The Ethics of Corporate Censorship of Information-Sharing Behavior: A Nonconsequentialist Perspective

Hanfeng Zhong and Paul A. Watters

Abstract

Recent research on information-seeking behavior (Watters and Ziegler 2016) has suggested a role for managing access to ensure consistency with local regulatory or policy requirements. However, sharing of self-generated, personal data—as facilitated by social-media companies—should be relatively free of information-sharing controls. While many studies have examined government censorship, the extent to which the private sector is complicit is often unclear. In this study, we examine whether censorship appears to occur on a number of social-media and related sites, including the transmission of sensitive keywords and URLs. The results indicate that some level of private-sector censorship is prevalent, often in breach of the technology companies' own terms and conditions. In some cases, apparently harmless information is overblocked. These companies need to be more transparent about their censorship mechanisms and subject their actual policies and procedures to scrutiny and public debate. Removing controls on information-sharing behavior is consistent with a nonconsequentialist perspective on privacy.

Introduction

Social networks play a significant role in our social lives; they have been used as a platform for political campaigning, mobilizing protests, fostering political expression, and cultivating debate (Jackson 2014). Recently, people have become more willing to share private information and viewpoints online, because social networks provide a global platform for discussion and debate. However, social networks have also been used to transmit child pornography and materials inciting extreme violence, presenting

LIBRARY TRENDS, Vol. 68, No. 4, 2020 ("Censorship and Information Behavior in the Age of Big Data," edited by Paul A. Watters), pp. 697–711.

governments with an information-control challenge (Watters and Ziegler 2016; Malamuth, Hald, and Koss 2012). An increasing number of countries have sought to enhance their powers to regulate social networks to ostensibly combat these threats. While some critics argue that government censorship is futile (Nabi 2014), others have demonstrated that government censorship online is common and may be effective in some cases (Watters 2015).

Why do governments seek to censor the Internet? Some potentially legitimate purposes of Internet censorship are to block sites that are related to radicalization of terrorists, promotion of hate speech, child pornography, and drug abuse or production (Duffy 2015). However, in most cases, and in many jurisdictions, government censorship may be viewed as heavy-handed and an unnecessary interference in the private lives of ordinary citizens. Indeed, as it is likely that the number of people who will protest against government censorship is likely to increase, if the censorship regime is harsh, arbitrarily enforced, and unfair, this may lead to unintended consequences. For instance, Internet users may be unwilling to blog due to fear of government persecution, which may influence people's preparedness to communicate (Roberts 2018).

To what extent are governments currently engaged in Internet censorship? In some jurisdictions, such as the United States, the level of censorship (and surveillance) is unclear, but historically variable (Singer and Friedman 2014). Changes in the degree of censorship are usually justified because of the potential adverse impact on national security, public interest, and justice (Kopel 2013). For US corporations, the extent to which they may engage in censorship on behalf of the government remains an open question: several former Facebook employees, for example, disclosed that Facebook routinely censors Conservative news stories, and that this may have had an influence on the US federal election outcome (Nunez 2016).

In contrast, countries like Japan appear very reluctant to censor the Internet. Materials such as child pornography and extreme violence appear relatively easy to access in Japan. In April 2009, bowing to international pressure, the Japanese government released a new policy to block children's access to porn sites via mobile devices, with enforcement devolved to Internet service providers (Aoyama, Utsumi, and Hasegawa 2012). Internet providers can confirm Internet users' age through their membership-authorization system. The Japanese government also provides free content-filtering software to the public, but its use is not mandatory.

On the other hand, Internet censorship in South Korea appears to be somewhat draconian: Internet users must submit their personal information in detail, including their real names, address, ID number, and occupation, if they want to apply for their accounts or an email address on domestically operated providers (Oh et al. 2010). Ostensibly, the government only wants to authenticate users to prevent hacking and identity

theft (Sung 2017), yet a culture of fear may prevail where users are prevented from freely expressing criticism of their governments.

Numerous studies have reviewed the role of government censorship. Yet, corporations can decide to regulate the Internet in their organization and limit the information released. This may be due to government encouragement, e.g., to block child pornography and extreme violence (Bambauer 2009). In some jurisdictions, such as Singapore, a coregulatory regime operates that shares responsibility for censorship across the public and private sector (Watters and Ziegler 2017). The broader question is—What are the ethical standpoints taken by these companies, and how are they informed? A nonconsequentialist perspective might suggest that, since the act of transmitting personally generated data may be more significant than the outcome, it is not clear why a third-party (the social network) feels compelled to censor, in the absence of a legal imperative.

Internet-filtering software has been widely used by schools and libraries to block some information that may not be suitable for school or library settings, such as pornography, advertising, chat, gaming, social networks, and online forum sites (Spacey et al. 2015). Yet very few studies have directly investigated the extent to which social networks routinely censor their users' public and private messages. In this study, we investigate corporate censorship on the most popular social networks, including Facebook, Twitter, Google, Reddit, and Imgur. The majority of censoring and blocking activities from social-network sites are based on keywords or phrase searching (Ho and Watters 2004; Yangyue 2014).

Many previous studies of Internet censorship focus on political censorship but not on less contentious (but perfectly legal) discourse. In this study, we investigate the nature of corporate social-network censorship by asking several related questions:

- Which content is frequently censored by the top five social-network sites?
- Which site practices the most content filtering?
- Do some social networks censor content in some countries more than others?
- Are private conversations censored more than public posts?

Literature Review

Why is censorship such a topical issue in relation to social networks? Recent history has shown that protestors and revolutionaries have made very effective use of (a) secure messaging technology away from the prying eyes of law enforcement and (b) geolocation-based sharing and topic-based broadcasting as a means to rapidly organize political protest. For instance, the Arab Spring movement that spread across the Middle East adopted Facebook and Twitter for organizing their protests (Frangonikolopoulos and Chapsos 2012). In 2011, the supporters of the Occupy movement

in the United States used social networks like Facebook and Twitter to organize protests against the government (Kwon, Moon, and Stefanone 2015). The number of participants who browsed more than four hundred Facebook pages about this movement reached 2.7 million, and there were in excess of one hundred Twitter accounts with tens of thousands of followers. The volume of data is staggering, and the recent protests in South Dakota (Whyte 2017) against a government-sanctioned oil pipeline saw attempts to "check-in" with geo-tagged coordinates in an attempt to increase the manual efforts of law enforcement to identify suspects and map their connections. Governments are fearful of the power of this new technology.

Aside from government suppression of dissent, a more difficult issue surrounds the use of "community standards" to remove offensive or illegal content. Facebook, for example, enforces a series of vague "community standards" that can assist in free-speech suppression. The human moderators employed by Facebook have wide powers to remove any posts that seem "inappropriate" without any reason, and users have no opportunity to see the notice or warning (Sirichit 2015). Facebook automatically removes users' comments and classifies them as "spam." As a result, most Facebook users cannot see the "Promote" feature at the bottom right corner of their posts, and they even don't know how many visitors read their post. If your post is deleted by a moderator, Facebook will log you out while they remove your posts and require you tick a checkbox to declare that you understand their "community standards"; otherwise, you cannot log back in. There have been numerous reports of Facebook removing breastfeeding photos and artwork (Ballard and McKinlay 2011; Blue 2015) while allowing sexualized advertising and content. Government censorship programs are usually grounded in a firm policy base, such as copyright law, trademark laws, and regulations governing publicity (Tushnet 2000). In the private sector, there are no such controls beyond the Terms of Service.

The arbitrary application of ill-specified moderation policies could ultimately lead to self-censorship (Sleeper et al. 2013). For instance, some Facebook users closely manage their self-representations in different social contexts because they are not willing to break away from the community's perceived social norms (Das and Kramer 2013). This could be due to a fear of government persecution. On some platforms, like Twitter, all posts are public by default; there are no fine-grained access controls like on Facebook (Driscoll and Walker 2014). The lack of protective measures can play a significant role in public discourse: an unknown attacker used 25,860 fraudulent accounts to send 440,793 tweets in an attempt to subvert political discourse following the announcement of Russia's parliamentary election results (Thomas 2013).

Google offers a range of social-media and searching services; all are

susceptible to censorship. Sometimes, this is at the request of host governments: only a minority of countries take direct actions to ban Google because they will not comply with government censorship requests (Rushe 2012). In other cases, Google willingly complies: for example, users cannot retrieve any search results when searching for "Nazi" on Google Germany because of local laws (Zittrain 2010). Sometimes, compliance is refused: Google once declined the request from Pakistan to remove six videos on YouTube that satirized their army and senior politicians. After that, some related Google services like Maps, Drive, and Play Store were banned because the same IPs are shared across all of these services (Nabi 2014). Recently, Pakistan blocked some popular websites like Google, Facebook, Wikipedia, and YouTube successively (Chaabane et al. 2014). So, there may be a commercial imperative for social networks to ultimately comply with government censorship requests in order to stay in business. In recent years, Google approved 47 percent of informal requests for content removal and 65 percent of court orders within six months (Rushe 2012). Even Google is concerned about the volume of requests to remove political content on its search results, especially since the majority of these requests came from users in Western democracies (Franks 2012).

Managing censorship with social networks is a technical challenge. Aside from user reporting, there may be algorithmic approaches, which can reduce human effort. Take Reddit, for example: it contains more than 240,000 "subreddits," also called "content categories." Fifty percent of these subreddits are inactive (i.e., they received fewer than five posts in total) (Mills and Fish 2015). Users' selections about browsing posts from the primary page type based on Reddit's "Hot" ranking algorithm and their subscription of posts on subreddits would be identified as the highest ranks of posts by this algorithm. Censors could simply use the "hot" rank to review the highest-impact discussions and censor them as desired. This is because users' positive or negative votes to different links and posts could impact the positioning of those postings on the front page. However, in order for posts to reach the front, there had to be enough positive votes and comments from users (Soha 2012). Reddit users appear to have demanded some level of censorship from the service: at some point, they began to ban certain subreddits promoting child pornography and "anti-women" posts (Workman 2014). Some users emphasize the process of interacting with other users on Reddit, while others post content merely for collecting "Karma points." The second approach is called "Karmawhoring," which involves the collection of Karma points on each post as allocated by Reddit's ranking and evaluation system. A user's contribution's value on Reddit is presented in forms of Karma points, and their upvotes and downvotes calculate the Karma points. However, individual posts that receive more Karma points do not mean they have good content quality (Richterich 2014).

Methods

Five leading social-network sites (Facebook, Twitter, Google, Reddit, and Imgur) were chosen for exploring censorship because of their market popularity. We tested a range of censorship scenarios by performing topic searches of hundreds of sensitive keywords on these social networks. We also recorded the results of whether these keywords were blocked or not.

To collect censorship data from different countries, we used a VPN to connect to the Internet from these countries: New Zealand, the United States, Russia, Japan, and South Korea. Among these countries, some of them implemented strict policies on social-network sites, while others were notoriously lax. We utilized different test data for collecting censorship data from five social networks because each site has different characteristics. We aggregated censorship data separately for each site. Additionally, we aggregated censorship data on comparative analysis between three social networks—Twitter, Reddit, and Imgur—because we had utilized the same test data on those networks.

Facebook

To assess Facebook censorship, we measured the censorship status of the top three hundred porn sites to see if they would be blocked on Facebook Messenger. Usually, users will receive warnings from Facebook if they send messages with information in a blacklist. To measure the censorship of Facebook, we recorded whether Facebook blocked each porn site in a private chat session.

Google

To collect censorship data from Google, we utilized a list of sensitive keywords for measuring censorship on five country-specific versions of Google; these keywords may have been added to a blacklist by Google. The purpose of this test was to verify whether these keywords were blocked by five country-specific versions of Google.

Twitter

Collecting censorship data from Twitter is more complicated. We utilized hundreds of sensitive keywords as the test data on Twitter, with eight types of sensitive keywords, and a total of 392 (Miller 2012). We collected the censorship data from Twitter by searching these keywords on Twitter and recorded whether they appeared in search results.

Reddit

We used the Twitter method for collecting the censorship data on Reddit. In this test, we searched the same test data (i.e., 392 sensitive keywords) on Reddit for exploring censorship. The aim of this test is to verify whether Reddit blocked such posts or subreddits that related to these keywords. We also recorded the results of this test.

Imgur

To collect the censorship data from Imgur, we utilized the Twitter test data (i.e., 392 sensitive keywords) and adopted the same method as well. We regarded certain keywords as blocked information on Imgur if no search results were returned. Finally, we recorded all blocked and unblocked keywords on Imgur.

Limitations of the Data Collection

Keyword searching is the simplest approach to exploring censorship on social-network sites. This method is easy and fast; it can accurately seek out the blocked information by examining a variety of keywords on each social network. It also seldom generates errors, because we can get accurate information on social networks by using the search box on front pages. Note that we are not sure we can capture all blocked information by using this method, because it has limitations in collecting this kind of data. For example, Facebook and Google return a warning when searching using sensitive keywords on country-specific versions of Google and sending test data on Facebook Messenger sessions. However, this approach may not be suitable for verifying censorship on Twitter, Reddit, and Imgur, since when no search results are returned, we do not know whether the results were officially removed or whether users did not post such information. Future studies would need to investigate validation and triangulation of these results using other means.

Results

Facebook

To measure censorship on Facebook, we utilized the top three hundred porn sites as the test data for Facebook Messenger. Some fifty-nine (19.66%) of these were blocked, as shown in table 1.

Google

The primary purpose of this test was to discover the blocked information on different country-specific versions of Google. We thought Google may enforce two types of censoring activities on keywords; first, if no keyword suggestions were available, and second, if keyword suggestions were available, but no results were returned. We utilized 441 keywords as the test data in this test. We choose five countries: New Zealand, United States, Russia, Japan, and South Korea. We used a VPN to connect to the Internet outside New Zealand for collecting censorship data from these countries.

The results of this test were surprising. None of these sensitive keywords were blocked in New Zealand, the United States, Russia, and Japan. Only in South Korea were 31 of these sensitive keywords blocked (7%). Users can view keyword suggestions by partially entering them on Google Korea, but cannot access any of the results. Table 2 shows the blocked keywords in Google Korea.

Table 1. List of Blocked Porn Sites on Facebook

motherless.com	lesbianpornvideos.com	homesexdaily.com
sunporno.com	eroxia.com	free2peek.com
alphaporno.com	iyottube.com	vankoi.com
yobt.com	yourfreeporn.us	tubethumbs.com
pornsharia.com	sexoasis.com	maxjizztube.com
slutload.com	pornjog.com	PornTelecast.com
xxxbunker.com	pornupload.com	sexfans.org
pornhost.com	kosimak.com	koostube.com
fux.com	lubeyourtube.com	amateursex.com
kporno.com	141tube.com	cantoot.com
secret.shooshtime.com	my18tube.com	sextube.si
mofosex.com	afunnysite.com	xfanny.com
userporn.com	megaporn.com	sex2ube.com
jizzonline.com	adultvideodump.com	adultvideomate.com
pornotube.com	silverdaddiestube.com	porntubesurf.com
pornative.com	xfapzap.com	darkpost.com
boysfood.com	dirtydirtyangels.com	xxxpornow.com
lubetube.com	bigerotica.com	specialtytubeporn.com
al4a.com	isharemybitch.com	public-sluts.net
jizzbo.com	thegootube.com	

Table 2. List of Blocked Sensitive Keywords in Google Korea

asian babe	black cock	cuckold
autoerotic	blonde action	deep throat
ball kicking	blonde on blonde action	deepthroat
ball licking	brunette action	dildo
ball sucking	bukkake	double dong
bangbros	busty	double penetration
bareback	camslut	dp action
bbw	clit	eat my ass
beaver lips	courtney trouble	electrotorture
big knockers	cream pie	fantasies
big tits		

Imgur

Imgur is a site for hosting images. To exploring censorship on Imgur, we also collected censorship data from New Zealand, United States, Russia, Japan, and South Korea. Thus, we used the free VPN for collecting the censorship data from countries outside New Zealand.

In the testing process, we utilized 392 sensitive keywords with different categories for measuring censorship on Imgur. Finally, we discovered these five countries had blocked the same 73 keywords on Imgur. These blocked keywords represented 18.62% of the total. Table 3 shows all blocked keywords on Imgur.

Reddit

Reddit censorship appears to operate differently from other social networks, since it appears to involve blocking activities in a particular subred-

Table 3. List of Blocked Keywords on Imgur, with Categories

Categories	Keywords	Keywords
Domestic Security	Maritime domain awareness	National preparedness
	National preparedness	Disaster medical assistance team (DMAT)
	Disaster assistance	
	DNDO (Domestic Nuclear Detection Office)	Domestic nuclear detection
	Bomb(squad or threat)	Organized crime
Pornography	18onlygirls	Ktr
	czechav	Rarbg
	czechcasting	twistys
Cyber Security	DDOS (dedicated denial of service)	Mysql injection
	Conficker, Rootkit	Cyber terror
	Brute forcing	
Weather/Disaster/ Emergency	Mud slide or Mudslide	Tsunami Warning Center
	Stranded/Stuck	
Terrorism	IED (Improvised Explosive Device)	ETA (Euskadi ta Askatasuna)
	Al Qaeda (all spelling)	Basque Separatists
	AQIM (Al Qaeda in the Islamic Maghred)	Eco terrorism
	FARC (Armed Revolutionary Forces Colombia)	Conventional weapon
	IRA (Irish Republican Army)	PLF (Palestine Liberation Organization)
	TTP (Tehrik-i-Taliban Pakistan)	
Infrastructure Security	Airplane (and derivatives)	Cartel Methamphetamine
	CIKR (Critical Infrastructure & Key Resources)	
Smuggling (smugglers)	Failure or outage	Cartel de Golfo
	Computer infrastructure	Gulf Cartel
	NBIC (National Biosurveillance Integration Center)	
Michoacana	Transport security	Arellano-Felix
	Service disruption	Torreon
	Southwest Border Violence	Beltran-Leyva
	Ciudad Juarez, Kidnap	Barrio Azteca
	Reyosa	Narco banners (Spanish equivalents)
	Mara salvatrucha	Artistic Assassins
	Tamaulipas	Mexicles, U.S
	MS13 or MS-13	
HAZMAT & Nuclear	Biological infection (or event)	Suspicious package/device
	Suspicious package/device	Nerve agent
	Blister agent	Hazardous material incident
Health Concern + H1N1	Center for Disease Control (CDC)	Agro Terror
	Foot and Month (FMD)	Norvo Virus
	Water/air borne	Viral Hemorrhagic Fever
	H5N1	

dit. It was difficult to determine the blocked information on Reddit. In the testing process, we used the same test data (i.e., 392 sensitive keywords) that were utilized in Imgur. In Reddit, the keywords suggestions on the front page are some subreddits and users' posts.

This test is to discover the blocked subreddits and blocked posts by searching those 392 sensitive keywords on Reddit. None of these keywords was blocked by any of the five countries on Reddit, as shown in table 4.

Twitter

To measure censorship on Twitter, we utilized the same test data (i.e., 392 sensitive keywords) that were utilized in Imgur and Reddit. Just one sensitive keyword appeared to be blocked by the five countries, namely, the IRA (Irish Republican Army), which belongs to the terrorism category, representing a blocking rate of 0.26% (see table 5). As is shown in table 5, 2% of terrorism words were blocked by Twitter, while Twitter blocked no other keywords.

Discussion

The results of this study indicate that social networks widely interfere with private and personal communications in the absence of any legal or regulatory requirements to refrain from doing so, where blocking may be consistent with information-seeking-behavior access controls (Watters and Ziegler 2016). This seems inconsistent with the categorical imperative underpinning a free and open Internet. The vague notion that social networks apply "community standards" to filtering content deserves much closer scrutiny.

Some previous studies suggest that Facebook censors some content that violates any of its "community standards." This content could include users' political opinions, nudity, violence, human-right abuses, homosexual references, and advertisement. Facebook also censors website links and any other external information posted on Facebook (Sirichit 2015). These previous studies only examined Facebook censorship of users' posts and comments, not private conversations. We confirmed that censorship also exists in users' private conversations when sharing included approximately 20 percent of the top 300 pornography URLs. While censoring these URLs may seem unconcerning, the explicit right to free speech provided by the First Amendment in the US does not match Facebook's practices in this case. While the Facebook Terms of Service are a private matter, nonetheless, the proliferation of social media means that social networks provide a voice to political and social views that may be unpopular while being perfectly legal. Is it really the role of social-network moderators to decide what private communications can proceed, or not? Further regulation of social networks may be necessary to ensure that legal protections for free speech are respected.

Table 4. Statistics of Blocked Keywords on Reddit

Categories	Blocked by New Zealand	Blocked by the United States	Blocked by Russia	Blocked by Japan	Blocked by South Korean	Volume of blocked keywords	Frequency
Domestic Security	0	0	0	0	0	52	0.00%
Pornography	0	0	0	0	0	47	0.00%
Cyber Security	0	0	0	0	0	24	0.00%
Weather/ Disaster/ Emergency	0	0	0	0	0	39	0.00%
Terrorism	0	0	0	0	0	50	0.00%
Infrastructure Security	0	0	0	0	0	96	0.00%
HAZMAT & Nuclear	0	0	0	0	0	35	0.00%
Health Concern + H1N1	0	0	0	0	0	49	0.00%
Total	0	0	0	0	0	392	0.00%

Table 5. Statistics of Blocked Keywords on Twitter

Categories	Blocked by New Zealand	Blocked by the United States	Blocked by Russia	Blocked by Japan	Blocked by South Korean	Volume of blocked keywords	Frequency
Domestic Security	0	0	0	0	0	52	0.00%
Pornography	0	0	0	0	0	47	0.00%
Cyber Security	0	0	0	0	0	24	0.00%
Weather/ Disaster/ Emergency	0	0	0	0	0	39	0.00%
Terrorism	1	1	1	1	1	50	2.00%
Infrastructure Security	0	0	0	0	0	96	0.00%
HAZMAT & Nuclear	0	0	0	0	0	35	0.00%
Health Concern + H1N1	0	0	0	0	0	49	0.00%
Total	1	1	1	1	1	392	0.26%

By its own policy, Twitter forbids hate speech (Ammori 2014). Some previous studies suggest that Twitter aims to prevent users involved in spam and fraud (Newland et al. 2011). Users' accounts have been deleted by Twitter because they sent too much spam. Recently, Twitter began to censor revenge porn and further censor any expressions that may promote terrorism or racism (Petrovic, Osborne, and Lavrenko 2013). In this

study, censorship by Twitter was very weak, with only one sensitive keyword blocked across five countries.

Google appears to implement its censorship on "country-specific" versions of Google, but it may also censor any content in violation of the US Digital Millennium Copyright Act (Cobia 2009), with some requests coming from countries to remove political content or terrorism keywords on its search results (Rushe 2012). However, only 31 sensitive keywords were blocked by Google South Korea; no keywords were blocked by any other countries. Corporations need to evaluate whether they wish to operate in markets that encourage heavy-handed censorship of this kind.

Censorship of Reddit often involves blocking activities on certain subreddits that have already caused harm and received criticisms (Workman 2014). Usually, Reddit restricts users in certain behaviors like posting spam messages, canvassing activities, posting personal information, and posting child pornography and any other sexual content (Anderson 2015). Reddit has also been forced to censor some cases of "Karmawhoring" (collections of Karma points on some posts) because certain posts have no good content quality. However, our results indicate that Reddit does not appear to routinely censor information using keywords.

In summary, this study has provided evidence of a range of corporate censorship practices undertaken on public posts and private messages by the most popular social networks. Given the ubiquitous nature of social media, to what extent—and in which contexts—is it appropriate for private corporations to decide whether a communication is transmitted or not? This is a fundamental question for the community to answer; clearly, further regulation of the sector is one strategy, to ensure that there is better alignment between government censorship requests and the more discretionary aspects of corporate censorship. Rather than blocking communications thought to be inappropriate, perhaps the use of auto-Internet warnings could be a mechanism that is more consistent with situational crime-prevention theory (Prichard et al. 2015).

References

Ammori, Marvin. 2014. "The 'New' *New York Times*: Free Speech Lawyering in the Age of Google and Twitter." *Harvard Law Review* 127 (8): 2259–95.

Anderson, Katie Elson. 2015. "Ask Me Anything: What is Reddit?" *Library Hi Tech News* 32 (5): 8–11.

Aoyama, Ikuko, Shoka Utsumi, and Motohiro Hasegawa. 2012. "Cyberbullying in Japan: Cases, Government Reports, Adolescent Relational Aggression, and Parental Monitoring Roles." In *Cyberbullying in the Global Playground: Research from International Perspectives,* edited by Qing Li, Donna Cross, and Peter Smith, 183–201. Chichester: Wiley-Blackwell.

Bambauer, Derek. 2009. "Cybersieves." *Duke Law Journal* 59 (3): 377–446.

Ballard, Susan, and Pamela McKinlay. 2011. *Art at Risk: Copyright, Fair Dealings and Art in a Digital Age. Part 2.* Otago Polytechnic. https://ako.ac.nz/assets/Knowledge-centre/RHPF-s0912-Art-at-Risk/f2912a0bfe/Art-at-Risk-copyright-fair-dealing-and-art-in-a-digital-age-Part-two-Background-Research.pdf.

Blue, Violet. 2015. *The Smart Girl's Guide to Privacy: Practical Tips for Staying Safe Online.* San Francisco, CA: No Starch Press.

Chaabane, Abdelberi, Terence Chen, Mathieu Cunche, Emiliano De Cristofaro, Arik Friedman, and Mohamed Ali Kaafar. 2014. "Censorship in the Wild: Analyzing Internet Filtering in Syria." In *IMC '14: Proceedings of the 2014 Internet Measurement Conference*, 285–98. New York: Association for Computing Machinery. https://doi.org/10.1145/2663716.2663720.

Cobia, Jeffrey. 2009. "Digital Millennium Copyright Act Takedown Notice Procedure: Misuses, Abuses, and Shortcomings of the Process." *Minnesota Journal of Law, Science & Technology* 10 (1): 387–411.

Das, Sauvik, and Adam Kramer. 2013. "Self-Censorship on Facebook." In *ICWSM 2013: Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media,* 120–27. Palo Alto, CA: AAAI Press. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/viewPaper/6093.

Driscoll, Kevin, and Shawn Walker. 2014. "Working within a Black Box: Transparency in the Collection and Production of Big Twitter Data." *International Journal of Communication* 8:1745–64.

Duffy, Natalie. 2015. "Internet Freedom in Vladimir Putin's Russia: The Noose Tightens." Washington, D.C.: American Enterprise Institute. https://www.aei.org/wp-content/uploads/2015/01/Internet-freedom-in-Putins-Russia.pdf.

Frangonikolopoulos, Christos, and Ioannis Chapsos. 2012. "Explaining the Role and the Impact of the Social Media in the Arab Spring." *Global Media Journal: Mediterranean Edition* 7 (2): 10–20.

Franks, Alexandra. 2012. "Censorship of Legitimate Content on the Internet." PhD diss., University of Otago.

Ho, Wai H., and Paul A. Watters. 2004. "Statistical and Structural Approaches to Filtering Internet Pornography." In *SMC 2004: Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics*, 4792–98. Piscataway, NJ: Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/ICSMC.2004.1401289.

Jackson, Benjamin F. 2014. "Censorship and Freedom of Expression in the Age of Facebook." *New Mexico Law Review* 44 (1): 121–67.

Kopel, Karen. 2013. "Operation Seizing Our Sites: How the Federal Government is Taking Domain Names without Prior Notice." *Berkeley Technology Law Journal* 28:859–900.

Kwon, K. Hazel, Shin-Il Moon, and Michael A. Stefanone. 2015. "Unspeaking on Facebook? Testing Network Effects on Self-Censorship of Political Expressions in Social Network Sites." *Quality & Quantity* 49 (4): 1417–35.

Malamuth, Neil, Gert Hald, and Mary Koss. 2012. "Pornography, Individual Differences in Risk and Men's Acceptance of Violence against Women in a Representative Sample. *Sex Roles* 66 (7–8): 427–39.

Miller, Daniel. 2012. "Revealed: Hundreds of Words to Avoid Using Online If You Don't Want the Government Spying on You (and They Include 'Pork', 'Cloud' and 'Mexico')." *Daily Mail*, May 26, 2012. http://www.dailymail.co.uk/news/article-2150281/REVEALED-Hundreds-words-avoid-using-online-dont-want-government-spying-you.html

Mills, Richard, and Adam Fish. 2015. "A Computational Study of How and Why Reddit.com Was an Effective Platform in the Campaign Against SOPA." In *Social Computing and Social Media: 7th International Conference, SCSM 2015,* edited by Gabriele Mieselwitz, 229–41. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-20367-6_23.

Nabi, Zubair. 2014. "~~Resistance~~ Censorship is Futile." *First Monday* 19 (11).

Newland, Erica, Caroline Nolan, Cynthia Wong, and Jillian York. 2011. *Account Deactivation and Content Removal: Guiding Principles and Practices for Companies and Users.* Cambridge, MA: Berkman Klein Center.

Nunez, Michael. 2016. "Former Facebook Workers: We Routinely Suppressed Conservative News." *Gizmodo,* May 9, 2016. http://gizmodo.com/former-*facebook*-workers-we-routinely-suppressed-conser-1775461006.

Oh, Yunsang, Takashi Obi, Joong Sun Lee, Hiroyuki Suzuki, and Nagaaki Ohyama. 2010. "Empirical Analysis of Internet Identity Misuse: Case Study of South Korean Real Name System. In *DIM '10: Proceedings of the 6th ACM Workshop on Digital Identity Management,* 27–34. New York: Association for Computing Machinery. https://doi.org/10.1145/1866855.1866863.

Petrovic, Sasa, Miles Osborne, and Victor Lavrenko. 2013. "I Wish I Didn't Say That! Analyzing and Predicting Deleted Messages in Twitter." Unpublished manuscript, last modified May 14, 2013. https://arxiv.org/abs/1305.3107.

Prichard, Jeremy, Paul A. Watters, Tony Krone, Caroline Spiranovic, and Helen Cockburn. 2015. "Social Media Sentiment Analysis: A New Empirical Tool for Assessing Public Opinion on Crime?" *Current Issues in Criminal Justice* 27 (2): 217–36.

Richterich, Annika. 2014. "'Karma, Precious Karma!' Karmawhoring on Reddit and the Front Page's Econometrisation." *Journal of Peer Production* 4:1–12.

Roberts, Margaret E. 2018. *Censored: Distraction and Diversion inside China's Great Firewall.* Princeton, NJ: Princeton University Press.

Rushe, Dominic. 2012. "Google Reports 'Alarming' Rise in Censorship by Governments." *Guardian*, June 17, 2012. https://www.theguardian.com/technology/2012/jun/18/google-reports-alarming-rise-censorship.

Singer, P. W., and Allan Friedman. 2014. *Cybersecurity: What Everyone Needs to Know.* Oxford: Oxford University Press.

Sirichit, Methaya. 2015. "Censorship by Intermediary and Moral Rights: Strengthening Authors' Control over the Online Expressions through the Right of Respect and Integrity." Unpublished manuscript, March 25, 2015. https://works.bepress.com/methaya_sirichit/1/.

Sleeper, Manya, Rebecca Balebako, Sauvik Das, Amber Lynn McConahy, Jason Wiese, and Lorrie Faith Cranor. 2013. "The Post That Wasn't: Exploring Self-Censorship on Facebook." In *CSCW '13: Proceedings of the 2013 Conference on Computer Supported Cooperative Work,* 793–802. New York: Association for Computing Machinery. https://doi.org/10.1145/2441776.2441865.

Soha, Michael. 2012. "Networked, Collaborative, and Activist News Communities Online: A Case Study of Reddit and Daily Kos." PhD diss., University of Massachusetts Amherst.

Spacey, Rachel, Louise Cooke, Claire Creaser, and Adrienne Muir. 2015. "Regulating Internet Access and Content in UK Public Libraries: Findings from the MAIPLE Project." *Journal of Librarianship and Information Science* 47 (1): 71–84.

Stabile, Mark, Sarah Thomson, Sara Allin, Seán Boyle, Reinhard Busse, Karine Chevreul, Greg Marchildon, and Elias Mossialos. 2013. "Health Care Cost Containment Strategies Used in Four Other High-Income Countries Hold Lessons for the United States." *Health Affairs* 32 (4): 643–52.

Sung, Minkyu. 2017. "Surveillance and Anti-Communist Authoritarianism in South Korea." *Surveillance & Society* 15 (3/4): 486–90.

Thomas, Kurt. 2013. "The Role of the Underground Economy in Social Network Spam and Abuse." PhD diss., University of California, Berkeley.

Tushnet, Rebecca. 2000. "Copyright As a Model for Free Speech Law: What Copyright Has in Common With Anti-Pornography Laws, Campaign Finance Reform, and Telecommunications Regulation." *Boston College Law Review* 42 (1): 1–80.

Watters, Paul A. 2015. "Censorship is ~~Futile~~ Possible but Difficult: A Study In Algorithmic Ethnography." *First Monday 20* (1).

Watters, Paul A., and Jacqueline Ziegler. 2016. "Controlling Information Behavior: The Case for Access Control." *Behavior and Information Technology* 35 (4): 268–76.

———. 2017. "The Failure of Public and Private Sector Regulation in Cyber Security: A Singaporean Case Study." In *Cyber Security and Policy: A Substantive Dialogue,* edited by Andrew Colarik, Julian Jang-Jaccard, and Anurahda Mathrani. Auckland, NZ: Massey University Press.

Whyte, Kyle. 2017. "The Dakota Access Pipeline, Environmental Injustice, and US Colonialism." *Red Ink: An International Journal of Indigenous Literature, Arts, & Humanities* 19 (1): 1–22.

Workman, Hattie McClung. 2014. "Formation of Safe Spaces in Gendered Online Communities: Reddit and 'The Front Page of the Internet.'" Senior thesis, Texas Christian University.

Yangyue, Liu. 2014. *Competitive Political Regime and Internet Control: Case Studies of Malaysia, Thailand and Indonesia.* Newcastle, UK: Cambridge Scholars.

Zittrain, Jonathan. 2010. "The Fourth Quadrant." *Fordham Law Review* 78 (6): 2767–81.

Hanfeng Zhong recently completed his master of information sciences studies at Massey University in New Zealand. His research involves measuring private- sector censorship in social media. He now works in the information-technology industry in Hangzhou, China.

Paul A. Watters is adjunct professor of cyber security at La Trobe University in Australia and founder of the 100 Point Cyber Check TM. He has written extensively about internet censorship, restriction, and access control. He currently holds an Australian Research Council Discovery Grant (DP160100601) to investigate "Auto-Internet Warnings to Prevent the Viewing of Minor-Adult Sex Images" and also receives funding from the Australian Centre for Combating Child Exploitation.