Computers in Human Behavior 51 (2015) 363–372

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

The impacts of identity verification and disclosure of social cues on flaming in online user comments

Daegon Cho [a,*], K. Hazel Kwon [b,1]

[a] Pohang University of Science and Technology (POSTECH), Dept. of Industrial and Management Engineering, 77 Cheongam-ro, Nam-gu, Pohang, Gyeongbuk 790-784, South Korea
[b] Arizona State University, School of Social and Behavioral Sciences, 4701 W. Thunderbird Rd. MC3051, Phoenix, AZ 85306, USA

ARTICLE INFO

Article history:

Keywords:
Disinhibition
Anonymity
Flaming
Online public discussions
Online comments
Profanity

ABSTRACT

While a growing body of literature attests to the relationship between user identifiability and inflammatory speech online, few studies have investigated the ways in which different anonymity control mechanisms affect the quality of online discussions. In this study, two mechanisms, a policy-driven and a voluntary approach, are examined for their conditional and interaction effects on reducing flaming in user comments online. Based on a large-scale, real-world data on political news comments in South Korea, the results suggest that whereas the policy-driven regulation does not reduce, and even increases, flaming, the voluntary approach significantly decreases it, especially among the moderate commenters. The findings are further speculated from an economic perspective by which transaction costs are perceived differently contingent on the ways in which anonymous commenting is regulated.

1. Introduction

The prevalent use of the Internet for political discussions has brought forth the question of how online anonymity affects the quality of public deliberation. On the one hand, scholars argue for constructive roles of anonymity in increasing a sense of equality, reducing normative pressure on conformity, and encouraging users to exert individual rights for free speech in a liberated manner (Dahlgren, 2005; Papacharissi, 2004; Ruiz et al., 2011). On the other hand, skeptics point out the detrimental effects of anonymity on aggravating flaming and trolling in online social interactions (Kushin & Kitchener, 2009; Lampe, Zube, Lee, Park, & Johnston, 2014). Both positive and negative perspectives are rooted in the same tenet that the degree of identification can make a difference in user behaviors.

As Herring et al. (2002) state, acrimonious comments, "while clearly problematic, are nonetheless widespread and often tolerated, due in part to the pervasiveness on the Internet of civil libertarian values that consider abusive speech a manifestation of individual freedom of expression" (p. 372). The ease with which the Internet anonymity may induce flaming and trolling has become a non-negligible issue especially along with growing concerns about civility crisis in contemporary politics (Mutz & Reeves, 2005). Inflammatory political campaigns, for example, have become pandemic in today's news headlines and have been found to delegitimize opposing point of views and political processes (Brooks & Geer, 2007), and decrease citizens' political trust (Mutz & Reeves, 2005).

While some scholars may argue that online profanity may not be as harmful as one might assume (Canter, 2013; Papacharissi, 2004), a recent survey indicates that the Internet and social media were cited as one of the leading causes of incivility (Weber, 2013). Coe, Kenski, and Rains (2014) also suggest the negative implication of online profanity by showing that the more uncivil comments (among which vulgarity and name-calling were subset) online users are exposed to, the greater they become reluctant to engage in discursive interactions. A study on Youtube flaming similarly suggests that the majority of users perceive flaming "annoying" and "[deviating from] an honest way of expressing disagreement" (Moor, Heuvelman, & Verleur, 2010, p. 1542). Inasmuch as negative responses elicited by flaming are commonplace, it may be a valid claim that flaming in online comments could pose a danger to the conditions for deliberative citizen-to-citizen discussions.

The relationship between flaming and online anonymity has drawn even more scholarly attention along with anonymity regulatory gestures recently put into high gear. Such gestures range from promoting voluntary self-disclosures, to crowdsourcing moderation, and to mandating identity verification processes. In particular, the current study distinguishes two qualitatively different mechanisms, one of which pertains to a policy-driven regulation and the other to a voluntary disclosure of social cues, and explore the extent to which each mechanism contributes to the reduction

* Corresponding author. Tel.: +82 54 279 2375; fax: +82 54 279 2870.
E-mail addresses: dgcho@postech.ac.kr (D. Cho), khkwon@asu.edu (K.H. Kwon).
[1] Tel.: +1 602 543 5676; fax: +1 602 543 6004.

http://dx.doi.org/10.1016/j.chb.2015.04.046
0747-5632/

of spiteful comments online. For empirical exploration, we leverage large, real-world data on news website comments in South Korea where the Internet Identity Verification Law was once enforced nationwide (Cho & Kim, 2012). While we focus on South Korea because its national law aptly exemplifies a policy-driven, top-down approach in clear contrast to a voluntary approach, we believe that this study should provide some generalizable conclusions resonating with the intensified debates surrounding anonymity regulations across different national and cultural contexts.

The contribution of this study is twofold. First, while extant literature has examined the relationship between the Internet anonymity and flaming, few have differentiated the effects of dissimilar anonymity control mechanisms. Thus, our study may provide additional theoretical as well as managerial insights for appropriating different regulatory systems, when needed. Second, whereas the related studies demonstrate analyses on a single site or company (e.g. Coe et al., 2014; Lampe et al., 2014; Moor et al., 2010; Rowe, 2014), we leverage a large dataset beyond a particular platform: The analyses cover online comments collected from 26 different news websites, and the focus is on election campaign periods during which exchanges of political opinions become increasingly crucial for electorates' decision making.

2. Literature reviews

2.1. Anonymity and flaming in online comments

Although widespread use of online forums has facilitated plural political debates, anonymity has been linked to the risk of depreciatory debate culture. Despite positive roles of the anonymous Internet such as facilitating free speech without the fear of social disapproval and even physical arrestment in sensitive political contexts, anonymity has nevertheless provided leeway for spamming, hate speech, deception and impersonation (Kling, Lee, Teich, & Frankel, 1999). Moreover, flaming and trolling have recently become increasingly worrisome (Herring, Job-Sluder, Scheckler, & Barab, 2002; Lampe et al., 2014).

In particular, flaming refers to a message sender's hostile emotional expressions characterized by using insulting, profane, or offensive languages, which may "inflict harm to a person or an organization resulting from uninhibited behavior" (Alonzo & Aiken, 2004, p. 205). On occasions, flaming may be tolerable and even liberating message senders by an emotional release. However, every flaming has a potential to negatively affect message receivers: Although a sender might flame for a harmless purpose, the message can be perceived offensive depending on who reads it (Moor et al., 2010). Empirical findings have shown that flaming can induce negatively emotional effects: Johnson, Cooper, and Chin (2009), for example, suggest that flaming downplay computer-mediated negotiation processes by invoking angers. Similarly, Coyne, Stockdale, Nelson, and Fraser (2011) find that media exposure to profanity, including the exposure to the Internet content, induces more frequent uses of profane languages, which subsequently increase physical and relational aggressive behaviors among adolescents. Not just having individual psychological and emotional effects, flaming may also play an unconstructive role in shaping online discussion culture: Verbal attacks can become reciprocal, creating "flaming norms" in an online community (Lea, O'Shea, Fung, & Spears, 1992; Moor et al., 2010, p. 1542). For example, Moor et al. (2010) contend that, despite the consensus that flaming is negative among users, inflammatory comments have become increasingly normative in Youtube and such flaming norms cause some users' withdrawal from content-generating activities.

The underlying rationales behind flaming are drawn from the literature of anonymity effect on disinhibition. According to Scott (1998), the concept of anonymity is a rich, multi-dimensional concept, generally defined as the condition to which a message source's true identity is absent or unacknowledged. The discussions of anonymity can take a different direction depending on whose perception is at the center of inquiry: self-anonymity refers to a sender's perceived anonymity to others, and other-anonymity pertains to the anonymity a receiver experiences during an interaction with unidentified source (Scott, 1998). This study focuses on self-anonymity, also defined as identifiability, because it directly plays a role in determining the message sender's own behaviors (Spears & Lea, 1994). Accordingly, the concept 'anonymity' implies self-anonymity hereupon and is used interchangeably with the term identifiability. Self-anonymity effect on disinhibition is based on the set of long-lasting theories on crowd behaviors and social conformity (e.g. Festinger, Pepitone, & Newcomb, 1952; McPhail, 1991; Postmes & Spears, 1998; Zimbardo, 1969). The simplest description of the rationale is: The less identifiable, the more disinhibited. Festinger et al. (1952) and Zimbardo (1969) introduce a widely known concept called "de-individuation" to refer to the extra- or anti-normative behaviors induced from the lack of personalized social cues.

Individuals become uninhibited because of the absence of accountability. On a positive note, being free from the real-self means less constraints, thus greater freedom and outspokenness during discussion interactions (Jessup, Connolly, & Galegher, 1990; Zarsky, 2003). The downside, however, is that this tendency can be translated into ill mannerism. Not surprisingly, negative online activities such as defamation of character, threats, and slanderous comments have often been found problematic and become a concern among law and policymakers (Cohen, 1995; Sunstein, 2014). A majority of empirical research on de-individuation has indeed highlighted negative effects, particularly in relation to aggressive, anti-normative behaviors (Christopherson, 2007). In social media environment, different platforms are characterized with different levels of affordances for identifiability and social cues, which may have disproportionate effects on user disinhibition. For example, Halpern and Gibbs (2013) compare Facebook and Youtube and find that Youtube contains more impolite comments due to its lower identifiability and networked information access than Facebook.

The issue of flaming has also been linked to the social identity model of de-individuation effects (SIDE). Among various CMC theories, this model is particularly suitable for analyzing the exchange of commenting behaviors in the online discussion platforms where "one-off, non-interactive, pseudonymous messages are common" (Walther, DeAndrea, Kim, & Anthony, 2010, p. 473). Specifically, the SIDE model proposes that reduced social cues in online settings may actually facilitate in-group normative behaviors by allowing communicators to treat one another equally as a member of the online group (Sproull & Kielser, 1991) and trivializing individuated unequal social status (Dubrovsky, Kielser, & Sethna, 1991). De-personalization and increased conformity to a group norm, which are the core dynamics of SIDE, occur when the process is "coupled with a salience of common (group) identity" (Lee, 2004, p. 235; Postmes, Spears, Sakhel, & de Groot, 2001). While SIDE may result in a stronger group cohesion and solidarity, the positive consequence is not always the case. Indeed, some experimental studies challenge the equalization hypothesis, claiming that anonymous group communication online does not significantly increase equality among communicators of different status (Coffey & Woolworth, 2004; Connolly, Jessup, & Valacich, 1990; Hollingshead, 1996; Straus, 1997). Moreover, scholars even contend that SIDE effect aggravates the tendency for group polarization (Sia, Tan, & Wei, 2002). For example, Lee's (2007) study

shows that the increased group identity under anonymous condition induces subjects' opinions to be more polarized. When flaming becomes a normative behavior within an online community, the SIDE suggests that users motivation to conform to group norm can result in more aggravated flaming behaviors (Lea, O'Shea, Fung, & Spears, 1992).

In sum, de-individuation thesis and SIDE model suggest that anonymous conditions affect the ways in which individuals behave. Although anonymity may help nurture a positive atmosphere of open mindedness in certain situations, the same condition can likewise result in negative consequences in a different context. Flaming is one aspect of online disinhibitions, which could impose negative psychological effects as well as depreciate the climate of online public discussions.

2.2. Regulating anonymity: Policy-driven versus voluntary approaches

As anti-normative behaviors such as flaming, trolling, and cyber-threats have become commonplace online, growing attention is being paid toward solutions that can increase user identifiability and accountability. The scope of anonymity regulating attempts vary in numerous ways: For example, many websites encourage users to voluntarily link their social networking site (SNS) profiles to the activities in the respective website (Rowe, 2014; Soni, 2013); Media companies often stipulate a policy that requires users to provide real names during the registration, which often triggers intricate issues on privacy and trust (Gross & Acquisti, 2005; Stephen & Galak, 2012; Youmans & York, 2012); Furthermore, some state governments and countries take more aggressive approaches by mandating laws that make users' real identities more easily traceable (Cho & Acquisti, 2013; Cho & Kim, 2012).

Such various attempts to control anonymity can be roughly categorized into two approaches: Policy-driven regulations, and promotion of voluntary disclosure of social cues. (We do not suggest, however, that the two approaches represent comprehensive and mutually exclusive list of anonymity control mechanisms.) Policy-driven regulations refer to enforcing formal rules that compels users to be accountable for their online behaviors. For example, companies like The Wall Street Journal and Facebook[2] require users to follow the real name policy, and theoretically have the authority to block any users who do not comply with the policy. On a state or national level, the government may either require web service providers to collect users' personal information, as in the case of South Korea and China, or endorse law enforcement authorities to gain access to the user data relatively easily, as exemplified by the recent legislature in Arizona (Newman, April 3, 2012). The main purpose of policy-driven approaches is to increase traceability of real identity: Even if users are allowed to use pseudo-names online, users may not perceive self-anonymity at least to surveillors of the website since each pseudo-name is translatable into true identity based on personal information provided (Kling et al., 1999). The premise behind implementing a policy-driven approach as a solution for online disinhibition is that the identity traceability would provoke an awareness of liability, which would in turn prevent users from irresponsible actions online.

Alternatively, voluntary approach promotes a culture of self-disclosure by taking advantage of recent social commenting technologies (Wang, 2013). Conventionally, the use of social commenting system is offered as a more efficient and convenient option for users when they log into a respective website: Users can skip all the sign-up steps, and immediately log into the website by simply clicking a social plug-in that automatically connects the activities on the website to an existing SNS account (Rowe, 2014). As a result, identity-disclosing social cues displayed in the SNS profiles become publicly visible on the website as well as the users' comments, which can likewise be shared by social contacts networked in the SNS. Although such content fluidity is likely to pose privacy concerns and peer monitoring risks (Humphreys, 2011; Rainie & Wellman, 2012), those who voluntarily opt in on the use of social plug-ins tend to give more weight toward the immediate benefits (e.g. convenience, ease of use, and time saving) over the cost of losing privacy (Smith, Dinev, & Xu, 2011). The underlying mechanism behind the voluntary disclosure approach is to foster "public self-awareness," which refers to concerns about one's appearances and images constructed in social situations (Lee, 2007). The premise behind social commenting approach as the remedy for online disinhibition is that public self-awareness evokes human desire for positive self-presentation. Motivated by social incentives associated with positive display of self, users should practice mindfulness when commenting in the networked environment.

In sum, various anonymity control mechanisms have been widely implemented as an attempt to reduce anti-normative messaging behaviors online. While policy-based and voluntary approaches both aim for the common result – to reduce negative behaviors – they each trigger subtly different sense of self-anonymity: Policy-driven approaches activate a perception that the self is not anonymous to the top-down monitoring body. Such self-awareness may induce a sense of liability and obligation, reminding users that they can be penalized for irresponsible actions online. Alternatively, voluntary approaches accentuate the identifiability to peer readers and evoke a social motives for positive impression management, encouraging users to maintain socially acceptable images online while avoiding irresponsible actions that will compromise self-representation in the face of increasing peer monitoring and subsequent social sanctions.

3. Research background and hypotheses

3.1. Research background

In order to reduce flaming and trolling from discussion culture on the Internet, the South Korean National Assembly enacted the Identity Verification Law in 2005, which was applicable to all political comments made on major websites. This regulation required service providers operating message boards on political websites to permit users to post comments only after they had successfully completed the identity verification process (Park & Greenleaf, 2012). Although another moniker of this law was a "real-name" policy, the law did not explicate that users' real names must be publicized on a discussion forum. Therefore, the use of pseudonym was allowed as long as the pseudonym is traceable to the offline real identity. In December 2011, the regulatory agency officially announced that the law would be repealed in 2012 due to a possible harm and chilling effect on free speech (Park & Greenleaf, 2012). Moreover, the court's ruling addressed that given the advent of new means of communication including mobile messaging and social networking sites (SNS), the identity verification law would only serve limited public interest across a narrow range of the Internet space. Consequently, the Constitutional Court overturned the law in August 2012.

History, as described above, indicates that General Election in April 2012 was held during the transitional period preceding the expiration of the law. Although the law was still in effect during the two-week campaign period, some websites decided to no longer require users to undergo the verification process. Given the random variation in the compliance level among websites during this period, the division among compliant and non-compliant websites was thought to qualify as the comparison groups. Set

[2] Facebook recently shows the action to loosen the real name policy. See the report in TechCrunch (January 30th, 2014) at http://www.techcrunch.com/2014/01/30/facebook-will-give-up-the-ghost-on-real-id-in-future-apps/.

aside the studies by Cho & Kim, 2012; Cho & Acquisti, 2013 that demonstrated short-term effectiveness of the policy based on two million comments across topical areas, little evidence exists regarding the policy effect, in particular during a political election campaign period. The investigation of political campaign period is of particular interest because the top-down monitoring via identity verification system primarily aimed to deter Internet users from political flaming, with the assumption that such messaging should harm the fairness in voters' decision-making.

3.2. Research hypotheses

The following hypotheses are developed to evaluate the influence of each anonymity control mechanism on user commenting behaviors. In particular, we focus on the likelihood of flaming as a dependent variable. The contrast between the policy-based regulation and the voluntary disclosure of social cues is at the center of our examination. As previously mentioned, both approaches aim to decrease anti-normative behaviors on the Internet. In particular, policy-driven regulations may effectively control the level of anonymity by increasing the traceability of real identity and subsequently increasing user accountability. Indeed, the intervention to change disinhibited behaviors via increased accountability is one of the main agendas for many existing Internet policies. Therefore, our first hypothesis is posited as follows:

H1. Formal policy regulations induce a sense of identifiability such that commenters under the policy regulations are less likely to write inflammatory comments than commenters who are not under the policy regulations.

As for the effects of voluntary approaches, consider that social motives are influenced by an individual's belief on how he or she is perceived by others. Comments that present the individual as sociable, fair-minded, or caring may yield a positive image, whereas comments that present the individual as unfair, aggressive, or greedy may not only reduce the positive image but even produce a negative image (Bénabou & Tirole, 2006). Thus, when social cues are attached to the comments being posted, each individual's identifiable social image is at stake depending on the contents of the comments. Consequently, settings where social images are at stake may encourage commenters to avoid anti-normative behaviors. This leads to our second hypothesis:

H2. Commenters who opt in on the voluntary disclosure of social cues (i.e. social commenting) are less likely to write inflammatory comments than commenters who did not choose a social commenting option.

Lastly, hypotheses H1 and H2 are combined to assess how the two suggested anonymity control mechanisms concurrently affect commenting behaviors. The online commenting conditions in which this study is conducted, described in further detail below, allow us to test the effects of these two approaches together. This hypothesis is formulated as follows:

H3. Commenters who use social commenting system under formal policy regulations are even less likely to make inflammatory comments than commenters who do not use social commenting system under formal policy regulations.

4. Research design

4.1. Research setting

This study is based on the two-week General Election campaign period in April 2012 in South Korea, which provides a natural, real-world setting where the two aforementioned mechanisms had co-existed to a varying degree across news websites. The so-called Identity Verification Law in South Korea had become the target for debates around the world as an exemplar of anti-anonymity governmental regulations.[3] In particular, concerns were associated with its efficacy on reducing trolling and flaming on the Internet. In addition to the identity verification law in practice, many South Korean websites had adopted a social commenting system, which allowed users to log into the website via their SNS accounts. Data was acquired from the third-party company that continues to provide domestic news websites with a social commenting system. This firm is the largest company of its kind in South Korea, serving major media organizations since 2010. Thus, the study context demonstrates an appropriate real-world setting with concurrence of the two anonymity control mechanisms.

Fig. 1 presents the 2×2 research design setting. The first factor, which measures whether or not commenters are members of the news sites compliant with the law, represents a policy-driven anonymity control mechanism. This factor categorizes sample commenters into two groups: commenter from the law-compliant websites and those from non-compliant websites. When users wish to share comments on the law-compliant websites, they must verify their real identity with a real name, Residential Registration Number (a national identification number), birthday, home address and phone number. In contrast, users are not required to undergo such processes prior to posting messages on the non-compliant websites. Note that the verification process does not indicate that a commenter's real name must be publicly displayed. Rather, personal information is saved in the website operators' servers in order to enhance the identity traceability and promote a sense of liability. The second factor represents the voluntary approach toward anonymity control. All of the examined websites offer users an option of whether or not to use a social commenting system. The amount of disclosed social cues differs based on the user preference: Opting-in social commenting implies an agreement to disclose more social cues, in particular by connecting a SNS profile with one's commenting behaviors. This process is consistent with the voluntary anonymity control mechanism that encourages self-disclosure of social cues. Therefore, the second factor categorizes sample commenters into two groups: whether or not to opt-in a social commenting system.

4.2. Data

A rich data set of online comments was collected in assistance from the largest third-party social commenting provider in South Korea. Our dataset contains a total of 13,219 political comments written by 5753 commenters threaded under the articles in the Politics and Election sections of 26 major news media sites. The time window of data collection was during the two-week period of the official election campaign (March 29–April 11). As shown above, commenters and comments are distinguished by whether the commenter's personal identity was verified (Factor 1) and whether an SNS account was used (Factor 2).

Although conceptualization of flaming is often vague and abstract, language uses have been at the center of its definition. Therefore, in order to identify a dependent variable of the study, comments that contain swearing or name-calling words were defined in this study as flaming comments. In order to properly

[3] The regulation pertains to Article 44-5, Section 1, Item 2 of the Act of Promotion of Information and Communications Network Utilization and Information Protection and Article 29 and Article 30, Section 1 of the Enforcement Ordinance of the same act. In 2009, the law was strengthened to require websites having a daily viewership of over 100,000 users to establish an identity-verification mechanism. In the summer of 2012, the Constitutional Court overturned the law, calling it unconstitutional.

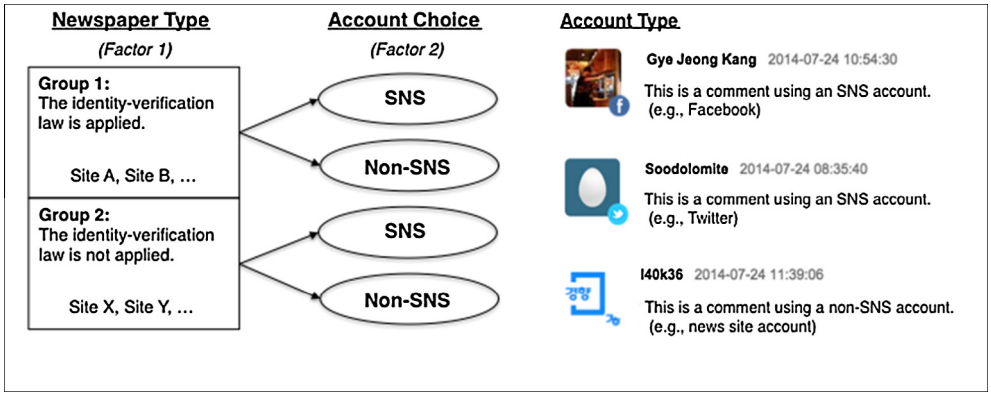

Fig. 1. Summary of research setting. Note: The identity-verification law is applied to newspapers in the Group 1 through which it is required to verify a real identity by submitting a national identification number. The law is not applied to newspapers in the Group 2. Commenters then choose an account between SNS and non-SNS accounts. SNS accounts include Facebook, Twitter, and other local SNS services. A non-SNS account includes a news site account. If a commenter signs in with any of SNS accounts, a user's current SNS profile picture that often contains a person's face image or other information that may be connected to a user's real identity. If others are interested in the commenter's profile, they can visit the commenter's personal SNS webpage by simply clicking the displayed image. If a user chooses a non-SNS account, the newspaper logo is appeared instead of a user's SNS profile picture. Followed by the image, a user name is displayed, which can be either a real name or a nickname.

classify such comments, contents were reviewed based on a dictionary of abusive languages. The dictionary consists of 651 commonly used strong swearwords and name-calling terms, which are the variations of 319 words designated as abusive by a transnational company *Nielsen Korea*, one of the largest audience and consumer research firms in South Korea. The selected terms included expletives and epithets (or commonly used pseudo-swearwords to bypass automatic filtering procedures) and other extreme anti-normative terms frequently used in online communities in South Korea. Two former journalists who worked at a major South Korean news media organization reviewed the dictionary to reassure the legitimacy of the dictionary. Exemplary flaming comments are translated into English for demonstration, with swearwords bold-faced.

> "These 6 year-old **birdbrains**, with no sense of the most basic principle that public officials should care about citizens' lives, keep avoiding their responsibility and are busy filling their own stomachs. Do they deserve to be the congressmen of our nation?"

> "Min-Tong-Dang (one of the political parties in Korea) is a melting pot **filled with bitches**. How dare they nominate **a bitch** like Mr. Kim? These **bitches** are **commies** in blood, and they don't care whether or not (a candidate is) an imposter, a liar, etc. as long as he has the **commie** taste."

> "President Rho (one of the former presidents in Korea) assisted your 'beloved' North Korea to arm with nukes. He should have been impeached at least 100 times. You **retards** need to get a sense back. . .or get hit by nuke!"

Inflammatory comments were automatically classified with the assistance of software. To verify the validity of our classification approach using the dictionary, 300 comments in the subset of flaming comments were randomly selected, and two human coders independently and manually reviewed each comment to determine if the comment is properly classified. Both coders agreed that over 97% of comments can legitimately be regarded as flaming comments by containing swearing terms, and a Cohen's kappa score of 0.82 suggested sufficient inter-coder reliability.

In addition, comment- and commenter-specific variables that may affect commenting behaviors were identified, and the descriptive statistics were produced, which are presented in Table 1.

5. Empirical analysis and results

As shown in Table 2, the majority of commenters used an SNS account ($N = 3908$, 67.93%) and the majority of comments were written in the news sites that were non-compliant with the identity verification law ($N = 9666$, 73.12%). This may be due to the fact that social commenting system was widely adopted due to its convenience, and that more news sites in our sample (21 out of 26 sites) were non-compliant sites. Despite unequal sample size, the website characteristics (e.g. organizational scales and political predispositions) varied randomly, eliminating the risk for systematic biases.

Using the 2×2 taxonomy described in Table 2, the ratios of inflammatory comments are calculated: Given that an SNS account is used, there is a higher likelihood of including flaming when comments are subject to the identity verification law. Given that a website is compliant with the law, however, there is a lower likelihood of including inflammatory comments when an SNS account is used (Fig. 2).

For formal hypotheses tests, we utilize the random effects panel Probit models (Poirier, 1980) in order to estimate the effects of treatments on the probability of flaming while acknowledging that the comments written by a single commenter may not be independent of one another. In other words, by establishing the unit of analysis as a commenter, where each comment constitutes a separate yet correlated data point, we control for the unobservable factors specific to each commenter. This approach assumes constant correlation across all comments written by a single commenter and accounts for that correlation in the variance–covariance matrix of the coefficients. While the estimated coefficients do not represent marginal effects of independent variables on the likelihood of inflammatory comments, the directionality of the effect estimated by the coefficient is important to note. Similarly, the magnitude of the marginal effect is not the primary focus of this study, although the true effect would be proportional to the magnitude of the presented coefficient. To highlight the effects of verification law and social commenting (Hypotheses H1–H3), the following regression frame is used:

$$Flaming_{ij} = \beta_0 + \beta_1 Veri_{ij} + \beta_2 SNS_{ij} + \beta_3 Veri_{ij} \times SNS_{ij} + \beta_4 AllComments_i + \beta_5 AllLengths_i + v_{ij}, v_{ij} = \alpha_i + u_{ij} \quad (1)$$

where i indexes the commenter and j indexes the comment. The dependent variable, *Flaming*, is equal to one if the comment includes designated swearing terms, and zero otherwise. Similarly, *Veri* is a dummy variable that indicates whether or not

Table 1
Descriptive statistics.

Variable	Description	Mean	Std. dev.	Min	Max
Flaming	Include inflammatory languages or not	0.1353	0.3421	0	1
Veri	Subject to the identity-verification law or not	0.2687	0.4433	0	1
SNS	Use of SNS account or not	0.5954	0.4908	0	1
AllComments	The number of comments by commenter	15.7737	36.0806	1	233
AllLengths	All lengths of comments by commenter	1226.9110	2383.8260	1	14,450

Table 2
The number of comments and commenters.

Comments (Commenters)		Factor 1		Sum
		Subject to the law	Not subject to the law	
Factor 2	SNS	1568 (570)	6303 (3338)	7871 (3908)
	Non-SNS	1985 (348)	3363 (1497)	5348 (1845)
Sum		3553 (918)	9666 (4835)	13,219 (5753)

a comment is subject to identity verification, and *SNS* denotes whether or not a comment is produced via SNS account. *AllComments* and *AllLengths* are discrete variables that control for commenter-specific characteristics and represent the number of all comments and the total comment length by a single commenter during the study period, respectively.[4] The unobserved variables, denoted by the error terms, are assumed to be normally distributed.

The variables of interest are *Veri* (or *SNS*), respective to hypotheses H1 (or hypothesis H2), and *Veri* × *SNS*, an interaction term between *Veri* and *SNS*. A negative coefficient on *Veri* (or *SNS*) would suggest that commenters are less likely to include flaming under the identity verification law (or under the use of an SNS account). A coefficient on *Veri* × *SNS* corresponds to hypothesis H3 and represents the compounding effect of the law and an SNS account.

The results of this study are displayed in Table 3. First, the positive and statistically significant coefficient for *Veri* in Column (1) suggests that the implementation of the law increases the likelihood of flaming. This result does not support the hypothesis H1 and is actually opposite of the policy's original objective. Conversely, the negative and statistically significant coefficient for *SNS* in Column (2) indicates that the use of an SNS account is associated with decrease in flaming, which supports hypothesis H2. The results from Column (3) include variables *Veri* and *SNS* as well as the interaction term *Veri* × *SNS*, and suggest that the estimated coefficient of *Veri* remains significant and positive while the coefficient on the interaction term is identified as negative and significant. This finding indicates that although the identity verification law is associated with elevated probability of flaming, the use of SNS accounts could counterbalance such likelihood and alleviate inflammatory commenting (Hypothesis H3 is supported). Lastly, the results from the control variables indicate that (1) there is a low likelihood of flaming as a commenter writes more comments, and (2) the length of comments is positively associated with the likelihood of flaming.

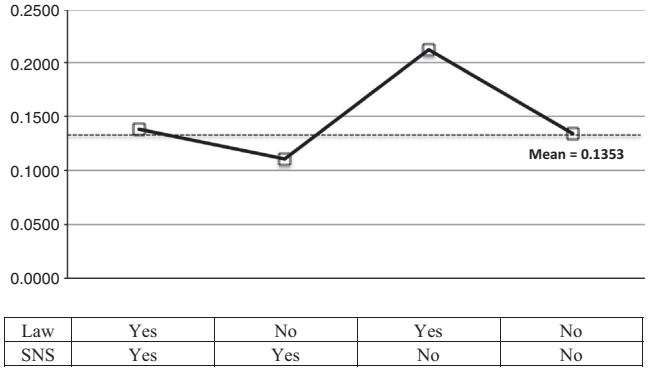

Fig. 2. The ratio of comments including flaming.

The analyses presented in the previous section assume random distribution of commenters across news sites as well as random distribution of SNS account users versus non-users. This assumption leaves room for a potential bias in which the composition of a particular group of commenters may be systematically different from that of other groups—so-called the selection bias. In particular, the significant results of commenter characteristics (*AllComments* and *AllLengths*) may invoke suspicions regarding any systematic bias induced by the characteristics of commenters.

To address this concern, additional modeling is conducted based on the commenter types. Specifically, the commenters are categorized into three groups – heavy, moderate, and light commenters – in order to account for commenter characteristics, which are then used in propensity score matching (PSM) (Dehejia & Wahba, 2002; Rosenbaum & Rubin, 1983). The PSM method provides a randomized, experiment-like setting that excludes the impact of unobserved heterogeneity by matching observational commenter-specific characteristics between the two groups of observations. We formally present the model as follows:

$$E(Flaming = 1 | Treatment = 1, \boldsymbol{X}) > E(Flaming = 1 | Treatment = 0, \boldsymbol{X}) \quad (2)$$

where the treatment groups are defined as commenters subject to the identity-verification law or those who use an SNS account. $\boldsymbol{X}$ is a vector covariate including *AllComments* and *AllLengths*. The variable, *SNS* (or *Veri*), is included as one of covariates when the treatment is the *Veri* (or the *SNS*).

The average treatment effects (ATE) using the PSM method are illustrated in Tables 4 and 5.[5] First, Table 4 takes a close-up look at flaming occurrences by comparing the composition of commenters and comments between identity verified and non-verified groups. As seen in the table, the heavy user ratio of commenter and comments in the identity verified group is much larger than that in the non-verified group. Furthermore, the proportion of flaming comments in the identity verified group is larger than that in the non-verified group. The ATE from the PSM test show that all group-specific coefficients are positive and statistically significant, suggesting that the implementation of the identity verification law is indeed associated with high probability of inflammatory language uses.

Table 5 compares between SNS account users and non-SNS account users. While the number of commenters is highly skewed to the group of light commenters, no significant difference is observed in terms of the proportion of flaming comments among

[4] We assume that commenters are randomly distributed across news sites in our sample, and news sites are randomly subject to the identity-verification law. These commenter-specific variables are included for the descriptive purpose to see how the type of commenters is related to offensive behaviors.

[5] We use the PSMATCH, propensity score matching module, in Stata 13 to match samples using the nearest 10 neighbors.

Table 3
The effects of identity verification law and social commenting on flaming.

Column	(1)		(2)		(3)	
	Coefficient	Std. error	Coefficient	Std. error	Coefficient	Std. error
Veri	0.5058**	(0.1015)			0.7981**	(0.1544)
SNS			−0.1978*	(0.0875)	−0.0337	(0.1009)
Veri × SNS					−0.5405**	(0.2031)
All Comments	−0.0176**	(0.0062)	−0.0169**	(0.0062)	−0.0185**	(0.0062)
All Lengths	0.0001**	(0.0000)	0.0002**	(0.0000)	0.0001*	(0.0000)
Constant	−2.6566**	(0.0782)	−2.4361**	(0.0966)	−2.6156**	(0.1068)
Prob > χ^2	0.000		0.000		0.000	
# of commenters	5753		5753		5753	
Number of comments	13,219		13,219		13,219	

Robust standard errors are in parentheses.
** $p < 0.01$.
* $p < 0.05$.

Table 4
The proportion of commenter, comment, and flaming between identity verified and non-verified group and propensity score matching test results (DV = Flaming).

	Subject to the law			Not subject to the law			ATE
	Commenter (%)	Comment (%)	Proportion of flaming comments (%)	Commenter (%)	Comment (%)	Proportion of flaming comments (%)	
(1) Light commenter	705 (76.8%)	1047 (29.5%)	15.4	4364 (90.3%)	5637 (58.3%)	12.6	0.052* (0.020)
(2) Moderate commenter	142 (15.5%)	791 (22.2%)	20.4	377 (7.8%)	1990 (20.6%)	14.8	0.137* (0.059)
(3) Heavy commenter	71 (7.7%)	1715 (48.3%)	18.5	94 (1.9%)	2039 (21.1%)	7.2	0.364** (0.074)
(4) Sum	918 (100.0%)	3553 (100.0%)	18.1	4835 (100.0%)	9666 (100.0%)	11.5	0.061** (0.019)

Note: Light commenters are those who left 1–3 comments, moderate who left 4–9, and heavy commenter who left more than or equal to 10; ATE = Average Treatment Effect. Robust standard errors are in parentheses under ATE.
** $p < 0.01$.
* $p < 0.05$.

Table 5
The proportion of commenter, comment, and flaming between SNS and Non-SNS account users and propensity score matching test results (DV = Flaming).

	SNS account users			Non-SNS account users			ATE
	Commenter (%)	Comment (%)	Proportion of flaming comments (%)	Commenter (%)	Comment (%)	Proportion of flaming comments (%)	
(1) Light commenter	3521 (90.1%)	4617 (58.6%)	13.0	1548 (83.9%)	2067 (38.6%)	13.1	−0.017 (0.016)
(2) Moderate commenter	316 (8.0%)	1685 (21.4%)	13.5	203 (11.0%)	1096 (20.5%)	20.7	−0.144** (0.051)
(3) Heavy commenter	71 (1.9%)	1569 (20.0%)	5.6	94 (5.1%)	2185 (40.9%)	17.2	−0.145 (0.075)
(4) Sum	3908 (100.0%)	7871 (100.0%)	10.7	1845 (100.0%)	5348 (100.0%)	17.0	−0.054* (0.021)

Note: Light commenters are those who left 1–3 comments, moderate who left 4–9, and heavy commenter who left more than or equal to 10; ATE = Average Treatment Effect. Robust standard errors are in parentheses under ATE.
** $p < 0.01$.
* $p < 0.05$.

light commenters. It is possible that the light commenters may write only one or a few comments extemporaneously when they want to express their opinions. Noticeable differences, however, are observed in terms of the proportion of flaming comments within moderate and heavy commenter groups, suggesting that moderate and heavy commenters may care about their social images represented by their commenting behaviors more than light commenters. Despite the ostensible difference of the proportion of flaming within the heavy commenter group, however, the difference is not so statistically significant after we control possible covariates (*p-value: 0.053*). This is due to the fact that some heavy commenters using a non-SNS account wrote the large number of flaming comments, whereas those using an SNS account wrote the very small number of flaming comments – for example, a heavy commenter using non-SNS account wrote 32 flaming comments out of 48 comments (66.7%), while a heavy commenter using an SNS account wrote no flaming comments out of 119 comments (0.0%). PSM results control these possible outliers, and focus on the likelihood of writing at least one flaming comment at the commenter-level. As a result, all group-specific coefficients are negative but not all coefficients are statistically significant, indicating that the use of an SNS account may decrease the probability of inflammatory comments especially for the moderate-level commenters.

6. Discussion and conclusions

The issue of anonymity has been at the center of scholarly discourses on ways to promote respectful climate for online public

discussions. In line with recent regulatory efforts to moderate anonymous commenting behaviors, this study explores the effects of different anonymity control mechanisms – policy-driven regulations and the promotion of voluntary self-disclosure – on inhibiting inflammatory comments. The policy-driven approach accentuates a sense of identifiability by the entities who are in charge of penalizing anti-normative behaviors, while the voluntary approach appeals to the identifiability by peer users, which pertains to the social incentives of positive self-presentation. To compare these two mechanisms, this study leverages a large, real-world dataset of newspaper website comments during an election campaign period in South Korea. The particular societal context is chosen due to the natural real-world setting created through the coexistence of the nationwide identity verification law (a policy-driven regulatory approach) and the widespread use of social commenting system (a voluntary approach).

Verifying identification credentials do not indicate that commenters' real name is displayed with their comments. Instead, as far as a user's identity is verified through the system of the website operator, it allowed users to maintain pseudonyms or screen names when they write their comment. Using an SNS account, however, indicates that a commenter's social cues are likely to be disclosed, because others can visit the commenter's personal SNS webpage. This study reveals that the enforcement of identity verification law is noticeably associated with an *increased* probability of inflammatory commenting, which is counterintuitive against the proposed policy goal. On the contrary, the use of an SNS account is negatively correlated with flaming. Our results suggest that commenters are more likely to be affected by the disclosure of their SNS profile with a comment than are they subject to the law. In other words, while both a policy-driven regulation and a voluntary disclosure of social cues are mechanisms to enhance the degree of identifiability, the public visibility of personal profile through using an SNS account tends to be related to reducing flaming in user comments online.

Additional PSM tests on commenter groups of different levels of activities (heavy, moderate, and light commenters) validate that the effects of the two mechanisms uniquely account for flaming across commenters with varying commenting behaviors. In particular, social commenting effect was the most prominent among the moderate commenters. This may be due to the fact that more or less frequent commenters may perceive their comments to affect their social image and be aware of language choices when they use an SNS account. We also observed a salient difference of the proportion of flaming comments between using an SNS account and a non-SNS account for heavy commenters, but the difference seems to result from a few outliers within the group of heavy commenters.

6.1. Theoretical implications

Considering that the identity verification law in South Korea had not required public display of real names, the ostensible anonymity, due to the use of pseudonyms, and the salient common political identity among highly opinionated likeminded users could have created a combined effect on inducing anti-normative commenting against out-group entities. This rationale is consistent with the SIDE model yet resonates with negative consequences. If this dynamic is truly occurring, the identity verification policy, and possibly any other top-down regulations that increase the transaction cost for political expressions, could prompt unfavorable conditions of public deliberations such as aggravation of political polarization and removal of moderate voices from the public opinion landscape. While this study could not investigate how the verification process – or other policy implementation that increase the transaction cost of discussion engagement – is related with the salience of group identity, future research is recommended to delve into the role of group political identity in polarization and intergroup flaming.

Our results of reduced flaming comments by using an SNS account are in line with findings in a majority of empirical research on deindividuation effect, indicating that the more identifiable, the less disinhibited. The voluntary disclosure of social cues is also related to social penetration theory, suggesting that increasing degree of self-disclosure is an outcome of natural evolution of interpersonal relationships (Altman & Taylor, 1973). As far as the minimum level of privacy is placed, privacy provides a foundation for self-disclosure, which allows users to build social capital and self-esteem and to achieve the desire for interaction, socialization, and recognition (Acquisti, Brandimarte, & Lowenstein, 2015). Commenters using an SNS account may believe that these motives are more essential than the protection of privacy, and they are likely to behave prosocially on the Internet in order not to produce a negative social image (Bénabou & Tirole, 2006).

6.2. Practical implications

Our results also have several policy and managerial implications. Most importantly, our findings suggest that the effectiveness of top-down interventions that increase the traceability of identity may be limited or even counterproductive on shaping political discussion culture. The results regarding the Identity Verification Law suggest that some potential commenters may withdraw themselves from writing comments due to the inconvenience and risks associated with the verification process, whereas others who undergo the verification process may do so because they perceive the benefits gained from becoming an online community member and expressing opinions to be higher than the costs incurred from the verification process. In such scenario, it is possible that the comments posted on a particular news site could over-represent the highly opinionated individuals who share political ideology promoted by the reader community of the news website.

This conjecture could be true especially in the South Korean context where most major newspapers reveal political preferences to a certain extent. Highly opinionated comments among likeminded users could polarize the political discussion climate, resulting in greater verbal aggressions against opposite opinions or individuals. The heightened transaction cost induced by the identity verification process could encourage such a pattern, resulting flaming towards non-likeminded others to be taken as a normative behavior within the likeminded community. The comparison of flaming proportions between identity verified and non-verified groups, as seen in Table 4 above, supports our inference in that the identity verified users are more likely to repeatedly write inflammatory comments and become aggressive against the opposite opinions.

While the Korea's Identity Verification Law may be regarded as one of a few special cases with heavy hand in regulating the Internet space by the government, preventing malicious use of the Internet has been of key issue in many countries. For example, the use of surveillance-oriented information technologies, of which the Internet monitoring is a central part, for national security is a growing phenomenon over the world. Accordingly, the enhanced level of identity traceability on the Internet has become a common practice across many governments. However, at the same time, the civil rights advocates worry that legal actions to eradicate the Internet privacy for the sake of security may affect online freedom of speech. Our results imply that forceful governmental intervention by implementing the verification process may not be a legitimate solution.

When it comes to the private sector, Google had an intention to force users to use their real name as they launched a new SNS,

Google+, in 2012, and it reversed the policy in 2014 after the substantial criticism. Facebook still holds the real name policy, and advocacy groups, such as the Electronic Frontier Foundation, keep challenging the policy, because the Facebook's policy led to non-trivial ongoing disputes, such as recent controversies against Native Americans and LGBTQ communities. These can be another evidences that the real name policy may provoke substantial problems, which is in accordance with our finding. Given that the issue of real-identity verification on the Internet is still controversial in various aspects, our study may provide practical implications to those countries and online service providers involved in the matters, suggesting that implementing the identity verification process as a remedy for the potential threat could result in unexpected policy outcomes. On another note, the positive association between the implementation of the law and flaming could also mean that the citizens have become desensitized to the regulations and thus no longer care about liability. This interpretation is consistent with the findings of short-term effects of the Identity Verification Law by the previous studies (Cho & Kim, 2012). Our results point out a fundamental top-down policy problem inherent in the Internet laws in general: As the use of social media becomes widespread at an accelerating speed, bureaucratic control of the far-reaching Internet space becomes extremely difficult and nearly impossible.

Meanwhile, our findings provide strong evidence for the possibility of effective implementation of voluntary means, in particular a social commenting system. The spillover of social cues seems to motivate users to express themselves in a responsible manner. Furthermore, users may choose to utilize the social commenting system because it lowers the transaction cost and increases convenience through bypass of the sign-up process. This is contrary to the identity verification process, and may possibly facilitate spontaneous participation in political discourses without the burden of establishing oneself as a member of the community. The current findings show the potential benefits of implementing voluntary approach in reducing anti-normative messaging.

However, it remains to be answered by future research whether the voluntary approach only reduces negative commenting behaviors or suppresses all types of discussion activities as a whole. There are anecdotal evidences suggesting that the disclosure of social cues does not necessarily decrease the volume of comments. For example, Huffington Post, which implements a social commenting system, has experienced steadfast increase of comments from 3 million in 2011 to 9 million in 2013 (Soni, 2013). However, descriptive examples do not validate a significant association between the voluntary approach and the general liveliness of the discussion culture. More systematic investigation is recommended for delving into broad, long-term implications of the self-disclosure culture on promoting public discussions. Furthermore, the voluntary mechanism potentially provokes other social issues such as privacy. Although discussion about such related issues is beyond the scope of the current study, these issues must be addressed by future research.

That said, we note that the study has several limitations. First and most importantly, we operationalized flaming somewhat narrowly, by classifying a comment as flaming if it includes an abusive language. Accordingly, a possibility to include false negative cases exists: The classification could neglect comments that are subtly abusive without the use of profane languages. While we used a computer-assisted classification approach due to the data size, future research might be interested in more qualitative, granular approach to explore the ways in which flaming is expressed in user commenting. Second, although we have attempted to address differences across news sites with and without the enforcement of the Identity Verification law as well as differences among commenters using an SNS account or a non-SNS account, we cannot completely rule out the possibility of selection bias, as is the case with any observational study. To strengthen the external validity, further investigations using an addition data set – for example, data from the non-election period after the expiration of the law, or data containing comments from news sites where the voluntary mechanism is not available – may provide additional insights. In addition, possible concerns exist on whether the news sites selected in the study could be biased. Nonetheless, we believe that our sample is an appropriate source from which to assess commenting behaviors of the population because our sample shows a sufficient variation in terms of types of readers, focus (e.g., business newspapers and local newspapers), and political orientations.

Despite some limitations, this study contributes to understand different anonymity moderation mechanisms in online news communities and how these mechanisms affect to reduce flaming in citizen-to-citizen public discussion. A policy-driven process could result in side effects by increasing transaction costs, while social commenting approach could facilitate discussion engagements in a respectful manner by decreasing transaction costs. The exploration of a large-scale real-world data has a unique contribution to understand the relationship between anonymity and flaming. Although limited factors were investigated in this study, the study does not dismiss other effects not included in the analyses. More extensive investigations on the content and valence of messages may help reveal new insights beyond the scope of this study.

References

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science, 347*(6221), 509–514.

Alonzo, M., & Aiken, M. (2004). Flaming in electronic communication. *Decision Support Systems, 36*(3), 205–213.

Altman, I., & Taylor, D. (1973). *Social penetration: The development of interpersonal relationships*. New York: Holt, Rinehart & Winston.

Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review, 96*(5), 1652–1678.

Brooks, D. J., & Geer, J. G. (2007). Beyond negativity: The effects of incivility on the electorate. *American Journal of Political Science, 51*, 1–16.

Canter, L. (2013). The misconception of online comment threads. *Journalism Practice, 7*(5), 604–619.

Cho, D., & Acquisti, A. (2013). The More social cues, the less trolling? An empirical study of online commenting behavior. *Working paper*.

Cho, D., & Kim, S. (2012). Empirical analysis of online anonymity and user behaviors: the impact of real name policy. In *Proceeding of the 2012 45th Hawaii international conference on system sciences (HICSS)* (pp. 3041–3050).

Christopherson, K. M. (2007). The positive and negative implications of anonymity in Internet social interactions: 'On the Internet, nobody knows you're a dog'. *Computers in Human Behavior, 38*(6), 3038–3056.

Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication, 64*(4), 658–679.

Coffey, B., & Woolworth, S. (2004). "Destroy the scum, and then neuter their families:" the web forum as a vehicle for community discourse? *The Social Science Journal, 41*(1), 1–14.

Cohen, J. E. (1995). Right to read anonymously: A closer look at copyright management in cyberspace. *Connell Law Review, 28*, 981–1040.

Connolly, T., Jessup, L. M., & Valacich, J. S. (1990). Effects of anonymity and evaluative tone on idea generation in computer-mediated groups. *Management Science, 36*(6), 689–703.

Coyne, S. M., Stockdale, L. A., Nelson, D. A., & Fraser, A. (2011). Profanity in media associated with attitudes and behavior regarding profanity use and aggression. *Pediatrics, 128*(5), 867–872.

Dahlgren, P. (2005). The Internet, public spheres, and political communication: Dispersion and deliberation. *Political Communication, 22*(2), 147–162.

Dehejia, R. H., & Wahba, S. (2002). Propensity score matching methods for nonexperimental causal studies. *The Review of Economics and Statistics, 84*(1), 151–161.

Dubrovsky, V., Kiesler, S., & Sethna, B. (1991). The equalization phenomenon: Status effects in computer-mediated and face-to-face decision-making groups. *Human–Computer Interaction, 6*(2), 19–146.

Festinger, L., Pepitone, A., & Newcomb, T. (1952). Some consequences of de-individuation in a group. *Journal of Abnormal and Social Psychology, 47*(2s), 382–389.

Gross, R., & Acquisti, A. (2005). Information revelation and privacy in online social networks. In *Presented at the Workshop on Privacy in the Electronic Society, New York, NY* (pp. 71–80).

Halpern, D., & Gibbs, J. (2013). Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression. *Computers in Human Behavior, 29*(3), 1159–1168.

Herring, S., Job-Sluder, K., Scheckler, R., & Barab, S. (2002). Searching for safety online: Managing "trolling" in a feminist forum. *The Information Society, 18*(5), 371–384.
Hollingshead, A. B. (1996). Information suppression and status persistence in group decision making: The effects of communication media. *Human Communication Research, 23*(2), 193–219.
Humphreys, L. (2011). Who's watching whom? A study of interactive technology and surveillance. *Journal of Communication, 61*, 575–595.
Jessup, L. M., Connolly, T., & Galegher, J. (1990). The effects of anonymity on GDSS group process with an idea-generating task. *MIS Quarterly, 14*(3), 313–321.
Johnson, N. A., Cooper, R. B., & Chin, W. W. (2009). Anger and flaming in computer-mediated negotiation among strangers. *Decision Support Systems, 46*(3), 660–672.
Kling, R., Lee, Y. C., Teich, A., & Frankel, M. S. (1999). Assessing anonymous communication on the internet: Policy deliberations. *The Information Society, 15*(2), 79–90.
Kushin, M. J., & Kitchener, K. (2009). Getting political on social network sites: Exploring online political discourse on Facebook. *First Monday, 14*(11). http://dx.doi.org/10.5210/fm.v14i11.2645.
Lampe, C., Zube, P., Lee, J., Park, C. H., & Johnston, E. (2014). Crowdsourcing civility: A natural experiment examining the effects of distributed moderation in online forums. *Government Information Quarterly, 31*(2), 317–326.
Lea, M., O'Shea, T., Fung, P., & Spears, R. (1992). 'Flaming' in computer-mediated communication: Observations, explanations, implications. In M. Lea (Ed.), *Contexts of computer-mediated communication* (pp. 89–112). London: Harvester Wheatsheaf.
Lee, E. J. (2004). Effects of visual representation on social influence in computer-mediated communication. *Human Communication Research, 30*(2), 234–259.
Lee, E. J. (2007). Deindividuation effects on group polarization in computer-mediated communication: The role of group identification, public-self-awareness, and perceived argument quality. *Journal of Communication, 57*(2), 385–403.
McPhail, C. (1991). *The myth of madding crowd*. New York, NY: Walter de Gruyter.
Moor, P. J., Heuvelman, A., & Verleur, R. (2010). Flaming on Youtube. *Computers in Human Behavior, 26*(6), 1536–1546.
Mutz, D. C., & Reeves, B. (2005). The new videomalaise: Effects of televised incivility on political trust. *American Political Science Review, 99*, 1–16.
Newman, J. (April 3, 2012). Arizona looks to outlaw Internet trolling. *Time*. <http://techland.time.com/2012/04/03/arizona-looks-to-outlaw-internet-trolling/> (Retrieved 25.08.14).
Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society, 6*(2), 259–283.
Park, W., & Greenleaf, G. (2012). Korea rolls back 'real name' and ID number surveillance. *University of New South Wales Faculty of Law Research Working Paper Series*.
Poirier, D. J. (1980). Partial observability in bivariate probit models. *Journal of Econometrics, 12*, 209–217.
Postmes, T., & Spears, R. (1998). Deindividuation and anti-normative behavior: A meta-analysis. *Psychological Bulletin, 123*, 238–259.
Postmes, T., Spears, R., Sakhel, K., & de Groot, D. (2001). Social influence in computer-mediated communication: The effects of anonymity on group behavior. *Personality and Social Psychology Bulletin, 27*, 1243–1254.
Rainie, L., & Wellman, B. (2012). *Networked: The new social operating system*. Cambridge, MA: MIT Press.
Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.
Rowe, I. (2014). Civility 2.0: A comparative analysis of incivility in online political discussion. *Information, Communication & Society*. http://dx.doi.org/10.1080/1369118X.2014.940365.
Ruiz, C., Domingo, D., Micó, J. L., Díaz-Noci, J., Meso, K., & Masip, P. (2011). Public sphere 2.0? The democratic qualities of citizen debates in online newspapers. *The International Journal of Press/Politics, 16*(4), 463–487.
Scott, C. (1998). To reveal or not to reveal: A theoretical model for anonymous communication. *Communication Theory, 8*(4), 381–407.
Sia, C.-L., Tan, B. C. Y., & Wei, K.-K. (2002). Group polarization and computer-mediated communication: Effects of communication cues, social presence, and anonymity. *Information Systems Research, 13*(1), 70–90.
Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: An interdisciplinary review. *MIS Quarterly, 35*(4), 989–1016.
Soni, J. (August 26, 2013). The reason HuffPost is ending anonymous accounts. *Huffingtonpost.com*. <http://www.huffingtonpost.com/jimmy-soni/why-is-huffpost-ending-an_b_3817979.html>.
Spears, R., & Lea, M. (1994). Panacea or panopticon? The hidden power in computer-mediated communication. *Communication Research, 21*(4), 427–459.
Sproull, L., & Kielser, S. (1991). *Connections: New ways of working in the networked organization*. Cambridge, MA: MIT Press.
Stephen, A. T., & Galak, J. (2012). The effects of traditional and social earned media on sales: A study of a microlending marketplace. *Journal of Marketing Research, 49*(5), 624–639.
Straus, S. (1997). Technology, group process, and group outcomes: Testing the connections in computer-mediated and face-to-face groups. *Human–Computer Interaction, 12*(3), 227–266.
Sunstein, C. R. (2014). *On rumors: How falsehoods spread, why we believe them, and what can be done*. Princeton, NJ: Princeton University Press.
Walther, J. B., DeAndrea, D., Kim, J., & Anthony, J. C. (2010). The influence of online comments on perceptions of antimarijuana public service announcements on YouTube. *Human Communication Research, 36*(4), 469–492.
Wang, S. S. (2013). "I Share, Therefore I Am": Personality traits, life satisfaction, and Facebook Check-Ins. *Cyberpsychology, Behavior, and Social Networking, 16*(12), 870–877.
Weber, & Shandwick (2013). Civility in America 2013. <http://www.webershandwick.com/uploads/news/files/Civility_in_America_2013_Exec_Summary.pdf> (Retrieved December 2014).
Youmans, W. L., & York, J. C. (2012). Social media and the activist toolkit: User agreements, corporate interests, and the information infrastructure of modern social movements. *Journal of Communication, 62*(2), 315–329.
Zarsky, T. Z. (2003). Thinking outside the box: Considering transparency, anonymity, and pseudonymity as overall solutions to the problems in information privacy in the Internet society. *University of Miami Law Review, 58*, 1028–1032.
Zimbardo, P. G. (1969). The human choice: Individuation, reason, and order versus deindividuation, impulse, and chaos. *Nebraska symposium on motivation* (Vol. 17, pp. 237–307). University of Nebraska Press.