall Solutions to the Problems of Privacy in the Internet Society," *University of Miami Law Review*, vol. 58, pp. 1028–1032, 2004.
[27] S. D. Reicher, R. Spears, and T. Postmes, "A Social Identity Model of Deindividuation Phenomena," *European Review of Social Psychology*, vol. 6, pp. 161–198, 1995.
[28] L. Jessup, T. Connolly, and J. Galegher, "The Effects of Anonymity on GDSS Group Process with an Idea-Generating Task," *MIS Quarterly*, vol. 14, 1990.
[29] J. Cohen, "Right to Read Anonymously: A Closer Look at Copyright Management in Cyberspace, A," *Connell Law Review*, vol. 28, p. 981, 1995.
[30] T. Postmes and R. Spears, "Deindividuation and anti-normative behavior: A meta-analysis.," *Psychological Bulletin*, vol. 123, no. 3, pp. 238–259, 1998.
[31] P. L. McLeod, R. S. Baron, M. W. Marti, and K. Yoon, "The eyes have it: Minority influence in face-to-face and computer-mediated group discussion.," *Journal of Applied Psychology*, vol. 82, no. 5, pp. 706–718, 1997.
[32] T. Postmes, R. Spears, and M. Lea, "Intergroup differentiation in computer-mediated communication: Effects of depersonalization.," *Group Dynamics: Theory, Research, and Practice*, vol. 6, no. 1, pp. 3–16, 2002.
[33] J. Wooldridge, *Introductory econometrics: a modern approach*. South-Western Cengage Learning, 2009.

Appendix

Table 4. The change of swear words and anti-normative expressions in Site 1.

Swear words		(1) Short term (60 days)				(2) Long term (6 months)			
		2007 (Treatment group)		2006 (Control group)		2007 (Treatment group)		2006 (Control group)	
Group	# of postings	Before law	After law	Period 1	Period 2	Before law	After law	Period 1	Period 2
Light User	1	195(19.03%)	166(16.06%)	178(23.78%)	210(23.32%)	171(20.91%)	334(17.92%)	128(20.38%)	216(24.14%)
	2	82(12.93%)	77(12.96%)	98(23.28%)	144(31.65%)	84(19.80%)	153(13.43%)	78(23.86%)	122(26.08%)
Middle User	3	71(18.80%)	53(12.88%)	81(25.23%)	105(30.97%)	61(18.20%)	104(12.66%)	42(19.59%)	101(28.29%)
	4 ~ 10	241(13.35%)	176(9.46%)	336(25.23%)	423(28.96%)	247(17.18%)	312(9.28%)	207(19.96%)	406(27.37%)
Heavy User	11 ~ 15	111(14.18%)	96(14.30%)	9(1.46%)	44(5.09%)	108(17.84%)	119(8.93%)	19(3.75%)	31(4.06%)
	16 +	323(2.11%)	150(1.03%)	749(3.58%)	1,049(4.10%)	355(2.33%)	330(1.01%)	421(3.20%)	1,068(4.64%)
Total		1,022(5.16%)	717(3.71%)	1,451(5.95%)	1,976(6.68%)	1,025(5.58%)	1,351(3.26%)	1,351(3.26%)	1,944(7.21%)
Statistics		χ^2=26.98***, p-value=0.000, df=5 Paired t-stat:2.32**, p-value=0.033		χ^2=17.93***, p-value=0.003, df=5 Paired t-stat: -2.69**, p-value=0.021		χ^2=26.35***, p-value=0.000, df=5 Paired t-stat: 4.69***, p-value=0.002		χ^2=20.32***, p-value=0.001, df=5 Paired t-stat: -2.87**, p-value=0.017	
Anti-normative expressions		**(1) Short term (60 days)**				**(2) Long term (6 months)**			
		2007 (Treatment group)		2006 (Control group)		2007 (Treatment group)		2006 (Control group)	
Light User	1	544(53.10%)	528(51.46%)	460(61.46%)	507(58.77%)	533(53.03%)	925(56.20%)	362(57.58%)	525(58.66%)
	2	296(60.91%)	302(51.80%)	243(57.72%)	293(64.40%)	296(55.79%)	557(52.22%)	187(57.06%)	286(61.05%)
Middle User	3	201(54.80%)	230(54.64%)	148(46.26%)	219(64.60%)	204(54.85%)	409(48.57%)	142(65.44%)	286(55.60%)
	4 ~ 10	847(46.88%)	881(44.13%)	843(63.33%)	1,083(74.09%)	847(49.59%)	1,604(40.54%)	608(58.48%)	968(65.27%)
Heavy User	11 ~ 15	387(49.43%)	344(42.96%)	437(70.85%)	505(58.47%)	387(51.08%)	631(35.62%)	294(57.72%)	462(60.50%)
	16 +	3,894(25.42%)	3,090(21.23%)	4,974(23.76%)	7,285(28.48%)	4,390(23.90%)	5,091(15.86%)	2,950(22.41%)	6,194(26.91%)
Total		6,173(31.19%)	5,374(22.73%)	7,106(29.16%)	9,893(33.47%)	6,656(29.27%)	9,215(22.27%)	4,543(28.60%)	8,721(32.00%)
Statistics		χ^2=72.00***, p-value=0.000, df=5 Paired t-stat:3.01**, p-value=0.014		χ^2=35.11***, p-value=0.000, df=5 Paired t-stat:-0.97, p-value=0.188		χ^2=6.03, p-value=0.196, df=5 Paired t-stat:2.58**, p-value=0.024		χ^2=7.32, p-value=0.119, df=5 Paired t-stat:-0.64, p-value=0.273	

* significant at 10%, ** significant at 5%, *** significant at 1%

Table 5. The change of swear words and anti-normative expressions in Site 2.

Swear words		(1) Short term (60 days)				(2) Long term (6 months)			
		2007 (Treatment group)		2006 (Control group)		2007 (Treatment group)		2006 (Control group)	
Group	# of postings	Before law	After law	Period 1	Period 2	Before law	After law	Period 1	Period 2
Light User	1	71(22.38%)	70(21.81%)	78(15.52%)	82(15.35%)	65(20.37%)	70(22.75%)	58(13.01%)	75(16.91%)
	2	51(27.72%)	50(23.81%)	50(22.22%)	42(16.94%)	42(26.94%)	43(23.44%)	34(18.80%)	46(21.63%)
Middle User	3	36(24.24%)	35(25.00%)	41(24.55%)	29(16.52%)	30(24.00%)	35(24.16%)	25(20.33%)	30(18.97%)
	4 ~ 10	75(9.25%)	61(6.74%)	98(14.51%)	116(11.57%)	79(12.83%)	62(7.88%)	76(13.68%)	104(13.90%)
Heavy User	11 ~ 15	0(0.00%)	7(1.74%)	14(3.75%)	0(0.00%)	6(1.67%)	15(3.86%)	12(4.41%)	9(2.44%)
	16 +	0(0.00%)	0(0.00%)	22(0.88%)	11(0.34%)	0(0.00%)	8(0.28%)	14(0.72%)	9(0.34%)
Total		233(4.66%)	222(4.71%)	301(6.88%)	278(5.17%)	222(5.18%)	233(4.98%)	217(6.32%)	272(5.92%)
Statistics		χ^2=8.25*, p-value=0.082, df=5 Paired t-stat:0.87, p-value=0.211		χ^2=21.15***, p-value=0.000, df=5 Paired t-stat:2.85**, p-value=0.017		χ^2=14.23***, p-value=0.000, df=5 Paired t-stat:0.46, p-value=0.330		χ^2=4.42, p-value=0.489, df=5 Paired t-stat:-0.565 p-value=0.298	
Anti-normative expressions		(1) Short term (60 days)				(2) Long term (6 months)			
		2007 (Treatment group)		2006 (Control group)		2007 (Treatment group)		2006 (Control group)	
Light User	1	113(35.87%)	115(35.83%)	127(25.43%)	151(28.44%)	98(30.71%)	124(39.70%)	115(26.05%)	130(29.55%)
	2	91(49.46%)	85(40.48%)	91(40.44%)	88(35.48%)	68(43.46%)	87(46.52%)	72(40.60%)	78(36.68%)
Middle User	3	49(33.33%)	75(54.35%)	69(41.82%)	76(44.35%)	61(52.79%)	67(46.08%)	56(45.93%)	70(45.34%)
	4 ~ 10	326(40.28%)	286(31.67%)	295(43.67%)	230(23.08%)	280(49.84%)	312(37.12%)	225(40.89%)	249(33.44%)
Heavy User	11 ~ 15	118(29.17%)	117(31.46%)	99(27.64%)	136(35.79%)	113(32.39%)	248(9.04%)	76(28.11%)	127(36.06%)
	16 +	243(7.77%)	154(5.58%)	265(10.80%)	277(9.11%)	119(29.88%)	214(7.47%)	162(8.73%)	302(11.29%)
Total		941(18.85%)	833(17.67%)	946(21.62%)	959(17.84%)	871(20.46%)	948(19.89%)	709(20.66%)	958(20.89%)
Statistics		χ^2=21.74***, p-value=0.000, df=5 Paired t-stat:-0.129, p-value=0.450		χ^2=16.52*** p-value=0.005, df=5 Paired t-stat:-0.55, p-value=0.302		χ^2=31.22***, p-value=0.000, df=5 Paired t-stat:1.64*, p-value=0.080		χ^2=21.87*** p-value=0.001, df=5 Paired t-stat:-0.15, p-value=0.442	

* significant at 10%, ** significant at 5%, *** significant at 1%

Note for Table 4 and Table 5
1. Each value is the average number per month. The percentage value in parenthesis is the proportion out of total postings in each group.
2. Chi-square test is based on average number per month. Paired t-test is based on the percentage value.
3. Paired t-test is based on each individual's posting before and after the reference point. (e.g. Real Name Verification Law(July, 27th, 2007 in Treatment group, July 31st, 2006 in Control group)

Table 7. The change of participation in Site 1.

Participants		(1) Short term (60 days)				(2) Long term (6 months)			
		2007 (Treatment group)		2006 (Control group)		2007 (Treatment group)		2006 (Control group)	
Group	# of postings	Before law	After law	Period 1	Period 2	Before law	After law	Period 1	Period 2
Light User	1	1,026(52.80%)	962(51.38%)	624(50.57%)	858(47.26%)	839(51.60%)	1,775(45.86%)	630(51.03%)	897(50.07%)
	2	251(12.93%)	265(14.15%)	169(13.65%)	243(13.39%)	219(13.44%)	572(14.77%)	165(13.30%)	235(13.10%)
Middle User	3	128(6.60%)	128(6.85%)	78(6.32%)	129(7.11%)	109(6.72%)	300(7.75%)	73(5.86%)	119(6.65%)
	4 ~ 10	291(14.98%)	281(15.02%)	173(13.98%)	250(13.75%)	239(14.67%)	649(16.76%)	176(14.26%)	243(13.59%)
Heavy User	11 ~ 15	59(3.04%)	60(3.19%)	38(3.04%)	64(3.53%)	51(3.13%)	146(3.76%)	41(3.28%)	60(3.34%)
	16 +	187(9.65%)	176(9.42%)	154(12.44%)	272(14.96%)	170(10.44%)	429(11.10%)	152(12.26%)	238(13.25%)
Total		1,943(100%)	1,871(100%)	1,234(100%)	3,629(100%)	1,626(100%)	3,870(100%)	1,235(100%)	1,791(100%)
Statistics		χ^2=1.67, p-value=0.892, df=5 Paired t-stat:-0.004, p-value=0.498		χ^2=6.13, p-value=0.293, df=5 Paired t-stat:0.00, p-value=0.500		χ^2=15.72***, p-value=0.007, df=5 Paired t-stat:0.004, p-value=0.498		χ^2=1.56 p-value=0.904, df=5 Paired t-stat:0.00, p-value=0.500	
Postings		(1) Short term (60 days)				(2) Long term (6 months)			
		2007 (Treatment group)		2006 (Control group)		2007 (Treatment group)		2006 (Control group)	
Light User	1	1,026(5.11%)	962(5.30%)	624(3.55%)	858(2.83%)	839(4.59%)	1,775(3.99%)	630(3.96%)	897(3.32%)
	2	502(2.50%)	530(2.92%)	337(1.92%)	486(1.60%)	438(2.39%)	1,143(2.57%)	329(2.07%)	470(1.74%)
Middle User	3	384(1.92%)	384(2.12%)	234(1.33%)	387(1.28%)	328(1.80%)	900(2.20%)	217(1.37%)	357(1.32%)
	4 ~ 10	1,758(8.76%)	1,682(9.28%)	1,009(5.74%)	1,525(5.03%)	1,430(7.83%)	3,894(8.76%)	1,042(6.55%)	1,485(5.50%)
Heavy User	11 ~ 15	809(4.03%)	700(3.86%)	468(2.66%)	821(2.71%)	645(3.53%)	1,859(4.18%)	511(3.21%)	765(2.83%)
	16 +	15,603(77.69%)	13,875(76.52%)	14,890(84.79%)	26,213(86.54%)	14,591(79.86%)	34,903(78.48%)	13,168(82.84%)	23,016(85.29%)
Total		20,083(100%)	18,132(100%)	17,561(100%)	30,289(100%)	18,269(100%)	44,473(100%)	15,895(100%)	26,987(100%)
Statistics		χ^2=14.30***, p-value=0.013, df=5 Paired t-stat: -1.22, p-value=0.138		χ^2=64.02***, p-value=0.000, df=5 Paired t-stat:0.00, p-value=0.500		χ^2=46.38***, p-value=0.000, df=5 Paired t-stat:0.00, p-value=0.500		χ^2=31.34***, p-value=0.000, df=5 Paired t-stat:0.02, p-value=0.491	

* significant at 10%, ** significant at 5%, *** significant at 1%

Table 8. The change of participation in Site 2

Participants		(1) Short term (60 days)				(2) Long term (6 months)			
		2007 (Treatment group)		2006 (Control group)		2007 (Treatment group)		2006 (Control group)	
Group	# of postings	Before law	After law	Period 1	Period 2	Before law	After law	Period 1	Period 2
Light User	1	315(44.87%)	321(44.52%)	613(62.85%)	543(54.75%)	320(50.38%)	314(44.70%)	444(59.80%)	441(53.46%)
	2	92(13.11%)	105(14.56%)	114(11.65%)	120(12.02%)	79(12.45%)	94(13.34%)	89(11.96%)	107(12.90%)
Middle User	3	50(7.05%)	46(6.38%)	54(5.54%)	60(6.03%)	39(6.12%)	49(6.97%)	41(5.53%)	52(6.29%)
	4 ~ 10	129(18.38%)	142(19.69%)	107(10.98%)	170(17.04%)	94(14.79%)	134(19.05%)	93(12.54%)	125(15.07%)
Heavy User	11 ~ 15	32(4.56%)	30(4.09%)	24(2.46%)	27(2.66%)	28(4.33%)	32(4.47%)	22(2.90%)	28(3.32%)
	16 +	85(12.04%)	78(10.75%)	64(6.52%)	75(7.49%)	76(11.93%)	81(11.47%)	54(7.26%)	74(8.96%)
Total		702(100%)	721(100%)	975(100%)	995(100%)	635(100%)	701(100%)	742(100%)	824(100%)
Statistics		χ^2=1.81, p-value=0.873, df=5 Paired t-stat:0.007, p-value=0.497		χ^2=19.90** p-value=0.000, df=5 Paired t-stat:0.001, p-value=0.499		χ^2=6.50, p-value=0.260, df=5 Paired t-stat:0.00, p-value=0.500		χ^2=7.032, p-value=0.218, df=5 Paired t-stat:-0.001, p-value=0.499	
Postings		(1) Short term (60 days)				(2) Long term (6 months)			
		2007 (Treatment group)		2006 (Control group)		2007 (Treatment group)		2006 (Control group)	
Light User	1	500(11.41%)	531(9.87%)	613(15.07%)	545(10.78%)	320(7.51%)	313(6.57%)	444(12.92%)	441(9.60%)
	2	225(5.14%)	248(4.61%)	227(5.58%)	239(4.73%)	158(3.71%)	187(3.92%)	174(5.17%)	218(4.63%)
Middle User	3	165(3.77%)	173(3.21%)	162(3.99%)	180(3.56%)	117(2.74%)	147(3.07%)	123(3.58%)	156(3.39%)
	4 ~ 10	678(15.43%)	999(18.57%)	629(15.48%)	1,030(20.38%)	563(13.23%)	843(17.67%)	552(16.08%)	743(16.28%)
Heavy User	11 ~ 15	360(8.22%)	380(7.07%)	306(7.53%)	347(6.86%)	352(8.25%)	399(8.36%)	273(7.93%)	355(7.72%)
	16 +	2,453(56.03%)	3,048(56.67%)	2,128(52.36%)	2,713(53.69%)	2,750(64.56%)	2,881(60.41%)	1,865(54.32%)	2,679(58.37%)
Total		4,378(100%)	5,378(100%)	4,065(100%)	5,052(100%)	4,259(100%)	4,769(100%)	3,432(100%)	4,590(100%)
Statistics		χ^2=26.81***, p-value=0.000, df=5 Paired t-stat:0.00, p-value=0.500		χ^2=32.37***, p-value=0.000, df=5 Paired t-stat:0.001, p-value=0.499		χ^2=4.97 p-value=0.289, df=5 Paired t-stat:0.00, p-value=0.500		χ^2=26.08*** p-value=0.000, df=5 Paired t-stat:0.00, p-value=0.500	

* significant at 10%, ** significant at 5%, *** significant at 1%

Note for Table 7 and Table 8:
1. Each value is the average number per month. The percentage value in parenthesis is the proportion out of total value in the last row.
2. Chi-square test is based on average number per month. Paired t-test is based on the percentage value.
3. Paired t-test is based on each individual's posting before and after the reference point. (e.g. Real Name Verification Law(July, 27th, 2007 in Treatment group, July 31st, 2006 in Control group)

Table 9. Result: Individual-level analysis

DV	Aggressive expressions	Participation	
	(1) Site 1 (N=577)	(2) Site 1 (N=4,054)	(3) Site 2 (N=1,519)
dTreat	-0.301*** (0.088)	-0.158*** (0.016)	-0.154*** (0.031)
dTreat · dGroup2	0.000 (0.144)	0.079*** (0.027)	0.196*** (0.057)
dTreat · dGroup3	0.000 (0.197)	0.117*** (0.037)	0.260*** (0.077)
dTreat · dGroup4	0.048 (0.109)	0.196*** (0.028)	0.554*** (0.044)
dTreat · dGroup5	0.312** (0.138)	0.237*** (0.058)	0.694*** (0.042)
dTreat · dGroup6	0.583*** (0.096)	0.489*** (0.032)	0.656*** (0.037)
Constant	0.301*** (0.022)	0.374*** (0.012)	0.426*** (0.017)
R-squared	0.1082	0.0625	0.1460

* significant at 10%, ** significant at 5%, *** significant at 1%,

Note:
1. Robust standard error is in parentheses.
2. Paired t-test is based on each individual's posting before and after the reference point. (e.g. Real Name Verification Law(July, 27th, 2007 in Treatment group, July 31st, 2006 in Control group)
3. +: This is not reported, because all users who posted with swear words before the law did not write postings with those uninhibited words after the law. Namely the mean before the is 1 without the variation, and the mean after the law is 0.