CENSORSHIP OF LEGITIMATE CONTENT ON THE INTERNET

A REFLECTION ON THE NEW ZEALAND LAW COMMISSION'S PROPOSALS FOR DEALING WITH HARMFUL DIGITAL COMMUNICATIONS

Alexandra Franks

A dissertation submitted in partial fulfilment of the requirements for the degree of Bachelor of Laws (with Honours) at the University of Otago, Dunedin, New Zealand

October 2012

Acknowledgements

Thank you to my supervisor, Colin Gavaghan, for your help and enthusiasm.

Thank you to my parents for all your love, support and encouragement; and to Matthew for introducing me to LAWS101.

Thank you to my friends for putting up with hearing about this all year, and thank you in particular to Shelby, for your love, support, patience and formatting skills.

Table of Contents

I Introduction

> *We live in oppressive times. We have, as a nation, become our own thought police; but instead of calling the process by which we limit our expression of dissent and wonder "censorship", we call it "concern for commercial viability".*
>
> David Mamet, *Writing In Restaurants*[1]

Indirect censorship of the internet is hidden insidiously beneath other concerns and not labelled as censorship. Instead it masquerades as protecting intellectual property, protecting national security or protecting a reputation. It occurs because mechanisms for controlling online content are not open, transparent, sufficiently narrow or publicly accountable. I aim to establish that indirect censorship is a problem which has been overlooked in discourse about internet censorship; explain how methods of indirect censorship have been used to block "legitimate" online content; and discuss the New Zealand Law Commission's (NZLC) proposals for dealing with harmful digital communications in the context of censorship of legitimate content. In particular, I will discuss the "Pressure and Persuasion" method of censoring legitimate content, which relates to the Law Commission's proposal for an Approved Agency, and the "Pretext" method, which relates to the proposal for a "takedown" power. The substantive values of blocked content categories, or whether censorship should or should not occur, are beyond the scope of my topic. The concern of my dissertation is with how censorship mechanisms can be and are used in a way that enables indirect and advantageous censorship, and how the NZLC's proposals might reflect these methods.

"Censorship" in this context means blocking access to content and removal of content. "Content" includes an entire website, a page of a website, search engine listings, comments and videos. "Content creator" means the person who has created the corresponding content. The term "intermediary" will be used to refer to both Internet Service Providers

[1] David Mamet *Writing In Restaurants* (Penguin Books, New York, 1987).

(ISPs) and Online Service Providers (OSPs). "Legitimate" content is not illegal. It does not infringe copyright, is not defamatory, illegal obscenity, racial hatred, or a valid threat to national security. I propose that if content removal and/or blocking mechanisms are used narrowly to meet only their stated aims, "legitimate" content should not be affected. If a content removal/blocking mechanism is specifically created to target "legitimate" content, then by virtue of that content being designated as undesirable it is no longer "legitimate". A State could designate "political opposition" as something to be blocked as, for example, Burma,[2] China,[3] and Tunisia[4] do,[5] and regardless of the substantive values of that decision any content that fits that category will be "illegitimate" content. For the purposes of this dissertation it would only cause concern if content that could not be classified as "political opposition" was blocked or removed on that basis. This would be censorship of "legitimate" content.

[2] OpenNet Initiative "Burma (Myanmar)" (6 August 2012) OpenNet Initiative Research Profiles <http://opennet.net>.

[3] OpenNet Initiative "China" (9 August 2012) OpenNet Initiative Research Profiles <http://opennet.net>.

[4] OpenNet Initiative "Tunisia" (7 August 2009) OpenNet Initiative Research Profiles <http://opennet.net>.

[5] Robert Faris and Nart Villeneuve "Measuring Global Internet Filtering" in Ronald Deibert, John Palfrey, Rafal Rohozinski, and Jonathan Zittrain (ed) *Access Denied: The Practice and Policy of Global Internet Filtering* (MIT Press, United States of America, 2008) 1 at 9.

II Establishing Indirect Censorship as a Problem

A Introduction

The purpose of this part of my dissertation is to establish that indirect censorship on the internet is an issue that is often overlooked in favour of a focus on overt censorship. Overt censorship is, and should be, a global concern. However, criticisms of overt censorship are incomplete without also considering indirect censorship. Indirect censorship is a concern because its use has increased as regulation of the internet has become more sophisticated. As Kreimer says: "proxy censorship of the internet […] is a growth industry of internet regulation."[6] The hidden nature of indirect censorship makes it potentially more dangerous to freedom of expression than overt censorship, as people at least know overt censorship is occurring and can respond accordingly. Critics of overt censorship who employ indirect censorship methods themselves undermine their criticism, incomplete as it is. Derek Bambauer sums up the concern with indirect censorship in the following quote:[7]

> Concern is not with Orwell's Oceania, with its overt control over communication, but instead with Orwell's Armchair, where the state eases people into a censored environment through softer, more indirect means.

B What is indirect censorship?

In *Orwell's Armchair*, Bambauer divides censorship mechanisms into "hard" and "soft" methods.[8] Hard censorship is where the State exerts direct control over the internet's infrastructure, or "forces" intermediaries to exercise direct control though law.[9] In comparison, soft censorship is the control of content through indirect means, such as

---

[6] Proxy censorship is censorship carried out by intermediaries, recognising that there are multiple points of control for the flow of information on the internet: Seth F Kreimer "Censorship by Proxy: the First Amendment, Internet Intermediaries, and the Problem of the Weakest Link" (2006) 155(11) U Pa L Rev .

[7] Derek E Bambauer "Orwell's Armchair" (September 2011) Social Science Research Network <http.ssrn.com> at 7.

[8] At 5.

[9] At 5.

"employing unrelated laws as pretext to block material, paying for filtered access, or persuading intermediaries to restrict content."[10] Bambauer adopts Lawrence Lessig's approach that more than law can influence behaviour, stating that soft censorship methods include constraining our actions through "architecture", market forces and social norms. "Architecture" is the "built environment" of cyberspace, the software and hardware that make the internet what it is and constitute a set of constraints on how one can behave.[11] Hard censorship involves using the law, while soft censorship is outside the law. Bambauer holds that hard censorship is the only legitimate type, applying his process-oriented legitimacy framework, and therefore the government should legislate for internet censorship.

While Bambauer's soft/hard division is attractive, I would argue there is significant overlap with "hard" forcing of intermediaries to exercise control and using the "soft" method of persuading them to exercise control. Both methods regulate intermediaries. Bambauer classes one as "direct" and the other "indirect", however, the hard method of deputising intermediaries also misdirects government responsibility, and the soft method of persuasion also deputises intermediaries. Lessig defines a direct control as one that tells people how to behave and threatens punishment for deviation from acceptable behaviour, and an indirect control as one that modifies another structure of constraint in order to reach the same end as the direct control.[12] His example of an indirect constraint modifying behaviour is indirect regulation of abortion by direct regulation of doctors' ability to offer that service.[13] From the patient's perspective it is the doctor that is preventing access to the service. From the point of view of the content creator, the intermediary is performing the blocking/removing action, regardless of whether the intermediary is legally obligated to do so or voluntarily agrees to do so at the behest of the government. The content creator is not directly punished for their

[10] Bambauer, above n 7, at 5.

[11] Lawrence Lessig *Code: And Other Laws of Cyberspace, Version 2.0* (Basic Books, United States of America, 2006) at 121.

[12] At 132.

[13] At 132: Lessig explains that *Roe v Wade* 410 US 113 (1973) recognised a woman's right to an abortion, making direct regulation of abortion unconstitutional. However, this does not prevent the government from preventing doctors in government-funded health clinics from discussing abortion with patients or restricting funding for abortions.

content, as would be characteristic of a direct control using Lessig's definition.[14] In both hard and soft censorship the intermediary acts as a structure of constraint through shaping that content (either through voluntary agreement or legal obligation). This means the hard/soft distinction does not tell the full story and it is not as simple as advocating for hard censorship, as Bambauer would do.

For the purposes of this section, "overt" internet censorship is typically accepted as being direct censorship which aligns with the hard method classification. Overt censorship involves direct government intervention that is not obscured as originating from a non-governmental source. Indirect censorship would then, on the face of it, align with the soft method classification. However, I have argued above that hard censorship can also be indirect censorship. What is clear, however, is that indirect censorship is never overt censorship as overt censorship does not obscure its origin.

The methods of indirect censorship that this dissertation will focus on are Pressure and Persuasion and the Pretext method. Pressure and Persuasion involves voluntary agreements between governments and intermediaries to block or remove content, negotiated in the shadow of the law. Pretext censorship involves use of pre-existing laws to block unrelated content.

C Overt internet censorship as a well publicised problem, particularly in relation to "Eastern" countries

Overt internet censorship is a well publicised issue as it is an obvious form of censorship. The most well known example is China's "Great Firewall".[15] Extensive filtering

[14] The original content uploader is not punished in the sense that any sanction ends with content removal/blocking. This does not account for those individuals who upload content which is removed or blocked and are then subject to some kind of prosecution for uploading that content. In the relevant examples of blocking/removal used in this dissertation, punishment does not typically go further than initial removal.

[15] China implements an extensive filtering system, blocking more than 300 IP addresses at the "international gateway level", meaning that access is blocked for all users within China regardless of the ISP an individual is subscribed to. China also filters by keywords appearing in the domain name or URL for a website: Faris and Villeneuve, above n 5, at 14-15.

and arbitrary censorship continue to be a focus of political discourse and condemnation, as such practices threaten freedom of expression affirmed in the First Amendment to the United States Constitution,[16] the International Covenant on Civil and Political Rights,[17, 18] the Universal Declaration of Human Rights,[19] and the New Zealand Bill of Rights Act 1990 (NZBORA).[20] The blatant nature of this form of censorship makes it an easy and popular target for political commentary. Countries that are the focus of this criticism fall broadly into the group of "Eastern" countries, comprising Asia and the Middle East. In her "Remarks on Internet Freedom" address, Hilary Clinton[21] named China, Tunisia, Uzbekistan, Vietnam and Egypt specifically as threatening the free flow of information that is characteristic of internet use.[22] Thomas Melia[23] told a conference on internet freedom in Dublin that "too many governments were filtering, censoring content, taking down sites and perpetuating internet shut downs".[24] Reporters Without Borders names 12 countries as "enemies of the internet", including China and Syria, and 14 countries as "under surveillance" in their 2012 Internet Enemies Report.[25]

---

[16] The protection of freedom of speech in the First Amendment is much stronger than in New Zealand's Bill of Rights Act 1990 s 14. The First Amendment protects hate speech (Frederick Schauer "The Exceptional First Amendment" (February 2005) Social Science Research Network <www.ssrn.com>), whereas in New Zealand, freedom of expression can be limited where to do so is demonstrably justified in a free and democratic society per s 5 NZBORA.

[17] International Covenant on Civil and Political Rights 999 UNTS 407 (opened for signature 19 December 1966, entered into force 23 March 1976), art 19.

[18] New Zealand is a party to the ICCPR: United Nations "Status International Covenant on Civil and Political Rights" (8 July 2012) United Nations Treaty Collection <http://treaties.un.org>.

[19] UN General Assembly *Universal Declaration of Human Rights* GA Res 217 A, III (1948), article 19.

[20] Bill of Rights Act 1990, s 14.

[21] Secretary of State of the United States of America.

[22] Hilary Clinton, Secretary of State of the United States of America "Remarks on Internet Freedom" (speech to The Newseum, Washington DC, 21 January 2010).

[23] Deputy Assistant Secretary of State of the Bureau of Democracy, Human Rights and Labor, United States of America.

[24] Genevieve Carbery "Governments 'filtering, censoring' content" *Irish Times* (online ed, Ireland, June 18 2012).

[25] Reporters Without Borders "Internet Enemies Report 2012" (12 March 2012) Reporters Sans Frontieres <http://march12.rsf.org./en/>.

The OpenNet Initiative (ONI) tests the extent of internet filtering occurring globally by identifying blocked URLs.[26] 59% (41/74) of countries tested by ONI over the period 2008-2011 engage in some form of filtering.[27] ONI's focus has been on direct, overt censorship, with testing for filtering of four specific categories: political content, social content,[28] use of internet tools (such as filtering circumvention tools and blogging platforms) and conflict/security.[29] However, the ONI testing methodology provides only a "snapshot of accessibility to a limited subset of the internet for a limited number of countries."[30] In particular, the testing methodology is unable to account for intermediary censorship, which is a partial focus of this dissertation through the "pressure and persuasion" censorship method.

Criticism of these countries has become standard and hardly revolutionary. I do not mean to say that the fact that censorship in these particular countries is well publicised means that people should stop drawing attention to their censorship practices, but the focus on established and well known censorship regimes obscures other important issues. For a criticism of any type of internet censorship to be complete, it should include indirect censorship.

D Failure to identify and publicise "Western" censorship

While publicising "Eastern" censorship with a focus on universal internet access, condemning extensive filtering and protesting the creation of localised internets is important, "Western" democracies fail to address the indirect and collateral censorship occurring in their

[26] In each country to be tested, ONI tests access to a 'global' list of popular websites, which may have controversial content, and also tests access to a country specific 'local' list of websites across multiple content categories. Tests are completed a number of times to account for variations in access at any given time and are carried out at multiple locations within the country: OpenNet Initiative "ONI Methodology, Tools, and Data FAQ" OpenNet Initiative <http://opennet.net/oni-faq>.

[27] OpenNet Initiative "Filtering Data" (8 November 2011) OpenNet Initiative <http://opennet.net/research/data>.

[28] Social content is content which is against societal norms.

[29] OpenNet Initiative "ONI Methodology, Tools, and Data FAQ", above n 26.

[30] OpenNet Initiative "ONI Methodology, Tools and Data FAQ".

own countries.[31] This is demonstrated by Google's Transparency Report, which provides bi-annual statistics on government requests for removal of political content. The countries that feature highly in sending these requests are those that are not typically associated with censorship, as shown in figures 1 and 2.[32]

July to December 2011

	Court Orders			Other Requests (Executive, Police, etc.)		
Country	Requests	Compliance	Items	Requests	Compliance	Items
Brazil	128	69%	397	66	26%	157
United States	117	40%	3,851	70	44%	2,341
Germany	60	80%	1,304	43	72%	418
Argentina	39	97%	247	7	86%	19
Turkey	22	64%	104	23	48%	70
Italy	20	70%	45	8	50%	51
Spain	18	78%	24	25	8%	283
France	12	67%	22	19	47%	39
United Kingdom	12	58%	97	37	54%	750
Switzerland	7	100%	65	1	100%	1

Figure 1. Top ten senders of Court Order requests

[31] See Bambauer above n 7, at 1, who says "America has begun to censor the Internet", and Derek Bambauer "Guiding the Censor's Scissors: a Framework to Assess Internet Filtering" (August 2008) Selected Works of Derek Bambauer <http://works.bepress.com/derek_bambauer> at 2.

[32] Dorothy Chou "More transparency into government requests" (18 June 2012) Google Official Blog <http://googleblog.blogspot.com>.

July to December 2011

	Court Orders			Other Requests (Executive, Police, etc.)		
Country	Requests	Compliance	Items	▼ Requests	Compliance	Items
India	5	80%	9	96	26%	246
South Korea	–	–	–	94	80%	249
United States	117	40%	3,851	70	44%	2,341
Brazil	128	69%	397	66	26%	157
Germany	60	80%	1,304	43	72%	418
United Kingdom	12	58%	97	37	54%	750
Spain	18	78%	24	25	8%	283
Turkey	22	64%	104	23	48%	70
France	12	67%	22	19	47%	39
Canada	4	50%	137	15	67%	25

Figure 2. Top ten senders of "Other Requests"

However, it should be noted that the higher numbers of requests coming from these countries compared to countries known to use extensive censorship could be precisely because those countries with censorship regimes already in place have their own mechanisms for content removal or blocking. China, for example, would not be sending content removal requests to Google because Google only has a 19% market share in China.[33] However, the fact that countries such as China would not feature highly in the Transparency Report because they have their own mechanisms of censorship does not undermine the point that the Report demonstrates censorship on the part of countries that have not been the focus of censorship condemnation. It should be noted that a high number of requests does not indicate abuse, as these requests could be legitimate. Low compliance from Google suggests that the items requested were not illegal or infringing from Google's point of view and in this respect low compliance is a better indicator of the nature of the requests, particularly as failure to comply with a legitimate request may expose Google to liability depending on the nature of the requested content.[34] However, using compliance as a measure of legitimacy of removal requests still involves making assumptions as to why compliance was low.

[33] OpenNet Initiative "China", above n 3.

[34] For example, if Google is alerted to defamatory material, failure to take action will not result in liability in the United States of America as s 230 of the Communications Decency Act (1996) provides immunity. However, if

The low profile role the United Kingdom's censorship scheme, Cleanfeed, played from 2004-2008 demonstrates failure to publicise western censorship while condemning overt censorship.[35] After discussing condemnation of filtering in China, Lilian Edwards notes that manipulation of internet content by private actors has "been widely disparaged in the West, yet until the Wikipedia incident in 2008 the role of the IWF was largely unknown except by a few industry and civil society commentators".[36] The "Wikipedia incident" Edwards refers to occurred when ISPs who had agreed to block content designated by the Internet Watch Foundation as falling within specific categories inadvertently blocked the ability to edit Wikipedia.org for all UK users. The block occurred because album art on one Wikipedia page involved a sexual image of an apparent child.[37] This incident drew attention to the existence of the Cleanfeed system to UK internet users who were largely unaware of it.[38] Edwards additionally referred to the Cleanfeed system as representing what could be "the most perfectly invisible censorship mechanism ever invented",[39] because of its lack of openness, transparency and public accountability.

Failure to identify indirect censorship as threatening the free flow of online information undermines condemnation of censorship practices and presents a weakness in arguments that censoring countries should adopt United States of America style First Amendment protections for speech, as if they will see those systems working in the United

---

Google is informed of copyright infringing content on their services and fails to take action, they are unable to engage the "safe harbour" liability prevention provisions in the Digital Millennium Copyright Act. Potential liability for inaction therefore depends on the type of content that is subject to removal request.

[35] This censorship scheme involves blocking content deemed by the Internet Watch Foundation as falling into certain content categories (child pornography, general obscenity, racial hatred). ISPs who agree to implement Cleanfeed automatically block the blacklisted content: Lilian Edwards and Charlotte Waelde (ed) *Law and the Internet* (3rd ed, Hart Publishing Ltd, Oxford, 2009) at 652-654.

[36] At 653.

[37] At 655.

[38] Dawn C Nunziato "How (Not) to Censor: Procedural First Amendment Values and Internet Censorship Worldwide" (2011) 42 Georgetown Journal of International Law 1123 at 1154.

[39] Lilian Edwards "From child porn to China, in one Cleanfeed" (2006) 3(3) SCRIPT-ed at 174.

States and be encouraged to eventually adopt them themselves.[40] To illustrate this point, in response to a United States report on human rights violations, China released its own report, *Human Rights Record of the United States in 2011*, which said that the US imposes fairly strict restrictions on the internet, and its approach "remains full of problems and contradictions" and that because of this "internet freedom" is just an excuse for the US to impose diplomatic pressure on other countries.[41] This perceived hypocrisy could prevent countries subject to condemnation for censorship practices from taking those criticisms on board.

E The role of the evolution of the internet in promoting awareness of indirect censorship

The evolutionary course of the internet demonstrates a sequential path of increased regulation that has led to norms of content filtering and taking down content rather than protecting information. In turn, these norms directly feed into the concept of indirect censorship combined with incentives of ISPs. The internet freedom movement has been a key factor in promoting awareness of the censorship which is the concern of this dissertation. The focus has not shifted from overt censorship to indirect censorship, but is expanding to include both. ONI has classified the evolution of the internet into four periods which build on each other.

---

[40] See Nunziato, above n 38, at 41. She argues that First Amendment style procedural values should be adopted internationally for protection of speech, and non-American countries are more likely to adopt these procedural values than substantive values. She argues that regardless of substantive content censored by any individual country, first amendment values should apply. She argues that other countries will more readily accept procedural values than substantive values. Also see Julie Adler "The Public's Burden in a Digital Age: Pressues on Intermediaries and the Privatization of Internet Censorship" (2011) 20 JL & Pol'y 231, where the argument is made that if the United States sets a good example for protecting citizens' digital rights, other countries will eventually follow that example.

[41] Mu Xuequan "China issues report on human rights in the US" (25 May 2012) Xinhuanet English News <http://news.xinhuanet.com>.

1 "Open Commons" 1960-2000

In response to the proposed Stop Online Piracy Act (SOPA) in the USA, a "manifesto" was posted online on behalf of the "internet generation", which stated, "the internet to us is not something external to reality but a part of it: an invisible yet constantly present layer intertwined with the physical environment."[42] This quote, claiming to represent a generation of internet users, demonstrates how the internet is no longer seen as a separate "space" that a user goes to ("going online"), as was idolised during the "Open Commons" phase of internet access and content regulation.[43] The predominant view of internet regulation was that there could be none, as whatever rules constrained people physically, no government could constrain the "virtual selves" living in cyberspace.[44] This view of the internet (described as "cyber-anarchist")[45] is seen most clearly in John Barlow's "Declaration of Independence for Cyberspace", where he says: [46]

> Governments of the Industrial World, you weary giants of flesh and steel, I come from Cyberspace, the new home of Mind. On behalf of the future, I ask you of the past to leave us alone. You are not welcome among us. You have no sovereignty where we gather.
>
> We have no elected government, nor are we likely to have one, so I address you with no greater authority than that with which liberty itself always speaks. I declare the global social space we are building to be naturally independent of the tyrannies you seek to impose on us. You have no moral right to rule us nor do you possess any methods of enforcement we have true reason to fear.

---

[42] Piotr Czerski "We, the Web Kids." (15 Feburary 2012) Pastebin <http://pastebin.com/0xXV8k7k>.

[43] Ronald Deibert , John Palfrey, Rafal Rohozinski, and Jonathan Zittrain "Toward the Fourth Phase of Cyberspace Controls" in John Palfrey Ronald Deibert, Rafal Rohozinski, and Jonathan Zittrain (ed) *Access Contested: Security, Identity and Resistance in Asian Cyberspace* (MIT Press, Cambridge, 2011) 3 at 7.

[44] Lessig, above n 11, at 302.

[45] Judge David Harvey *internet.law.nz* (3rd ed, Lexis Nexis, Wellington, 2011) at 60.

[46] John Perry Barlow "Declaration of the Independence of Cyberspace" (8 February 1996) Electronic Frontier Foundation Projects <http://projects.eff.org>.

This excerpt from the declaration illustrates the view of the internet as a separate space, where government interference is invalid and regulation impossible. The "myth of openness" which characterised this lack of regulation is now, for some, an aspirational model for the future of the internet.[47] The NZLC recognises this view as still held by some,[48] but discounts it as no longer representing reality.[49]

2 "Access Denied" 2000-2005

In this period the view of the internet as a separate space was eroded. [50, 51] In New Zealand, *O'Brien v Brown* held that there was no extra allowance for defamatory statements posted online merely because they are part of a separate internet "culture".[52] Filtering of content became a growing method of control buoyed by more assertive government intervention as the risks of the internet received increased publicity, with a particular focus on the dangers of pornography and children's unrestricted access to the internet. This phase marked the beginning of a global norm of content filtering.[53] In contrast to the "Open Commons" phase, government intervention was increasingly seen as legitimate and necessary. This phase was also characterised by "mission creep",[54] where filtering systems adopted for one reason were then used for additional classes of content.

---

[47] Lessig, above n 11.

[48] Law Commission *Harmful Digital Communications: The adequacy of the current sanctions and remedies* (Ministerial Briefing Paper 2012) at 3.7.

[49] At 3.8.

[50] Deibert, Palfrey, Rohozinksi and Zittrain, above n 43, at 8.

[51] But see Harvey, above n 45, at 66, who continues to expound the view of the internet existing in a separate space: "the internet exists in a virtual world, cyberspace, rather than in the real or geographical world." However, Judge Harvey focuses on the "internet" as being in a separate space, not the *internet user* going to a separate space. Judge Harvey then goes onto say "the fact of the matter is that 'virtual' actions are grounded in the real world," at 175.

[52] *O'Brien v Brown* [2001] DCR 1065 at 7.13.

[53] Deibert, Palfrey, Rohozinksi and Zittrain, above n 43, at 10.

[54] At 9.

3 "Access Controlled" 2005-2010

This period was characterised by non technological methods of shaping content, such as registration and identity requirements promoting self censorship,[55] combined with selective filtering at increasing points of control.[56] During this period "just in time blocking", where an authority makes an informal request of a private company for content removal, gained popularity.[57] This period emphasised shaping and controlling access to information rather than outright blocking.

4 "Access Contested" 2010-present

"Open Commons", "Access Denied" and "Access Controlled" demonstrate a pattern of increasingly sophisticated control. Combined with the vast growth of the internet, this pattern has led directly to the current period, classified as "Access Contested".[58] The "free" nature of the "Open Commons" period led directly to the "Access Denied" assertive intervention against perceived dangers of the prior open nature, as the more sophisticated filtering of the "Access Controlled" period built on "Access Denied". The growing position that government intervention on the internet is now desirable, or at least inevitable, has meant that contests over internet access and regulation between governments and advocates for a free/open internet are now more apparent. There is a stark difference between the claim of the "Open Commons" phase that the internet could not be regulated and law was therefore meaningless, and the current dilemma of the "Access Contested" phase, which is not if the internet can be regulated (the "Access Contested" period accepts that it can and has been), but how it should be regulated and whose values should guide regulation.[59] ONI summarises the current nature of the internet as being a "crisis of authority", reflecting fundamental disagreement over all aspects of the space.[60] This is especially clear in respect of moves

[55] At 10.

[56] See Kreimer, above n 6, for a discussion of different points of control.

[57] Deibert, Palfrey, Rohozinski and Zittrain, above n 43, at 12.

[58] At 14.

[59] At 17.

[60] At 35.

toward increasing territoriality of the internet.[61] As all aspects of the internet are up for debate due to the contested nature of the space, ONI argues that the "lid is lifted on the internet, allowing for a closer examination of what goes on beneath the surface",[62] allowing for an expansion in focus from overt censorship to include indirect censorship. Indirect censorship is by its nature obscured by other concerns such as security or intellectual property enforcement, which have been key drivers of content regulation.

Showing a culmination of the contest between advocates for increased regulation and advocates for internet freedom, a coalition of websites and organisations launched the Declaration of Internet Freedom on 2 July 2012. The Declaration promotes establishing and defending five basic principles: expression, access, openness, innovation and privacy.[63] This Declaration represents a conscious choice to craft the constitution of the internet. The Declaration is sparse, focusing on "principles not policy",[64] because it is intended to act as a stepping-stone to promote open discourse and collaboration in the contest over internet access and control. This Declaration represents a point of difference because it is intended to engage with indirect censorship. The goal of the Declaration is to promote discussion and encourage people to engage policy makers over internet freedom issues.[65]

F Conclusion

The purpose of this section was to establish that indirect censorship is a problem that has been overlooked due to a focus on overt censorship. Overt censorship is by its nature an obvious form of censorship, making it a target of condemnation. The inherent nature of indirect censorship means that it is obscured by other concerns. It is therefore harder to

---

[61] See Harvey, above n 45, at 66; OpenNet Initiative "North Korea" (10 May 2007) OpenNet Initiative Research Profiles <http://opennet.net> ; Samuel Blackstone "Iran Plans To Stop Using The Internet By 2013" (9 August 2012) Business Insider <www.businessinsider.com>.

[62] Deibert, Palfrey, Rohozinksi and Zittrain, above n 43, at 35.

[63] "Declaration of Internet Freedom" (2 July 2012) Internet Declaration <www.internetdeclaration.org>.

[64] Mike Masnick "Can't We All Get Along: Principles Over Policy; Ideas Over Ideology" (6 July 2012) Techdirt <www.techdirt.com>.

[65] TC Sottek "The Declaration of Internet Freedom: how the net's minutemen plan to protect the future" (2 July 2012) The Verge <www.theverge.com>.

identify (as it is hidden until one goes looking for it), harder to assess and harder to understand. Indirect censorship is not open, transparent, or publicly accountable. Criticisms of internet censorship are incomplete without also identifying and addressing indirect censorship. The "Access Contested" nature of the internet has enabled identification of indirect censorship as an issue and it is now a focus of the wider internet freedom movement.

III Incentives of Intermediaries Relating to Content Removal

A Introduction

Indirect censorship methods draw on incentives of intermediaries in order to achieve a result. In relation to the two methods discussed in this dissertation, Pressure and Persuasion draws on the incentives of intermediaries to avoid liability even when under no legal obligation to censor content. The Pretext method draws on intermediaries' propensity to overblock also in the context of being liability-shy. My argument is that the way that law interacts with the incentives of intermediaries instils a norm of content removal and overblocking. This has occurred through conscious choices over where liability should fall.

B Divergent costs and benefits of intermediaries

Intermediaries have a low commitment to content that they facilitate, whereas users have a strong commitment to content they create and access. The dominant incentives of intermediaries are to protect themselves from liability rather than protect any one user's content. Content creators may receive benefits from posting, such as increasing their reputation in a community, revenge, fulfilling social obligations to give back to a community from which they have benefitted (for example, someone who receives helpful information from a review site may feel obliged to post their own reviews to further help someone else) and self expression which has intrinsic value without any further benefit.[66] The costs to the user are low – all that is necessary is an internet connection and then any time/resources spent in creating content. If the user's content is removed/blocked, they lose all these perceived benefits whereas the intermediary receives none of these benefits in the first place and therefore has far lower incentive to protect a user's content. The benefits an intermediary may receive from any one user are revenue, in the form of a subscription fee if one is set, advertising revenue from a user's site/profile and increased hits. However, if the intermediary blocks that user's content they lose practically nothing because that user is only one of many, instead gaining valuable protection from liability. The intermediary may suffer a cost in the

[66] Felix T Wu "Collateral Censorship and the Limits of Intermediary Immunity" (2011) 87(1) Notre Dame L Rev 293 at 305-306.

sense of damage to their reputation, but overall the incentive to censor content will overcome any lost revenue from lost customers.[67] The benefit of censorship outweighs the costs, whereas for a user, the benefits of expressing themselves online outweigh the minimal costs in doing so.

C Incentive to remove without assessing complaint

The threat of notice-based liability creates an incentive to remove content without assessing a complaint's merits.[68] The law of defamation incentivises intermediaries to remove content as soon as they receive a complaint. A plaintiff alleging defamation must establish that a defamatory statement was made about him/her, and the defendant "published" this statement.[69] The New Zealand High Court has applied the *Byrne v Deane* test for liability to online publication,[70] which is whether, having regard to all the facts of the case, the proper inference is that "by not removing the defamatory matter, the defendant really made himself responsible for its continued presence".[71] The factors for liability under a *Byrne* analysis are knowledge of the existence of a defamatory statement, the means to control the existence/display of the statement and an "apparent unwillingness", deduced from the circumstances, to end that state of affairs.[72] The law therefore incentivises intermediaries to remove allegedly defamatory statements as soon as they are bought to their attention, as it is the "apparent unwillingness" to remedy the situation which results in liability for publication. In *Godfrey v Demon Internet*, removal 10 days after notification was not fast enough.[73] Intermediaries are therefore incentivised to remove content immediately upon notification. In

---

[67] See Christian Ahlert, Chris Marsden and Chester Yung "How 'Liberty' Disappeared from Cyberspace: The Mystery Shopper Tests Internet Content Self-Regulation" (2004) at 12 for further discussion of incentives of intermediaries to take down.

[68] Wu, above n 66.

[69] *Laws of New Zealand* Defamation (online ed) at [82]. There is no presumption that material appearing online has been published per *Al Amoudi v Brisard* [2006] 3 All ER 294 (QB) and *Nationwide News Pty Ltd v University of Newlands* [2005] NZCA 317, however third party access will be relatively easy to prove through server logs showing hits.

[70] *Solicitor-General for New Zealand v Siemer* HC Auckland CIV-2008-404-472, 8 July 2008.

[71] *Byrne v Deane* [1937] 1 KB 818.

[72] *Sadiq v Baycorp (NZ) Ltd* HC Auckland CIV 2007-404-6421, 31 March 2008 at [50].

[73] *Godfrey v Demon Internet* [1999] EMLR 542.

the United Kingdom, intermediaries who are informed of the existence of defamatory content but do not remove it cannot rely on the defence in s 1 of the Defamation Act 1996,[74] and are therefore incentivised to remove content on notification.[75]

The Digital Millennium Copyright Act also incentivises intermediaries to remove content without assessing a complaint's merits.[76] When an intermediary receives a notice of allegedly infringing content, they must remove that content in order to avail themselves of the "safe harbour" immunity in the Act. If they do remove content in good faith that is later found to not be infringing, they will not be liable for doing so.[77] Content removal is emphasised over content protection. The DMCA does have a put back provision, as a person may file a counter notice and have their content restored 10-14 days after takedown if a lawsuit is not initiated. The notice and takedown system in the EU e-commerce directive does not have a similar provision, and "there have been many claims that the EU regime creates incentives for ISPs to remove items first without even bothering to ask questions afterwards".[78] Although the DMCA has a put back provision, this criticism can also be levied at it, as an intermediary does not have to "ask questions afterwards". Nothing is required of the intermediary unless a counternotice is filed. The EU directive holds that an intermediary

---

[74] Defamation Act 1996 (UK), s 1

> In defamation proceedings a person has a defence if he shows that—
>
> (a) he was not the author, editor or publisher of the statement complained of,
>
> (b) he took reasonable care in relation to its publication, and
>
> (c) he did not know, and had no reason to believe, that what he did caused or contributed to the publication of a defamatory statement.

[75] *Godfrey v Demon Internet*, above n 73.

[76] Digital Millennium Copyright Act 17 USC.

[77] See Appendix 1 for relevant provisions of the DMCA.

[78] Tyler Moore and Richard Clayton "The Impact of Incentives on Notice and Take-down" (13 June 2008) Computer Laboratory, University of Cambridge <www.cl.cam.ac.uk> at 4.

will not be liable for hosting information if they "expeditiously" remove it upon actual knowledge of illegal activity.[79]

The "Mystery Shopper" test clearly demonstrates intermediaries' propensity to remove content without assessing the merits of a complaint.[80] In this test, experimenters complained to ISPs about public domain text from Mill's "On Liberty" found online, posing as the John Stuart Mill Heritage Foundation, which does not exist.[81] The UK ISP removed the content immediately upon notification. The US ISP did not, but not because they recognised the content as public domain or the complaint as vexatious. The complaint sent by the experimenters was not in the prescribed form required by the DMCA, because the experimenters were not prepared to file a false DMCA complaint "under penalty of perjury". Had they been, the content would have been removed.[82]

D Incentive to overblock

Related to the incentive to immediately remove content upon notification, the law incentivises intermediaries to overblock content. Felix Wu explains the incentive to screen out borderline content:[83]

> The problem with harnessing the power of intermediaries by imposing liability on them is that the fear of liability may induce intermediaries to block or eliminate too much content, including desirable lawful content […] The incentive is to screen out any type of marginal content, any content that significantly increases risk of liability. The marginal content excluded is likely to include a substantial amount of lawful speech.

---

[79] Directive 2000/31/EC art 14(1)(a), 14(1)(b).

[80] Ahlert, Marsden and Yung, above n 67.

[81] The experimenters used free webhosts and created websites hosting the "On Liberty" text anonymously so that it would not appear that they were complaining about content on their own sites.

[82] The DMCA requires a specific formulation to be included in takedown notices. See Appendix 1 for the relevant provisions of the DMCA relating to the notice and takedown regime, and Appendix 2 for an example takedown notice.

[83] Wu, above n 66, at 300.

If something specific needs to be blocked, for example one section from a larger, more complex website, an intermediary will often respond by blocking the entire site. This takes less time and is less costly. Even if the intermediary is committed to targeting only illegitimate content, "collateral damage" occurs because broad overblocking is easier to do than specifically targeted blocking.[84] The incentive to remove content immediately upon notification means there is pressure to act quickly, making overblocking more attractive. The NZLC's emphasis throughout the issues paper, briefing paper and draft Bill on a speedy response does not recognise the line to be drawn between a fast response and a more mediated response. Requiring a speedy response further enforces the norm of content removal. The NZLC is coming from the point of view of speed being a positive thing in these harm scenarios, which it undeniably is in respect of legitimate victims, but speed can be a bad thing too when it decreases caution and encourages overblocking. The difficulty of distinguishing legitimate content from illegitimate content, uncertainty of the judicial process and inherent expense involved in litigation mean that the intermediary is likely to "abandon the effort to avoid errors and adopt a conscious policy of prophylactic self censorship that blocks any content that could precipitate the threat of sanctions."[85] The propensity of intermediaries to err on the side of caution and deliberate misuse of this propensity by governmental actors is a concern for the UN Special Rapporteur on Freedom of Expression.[86]

The NZLC states that intermediaries will only be required to take "reasonable steps" to remove content. This could represent recognition of the propensity to overblock. However, in the context of the issues paper commentary it appears that this is only included in order to recognise that an ISP's powers are limited and they might not be able to specifically target content, not to recognise that in receiving an order they might be incentivised to overblock and therefore should only take "reasonable steps" to avoid this outcome. As the uncertainty of the judicial system is one aspect which incentivises intermediaries to overblock, a requirement to take only "reasonable" steps may not help the situation as there may be

[84] "Collateral censorship" is defined as refusal to distribute content out of fear of liability: Wu at 296.

[85] Kreimer, above n 6, at 28.

[86] Frank La Rue *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression* A/HRC/17/27 (2011)43.

uncertainty as to what is "reasonable", and overcautious intermediaries may therefore still overblock.

E Conclusion

The way that law interacts with the incentives of intermediaries has resulted in a norm of content removal rather than content protection, which has become a value of the internet. Intermediaries are incentivised to be over cautious, remove without assessing the merits of a complaint and overblock. The NZLC's proposals further instil these values because the Law Commission would allow takedown orders to be made against intermediaries regardless of legal responsibility for the content in question. If an intermediary receives a takedown order, they will be accountable for failing to comply if "clear notice has been given to them of exactly where the material is located and the content of it", and "they do not do all that is reasonable to remove the material".[87] Failure to comply will result in a fine not exceeding $5000.[88] This incentivises removal of content. However, this is not actually a problem when a takedown order comes from a judicial source that has considered the associated risks to freedom of expression, as the Tribunal making a takedown order should have done. The point is that in incentivising intermediaries to remove content, a conscious choice has been made further embedding a value of content removal.

[87] Law Commission, above n 48, at 5.87.

[88] Communications (New Media) Bill 2012, cl 22(2).

IV Censorship Assessment Framework

Bambauer has identified problems with assessing censorship using substantive values due to contradictions in any one State's criticism of another State's censorship practises. In particular, any non-biased assessment of censorship was unable to adequately account for censorship by democratic countries. Bambauer highlighted this problem with the following example: "American government officials criticise search engines when they help censor political speech in China, and when they fail to censor copyrighted material there."[89] In *Cybersieves*, he developed a "process-oriented" framework for assessing the legitimacy of internet filtering, in order to alleviate the problems faced by value led criticism.[90] If a State's censorship is "legitimate" on Bambauer's analysis, then criticism cannot be made of that State's censorship processes and instead should be directed at policy.

Bambauer's framework focuses on four elements of filtering to test legitimacy:[91]

1. Openness: does a State admit that it censors content and why?
2. Transparency: is the State transparent about what content is filtered, and specific about criteria used to determine blockage?
3. Narrowness: how closely does what the State say it blocks (under Transparency heading) match what is blocked in reality? Does the State over or under block?
4. Public Accountability: is there any way for public involvement in decisions of what to block, and is there any aspect of appeal/recourse after something has been blocked?

The framework was developed specifically for application to filtering. However, I will be applying it as a test for legitimacy of internet censorship generally. Filtering is not the main focus of this dissertation as the NZLC has not recommended implementation of a filtering system to deal with harmful digital communications.

---

[89] Derek E Bambauer "Cybersieves" (2009) 59(3) Duke LJ 377 at "A Series of Filtered Tubes".

[90] At "A New Hope".

[91] At "A New Hope".

The United Nations Special Rapporteur on Freedom of Expression emphasises that article 19 of the ICCPR is applicable to the internet as it was drafted with foresight to accommodate developments in exercising freedom of expression.[92] Any restriction to the right of freedom of expression must meet the criteria in article 19(3). The restriction must be provided by law, necessary and the least restrictive means. If censorship is legitimate applying Bambauer's framework, then these conditions for limiting freedom of expression will be met.

[92] Frank La Rue, above n 86, at 21.

V The Law Commission's Recommendations to deal with "Harmful Digital Communications"

A Introduction

This section will provide an overview of the NZLC's proposals in its issues paper, *The News Media meets 'New Media': Rights, Responsibilities and Regulation in the Digital Age*, and ministerial briefing paper, *Harmful Digital Communications: The adequacy of the current sanctions and remedies*. The proposals provide the context for considering censorship of legitimate content on the internet from a New Zealand perspective and will be discussed in greater detail in the sections that follow.

B Issues paper

As part of its issues paper on news media regulation, the NZLC was asked to report on "speech harms" arising from internet use. The specific question the NZLC was asked to address in its terms of reference was:[93]

> Whether the existing criminal and civil remedies for wrongs such as defamation, harassment, breach of confidence and privacy are effective in the new media environment and if not whether alternative remedies may be available.

The "new media environment" is the read/write web where any user can create and publish content. "Speech harms" are those wrongs embedded within the question, as well as the specific concern of "cyberbullying". The NZLC's preliminary conclusion was that these harms are able to cause "significant harm" when combined with the unique characteristics of the internet.[94] Factors the NZLC identified as amplifying potential harm were Web 2.0 social media,[95] search engine popularity,[96] anonymity,[97] and permanence of information. Judge

[93] Law Commission *The News Media Meets 'New Media': Rights, Responsibilities and Regulation in the Digital Age* (NZLC IP27, 2011) at 7.6.

[94] At 7.172.

[95] At 7.12.

Harvey additionally notes the one-dimensional nature of internet communication and relative cheapness of interaction as factors that facilitate harassment online.[98] These characteristics mean that speech harms have unique potential to cause "significant psychological harm".[99]

The nature of the internet compounds the problem of our slow legal system. The NZLC emphasises that victims of speech harms are unable to rely on the courts for effective remedies due to an inability to access justice and the inevitable delay involved in using court processes. Speedy removal of content is typically a victim's main aim.[100] Proposing remedies that allow for a rapid response is a key thread throughout the issues paper, illustrated by repeated emphasis on the importance of "speedy, efficient and relatively cheap justice"[101] and "quick and efficient justice in a more informal manner".[102] To meet these aims, the NZLC proposed giving the courts a statutory power to make "takedown" orders, or establishing a Communications Tribunal with this takedown power. The NZLC also proposed establishing a Communications Commissioner, with informal persuasive powers, as an alternative.

C Briefing paper

The NZLC's proposals for dealing with speech harms were only one part of the wider issues paper focused on "New Media", discussed above. However, concern over cyberbullying has caused this part of the project to be fast-tracked, resulting in a ministerial briefing paper, *Harmful Digital Communications: The adequacy of the current sanctions and remedies*, to accompany a Communications (New Media) Bill. As this paper focuses solely on "harmful digital communications" (the favoured term over "speech harms") and responds to, and incorporates, submissions on the issues paper, it is more detailed.

---

[96] At 7.13.

[97] At 7.15.

[98] Harvey, above n 45, at 352.

[99] Law Commission *The News Media Meets 'New Media'*, above n 93, at 8.38.

[100] At 8.38.

[101] At 8.44.

[102] At 8.49.

Throughout the paper, the NZLC continues to emphasise the importance of a speedy response. The underlying thread for justifying the proposals in the Bill is that victims of harmful digital communications find that a response to harm is not fast enough and are frustrated and sometimes further harmed by a lack of response.[103] This paper expands on the issues paper's claim that access to justice is a barrier for victims, citing the reasons for this as lack of knowledge about the law and possibility of redress, complexity of identifying defendants, breadth and speed of the spread of information on the internet and problems establishing jurisdiction.[104] NetSafe is quoted as submitting, "other than when police officers act informally, it is unlikely the law can produce redress in a timeframe that is effective."[105]

Refining the proposals outlined in the issues paper, the NZLC proposes the establishment of a Communications Tribunal with various powers, including a power to order content takedowns. The Communications Commissioner idea has not been adopted, however the NZLC proposes allowing for an "Approved Agency" which would carry out the functions that were initially proposed for a Commissioner. The briefing paper recommends that NetSafe become an approved agency.

D Indirect censorship and the Law Commission

The issues paper does not discuss indirect censorship, having approached the problem of speech harms with an understandable bias toward identifying solutions, given the terms of reference. The issues paper does not discuss the potential censorship risk involved in these recommendations and mentions censorship only as a contrast to speech harms, as seen in the statement: "censorship is not the only enemy of free speech. Those who exercise their free speech to intimidate, bully, denigrate and harass others on the internet lessen the credibility of free speech arguments."[106] In neither the issues paper nor the briefing paper, does the Commission discuss censorship risk as a potential societal harm to any level compared with the discussion of harm caused by speech harms. Neither paper addresses the significant

[103] Law Commission *Harmful Digital Communications*, above n 48, at 3.59.

[104] At 5.2.

[105] At 5.9.

[106] Law Commission *The News Media Meets 'New Media'*, above n 93, at 7.5.

problem faced by users who have their content blocked or removed maliciously through advantageous use of censorship mechanisms. Failure to address this censorship based harm makes the NZLC's considerations incomplete. From the NZLC's perspective, reporting mechanisms and OSP terms of use are not useful in removing unwanted content. However, I take the opposite view and aim to show that sometimes these mechanisms are too effective as they result in the removal of legitimate content as well as targeted illegitimate/harmful content. The briefing paper does discuss freedom of expression,[107] and the Communications (New Media) Bill states that in exercising its functions, the Tribunal must have regard to the importance of freedom of expression.[108]

E Law Commission proposals will embed values in the "constitution" of the internet

As the focus of this dissertation is censorship of legitimate content, the NZLC's discussion and proposals are only relevant to the extent that they impact upon this type of censorship. To this end, the NZLC's emphasis on speedy, informal resolution of problems has the potential to increase censorship of legitimate content. The NZLC treats the possibility of a fast takedown process as only a positive thing, whereas this dissertation aims to demonstrate the problems associated with notice and takedown schemes, related to the overall claim that legitimate censorship is open, transparent, sufficiently narrow and publicly accountable.

In his book, *Code and Other Laws of Cyberspace*, Lawrence Lessig argued that cyberspace needs a constitution, structuring the limits of legal power and protecting fundamental values, in order to preserve the "liberty" present during the open commons period.[109] He argued the nature of the internet is not innate liberty, but is set by the architectures used which embed certain values. By architectures, he meant the software and hardware that makes the internet how it is at any given point in time. The architectures used during the "Open Commons" phase allowed for values of expression, openness and innovation, but as architectures change, the values change too. The constitution is built by the

---

[107] Law Commission *Harmful Digital Communication*, above n 48, at 5.80.

[108] Communications (New Media) Bill 2012, cl 16(4).

[109] Lawrence Lessig *Code and Other Laws of Cyberspace* (Basic Books, United States of America, 1999).

"invisible hand of cyberspace", which at the time of writing his book, Lessig said was "building an architecture quite the opposite of what it was at cyberspace's birth".[110] Lessig's constitution argument was about how the internet should be architected through conscious choices, to protect some values over others, because "left to itself, cyberspace will become a perfect tool of control", obviously a bad thing from his point of view.[111]

Without any direct intervention to protect fundamental values, the "constitution" of cyberspace has become one of over efficient control. The internet has developed in such a way that overbroad filters are normal and acceptable, and a norm of content removal rather than content protection has developed. If the values of the "Open Commons" period are important, then a "constitution" needs to be consciously crafted in order to restore those values. This is relevant to the NZLC's proposals as if they are implemented, they will embed values in the internet so we must be aware of what those values are, and identify what causes them. Judge Harvey also notes, citing Marshall McLuhan, that "we shape our tools and thereafter our tools shape us."[112]

[110] At 6.

[111] At 6.

[112] Harvey, above n 45 at preface V.

VI The Law Commission's Proposal for an "Approved Agency" and the "Pressure and Persuasion" Censorship Method

The purpose of this section is to discuss the "Pressure and Persuasion" method of censorship and provide examples of this occurring. The NZLC's proposal for an "Approved Agency" will then be discussed in the context of this method. Pressure and Persuasion will then be assessed for legitimacy using Bambauer's process-oriented framework.

A "Pressure and Persuasion" censorship method

1 Definition

Pressure and Persuasion involves an intermediary being persuaded/pressured to enter a voluntary agreement to censor content. However, the persuasion is often accompanied by a threat that if a voluntary agreement is not entered into, a censorship scheme will be legislatively mandated. Pressure is also used to attain removal of content without having a voluntary agreement in place, called "just in time blocking", because it is used advantageously to attain removal at key moments.[113] Pressure and Persuasion is also reflected in the emerging tactic of law enforcement targeting ISPs with "requests" that they take down content.[114] New Zealand police use this tactic, as police submitted to the NZLC: "when the goal is to have offensive content taken down from a website, rather than to initiate a prosecution, some social media sites will respond after receiving a formal request on police letterhead."[115]

2 Soft/hard classification

Pressure and Persuasion is a "soft" method of censorship because it is a non-legal measure. It is an indirect method and misdirects State responsibility.

---

[113] Deibert, Palfrey, Rohozinksi and Zittrain, above n 43, at 12.

[114] Kreimer, above n 6, at 76.

[115] Law Commission *The News Media Meets 'New Media'*, above n 48, at 7.168.

3 Examples

The typical example is a voluntary censorship agreement, negotiated under the threat of a legislatively enforced scheme. The implementation of the Internet Watch Foundation's 'Cleanfeed' blocklist, discussed above in the context of Western censorship being overlooked, fits this. Many ISPs voluntarily implemented this blocklist. Following a press release that stated that all UK ISPs must implement the blocklist or the government would legislate, the rest followed suit.[116] Edwards explains the problem with this approach, and "pressure and persuasion" as a censorship method generally:[117]

> This censorship needs no laws to be passed, no court to rule, with the publicity that entails. It only needs the collaboration, forced or otherwise, of ISPs. ISPs are not public bodies; their acts are not subject to judicial review.

Nunziato claims that the UK government maintains that this system is voluntary, privately implemented, and does not involve government coercion.[118] She describes the consequence of this approach:[119]

> As a technical matter, the government […] is not technically mandating such speech restrictions and because such restrictions are "voluntarily" undertaken by private ISPs at the behest of the government in cooperation with "private" organisations like the IWF, these speech restrictive actions are technically outside the scope of applicable national laws protecting citizens' free speech rights.

A further example clearly fitting the voluntary agreement/pressure of legislation trope is the "six strikes" copyright enforcement agreement in the US. The Recording Industry Association of America began advocating voluntary agreements to fight music piracy from

---

[116] Lilian Edwards "Pornography, Censorship and the Internet" in Lilian Edwards and Charlotte Waelde (ed) *Law and the Internet* (Hart Publishing Ltd, Oxford, 2009) 623 at 653.

[117] Edwards, above n 35.

[118] Nunziato, above n 38, at 1138.

[119] At 1138.

2008.[120] The RIAA president stated that he was more interested in finding a marketplace way of resolving piracy than legislating to increase ISP responsibility for infringement.[121] The reasoning behind this approach can be seen in a comment from the Directors Guild of America that "litigation is slow and the internet is fast."[122] This comment also reflects the NZLC's emphasis on the need for a quick remedy, as the legal system is unable to work in "internet time".[123] The culmination of this occurred in 2011, when major US ISPs agreed to a "six strikes" copyright enforcement plan to be implemented mid 2012. This was a voluntary agreement, however the negotiations were heavily supported by the Obama administration,[124] and multiple sources report that the threat of legislation was used to promote agreement.[125]

In Pennsylvania, legislation required ISPs to block access to websites designated by the State Attorney General as providing access to child pornography. This does not represent pressure/persuasion in the sense above, as receiving an order to block content pursuant to use of a statute is not a non-legal measure of censorship and would instead be hard direct censorship. However, the ISPs did not receive an order, and were instead persuaded to block access to 1.1 million sites (targeting 400 child pornography sites) because of a letter from the

---

[120] Annemarie Bridy "Graduated Response and the Turn to Private Ordering in Online Copyright Enforcement" (2010) 89 Or LR 81.

[121] Anne Broache "RIAA: No need to force ISPs by law to monitor piracy" (30 January 2008) CNET News <http://news.cnet.com>.

[122] Bridy, above n 120, at 81.

[123] Law Commission *Harmful Digital Communications*, above n 48, at 5.6.

[124] Matthew Lasar "Big Content, ISPs nearing agreement on piracy crackdown system" (24 June 2011) Ars Technica Law & Disorder <http://arstechnica.com/tech-policy>; David Kravets "U.S. Copyright Czar Cozied Up to Content Industry, E-Mails Show" (14 October 2011) Wired Threat Level <www.wired.com/threatlevel>; Mike Masnick "Worst Kept Secret Now Confirmed: Government Was Very Involved Helping RIAA/MPAA Negotiate Six Strikes" (14 October 2011) Techdirt <www.techdirt.com>.

[125] Jason Mick "Obama Conscripts ISPs as 'Copyright Cops', Unveils 'Six Strikes' Plan" (8 July 2011) Daily Tech <www.dailytech.com>; Mike Masnick "White House's New Report on Intellectual Property Enforcement Should Get A Copyright As A Creative Work Of Fiction" (30 March 2012) Techdirt <www.techdirt.com>; Derek E Bambauer "Orwell's Armchair" (September 2011) Social Science Research Network <http.ssrn.com>, above n 7, at 31.

Attorney General *threatening* to invoke the statute.[126] One ISP refused and demanded a judicial order. The Attorney General issued a press release accusing that ISP of aiding and abetting paedophilia.[127] Similarly, an attempt by the New York Attorney General to get ISPs to censor alleged child pornography on Usenet groups resulted in ISPs dropping the Usenet service entirely.[128]

The pressure of compliance through the threat of a negative reputation is also used. In Utah, a proposal to label ISPs as "community conscious" if they refused to publish obscene content was met with approval.[129] The clear implication is that the proposal would coerce ISPs into appearing pro-active against pornography through "voluntarily" implementing filters as otherwise their reputations would be tarnished by not having a "community conscious" label. However, some users may be drawn to an ISP that does not have this label.

A final example of Pressure and Persuasion involves WikiLeaks. Amazon stored WikiLeaks on its cloud service, and was contacted by a senator, which resulted in Amazon terminating its relationship with WikiLeaks, citing unspecified violations of terms and conditions as the reason for doing so.[130] Paypal and Mastercard stopped processing donations to WikiLeaks after a letter from a state department legal adviser. According to Bambauer, it is doubtful whether the government could have obtained a court order to these ends.[131] This situation shows informal pressure achieving what formal legal action likely could not have, and therefore what hard, direct censorship could not have achieved.

B How advocating for an "Approved Agency" aligns with "Pressure and Persuasion"

[126] Kreimer, above n 6, at 31.

[127] At 32.

[128] Bambauer "Orwell's Armchair", above n 7, at 31-32.

[129] At 29.

[130] Adler, above n 40, at 4.

[131] Bambauer "Orwell's Armchair", above n 7, at 27.

This section will explain how the NZLC's proposal for an "Approved Agency" reflects and draws on the Pressure and Persuasion method. The initially proposed Communications Commissioner would have had the role of resolving speech harms informally through contacting intermediaries to draw attention to questionable material and acting as an "early warning system" for website administrators.[132] This was seen as a viable proposal for harm resolution because a Commissioner would have the advantage of being a recognised authority figure, and would be able to build up a relationship with intermediaries. The perceived advantage of this was that the Commissioner's requests for content removal would likely be treated more seriously than any other individual's requests (such as the victim), advancing the aim of speedy resolution of harms. This draws on Pressure and Persuasion because the basis for a Commissioner being able to effectively resolve complaints is that, as a person of authority, a Commissioner will be able to persuade an intermediary to block or remove content without a direct legal measure being necessary. The request coming from a person of authority ties into the liability-shy nature and low commitment to user content of intermediaries discussed above. In the briefing paper, the Commissioner proposal was integrated into a recommended model encompassing an "Approved Agency" which would feed complaints to the Communications Tribunal, with the Tribunal acting as a final "backstop".[133] The Agency will have the same role and functions as the proposed Communications Commissioner would have had, without requiring a new structure to be established.[134]

The Agency's role is to receive complaints[135] and try to resolve them by a "process of negotiation, mediation and persuasion".[136] In particular, the Agency will have the function under cl 9(d) of the Communications (New Media) Bill to "liaise with website hosts and

---

[132] Law Commission *The News Media Meets 'New Media'*, above n 93, at 8.80 - 8.83.

[133] Law Commission *Harmful Digital Communications*, above n 48, at 5.36, 5.40.

[134] The Communications (New Media) Bill 2012 allows for the appointing of more than one agency, however the Briefing Paper recommends that there be only one approved agency, NetSafe (Law Commision *Harmful Digital Communications* at 5.107).

[135] See Appendix 3 for relevant provisions of the Communications (New Media) Bill relating to the Approved Agency.

[136] Law Commission *Harmful Digital Communications*, above n 48, at 5.45.

[ISPs] and, if appropriate, to request them to take down or amend posts that are clearly offensive".[137] The NZLC submits that an Agency acting as a filter is necessary because "many complaints will be much better handled by less formal means."[138] In the context of the NZLC's emphasis on speedy recourse for harms, one of the major advantages of a Tribunal is its ability to act quickly and informally. The subtext that an Agency may more easily resolve complaints because it is even less formal than the informal Tribunal seems out of place, and possibly undermines the necessity of a Tribunal.

When criticised that the Tribunal presents a risk to free speech, Steven Price responded that it will be very hard to get complaints to the Tribunal anyway.[139] This is because the Agency will only refer a complainant to the Tribunal if the complaint cannot be resolved at the Agency level, and the Agency considers it to be sufficiently serious. However, a complainant can make a complaint to the Tribunal even if the Agency has decided to take no action.[140] The Tribunal is thus used as a threat to encourage dispute resolution, illustrated with the following statement from the briefing paper: "Persuasion, with the possibility of tribunal proceedings in the background, should be an effective tool in many cases."[141] This clearly reflects Pressure and Persuasion. However, it could be said that this applies to any settlement negotiation, which will always occur in the context of the threat of litigation if an agreement cannot be met.

Steven Price's comment that it will be very hard to get a complaint to the Tribunal is probably correct. However, this is not necessarily a good thing from a free speech perspective, when it means complaints will instead be resolved at the Agency level. From the

---

[137] Communications (New Media) Bill 2012, cl 9(d).

[138] Law Commission *Harmful Digital Communications*, above n 48, at 5.103.

[139] New Zealand Law Foundation Centre for Law and Policy in Emerging Technologies "Cyberbullying: Do we need new law for cyber-bullies?" Public Discussion (University of Otago, 6 September 2012).

[140] Per cl 11, an Agency may decide to take no action on a particular complaint. However, Communications (New Media) Bill 2012, cl 14(3) states that a complaint must not be made to the Tribunal unless the Agency has considered the complaint and determined what action to take, which could be no action. This means the Agency can decide to take no action under cl 11, yet a complaint can still be made to the Tribunal.

[141] Law Commission *Harmful Digital Communications*, above n 48, at 5.103.

NZLC's point of view, resolution of complaints at the informal Agency level is a good thing because recourse to the Agency will be easily accessible, cheap, and relatively quick. In most cases, resolution of complaints at the Agency level *will* be a good thing because real harms will be mitigated. However, if a complaint gets to the Tribunal, the Tribunal must have regard to the importance of freedom of expression in exercising its functions, per cl 16(4) of the Bill. There is no express requirement on the Agency to do the same. The Bill also includes in cl 16(3) a list of things that the Tribunal must take into account in deciding whether to make an order, which are partially aimed at protecting freedom of expression, in particular factor (g), the extent to which the communication is in the public interest. There is no requirement for the Agency to take these same factors into account when attempting to resolve complaints. The Agency must take into account the communication principles set out in cl 7 when performing its functions.[142] However, none of the 10 principles refer to freedom of expression.

The protections on freedom of expression at the Agency level are therefore far less stringent than they are at the Tribunal level, presenting an inconsistency. With regards to the power of the Tribunal to order takedowns, the NZLC is clear that as this is effectively an injunction, "the requirements of [NZBORA] would have to be vigilantly observed".[143] However, in giving the Agency the function of liaising with intermediaries to request takedowns, the same freedom of speech risks exist, though in the Agency's case, with far less controls. Of course, an intermediary does not have to take anything down based on the Agency's request and not taking the content down will not be an offence. However, the whole point of the Agency is to attempt to resolve complaints so that they do not need to go to the Tribunal, and to set up an authority that intermediaries may take more seriously than the average user. This means that, at least in part, the NZLC is anticipating that the Agency's takedown requests will be followed. The fact that if the Agency cannot resolve the complaint it will go to the Tribunal (if sufficiently serious) means that an intermediary is incentivised to takedown content based on the Agency's request, as if they do not, they might then be subject

[142] See Appendix 3 for a list of the communication principles.

[143] Law Commission *Harmful Digital Communications*, above n 48, at 5.77.

to an order to takedown from the Tribunal. Based on the incentives of intermediaries section of this dissertation, I would argue that intermediaries are likely to respond to the Agency request to prevent further involvement.

Pressure and Persuasion works because intermediaries do not have incentives to protect speech. I am arguing that the reasoning behind the perceived usefulness of an Agency is linked to the this method, meaning the likely result of an Agency request for takedown will be a takedown. This is a concern because the draft Bill does not control the Agency's use of what is effectively the same takedown power that the Tribunal has. As this is the case, the Agency could potentially request takedown of "legitimate" material. There is no way to assess the likelihood of this actually happening, but the ability for them to do this is there due to the lack of controls in the Bill. The inconsistent freedom of expression controls in the NZLC's proposals and draft Bill therefore constitute a mechanism of censorship which could be used to remove "legitimate" content. However, NZBORA will apply to an Agency per s 3(b), as the Agency's functions will be imposed on it pursuant to law through the power in cl 8 of the Bill for a Minister to appoint an organisation as an Approved Agency. The Agency must not breach s 14 NZBORA, unless doing so is demonstrably justified in a free and democratic society per s 5.

The NZLC's proposals for the powers of the Tribunal also reflect Pressure and Persuasion, in particular the claim that the power to make a declaration that a communication breaches statutory principles "would have significant persuasive power, even if not mandatory authority, in relation to websites operating out of the jurisdiction."[144] The NZLC also says that overseas intermediaries "would [likely] regard such determinations as sufficient reason to take the required action."[145]

C Assessing "Pressure and Persuasion" using Bambauer's legitimacy framework

I have argued that indirect censorship is often overlooked as overt, direct censorship presents a more obvious threat to freedom of expression. Focus on overt censorship is able to

[144] At 5.77(e).

[145] At 5.40.

obscure indirect censorship. I claimed above that this was a problem as indirect censorship is also a threat to freedom of expression. Using Bambauer's legitimacy framework, I will show that Pressure and Persuasion is not legitimate and through the different elements of the framework explain why this is a problem.

I will discuss Pressure and Persuasion generally, referring to the examples above, and specifically in relation to the Agency.

1 Openness

Pressure and Persuasion fails Bambauer's openness requirement. The test for this was whether a State admits that it censors and why. Negotiations for voluntary agreements go on "behind closed doors", as do agreements for "just in time blocking". They are negotiated "if not in the shadow of the law, then in the threat of such shadow."[146] In some scenarios a government may publicise a voluntary agreement. However, in doing so, the government is unlikely to take responsibility for the censorship, as this defeats the purpose of having a voluntary agreement or privately persuading an intermediary to remove specified content. A lack of openness is what makes Pressure and Persuasion an attractive method of censorship to employ, as the government's involvement is invisible. This factor means that Pressure and Persuasion can be used to censor legitimate content, because technically, no one knows it is actually the government doing so, and it is justified as a private intermediary's free choice over what content is and is not allowed on their service.

Lessig discusses regulation of code as a way to indirectly regulate user behaviour, stating that government regulates behaviour "indirectly by directly regulating technologies that affect the behaviour. Those regulated technologies in turn influence […] the targeted behaviour directly."[147] Code, or the architecture of the internet, affects how an individual is able to use and interact with the internet. If a government regulates the architecture of the internet, for example by requiring user identification, it also regulates the user. However, the requirement for identification will seem to come directly from the website being used.

[146] Bambauer "Orwell's Armchair", above n 7, at 31.

[147] Lessig, above n 11, at 136.

Intermediaries are essential to being able to use the internet. If the government enters a voluntary agreement with an ISP that that ISP will filter the user's content, the user can no longer access that blocked content. However, the blockage seems to emanate from the ISP. This is "indirection", which Lessig argues is the enemy of transparency.[148] In the context of regulating code writing, Lessig argues that "government can achieve regulatory ends, often without suffering the political consequences that the same ends, pursued directly, would yield."[149] Lessig further argues, "in a constitutional democracy, [the state's] regulations should be public. And thus, one issue raised by the practice of indirect regulation is the general issue of publicity. Should the state be permitted to use nontransparent means when transparent means are available?"[150] This same problem arises with Pressure and Persuasion. Using this method can result in censorship that would not be achievable if pursued directly, such as what occurred in the WikiLeaks example above.

The openness test takes only the State's openness into account in testing legitimacy. However, Pressure and Persuasion is also not open in a secondary way, as intermediaries might not disclose censorship.

Approved Agency

The Pressure and Persuasion that the NZLC anticipates the Agency employing is open. This is because, assuming the draft Bill passes, the government will be admitting openly in public legislation that the Agency will have the function of requesting takedowns and the wider Bill is largely dedicated to allowing censorship in appropriate cases. The reason for censorship will also be publicly available as it is included in cl 3 of the Bill: "the purpose of this Act is to mitigate harm caused to individuals by electronic communications".

2 Transparency

As a method of censorship, pressure and persuasion cannot meet Bambauer's transparency requirement. The test for this was whether the State is transparent about what

[148] Lessig, above n 11 at 133.

[149] At 136.

[150] At 135-136.

content is filtered, and specific about criteria used to determine blockage. Transparency requires that the user be provided with information that content has been blocked, and why.[151] Pressure and Persuasion is the antithesis of transparency. In the case of a voluntary agreement, officially the intermediary will be deciding what to block and applying their criteria to content. The government is not responsible for applying the blockage criteria and this makes transparency hard to achieve. The fact that Pressure and Persuasion is not transparent is what makes it an attractive method of censorship. The advantage of "just in time blocking" is that when content becomes unavailable with no explanation, a user might assume that there is a problem with the server, or their connection, or any of a myriad of reasons. The site the user is trying to access will not transparently report that the content is blocked because of a governmental request. The State can avoid claims of censorship and display plausible denial, because content could truly be down for any number of reasons, especially when it is later returned (as is characteristic of "just in time blocking").

Approved Agency

Censorship can be open but not transparent and vice versa. The informal nature of the Agency may mean its use of this method, while "open", would fail under the transparency check. When the Tribunal makes a takedown order, its reasons for doing so must be published per cl 16(5) of the Bill. The conception of the Agency is to be even more informal than the Tribunal. There is no corresponding requirement on the Agency to publish reasons if they decide to persuade an intermediary to takedown content. Any negotiation between the Agency and an intermediary, while the Agency is trying to resolve a complaint before it gets to the Tribunal, will be private. In this respect there will be no transparency about content that is censored.

3 Narrowness

The test for narrowness was how closely does what the State say it blocks match what is blocked in reality. As Pressure and Persuasion fails the transparency element of the legitimacy framework, there is no way of knowing whether content blocked under voluntary

[151] Nunziato, above n 38 at 1129.

agreements, or through the use of the method generally, matches what is intended to be blocked. There is no way to assess whether content is over or under blocked as we do not know what was meant to be blocked in the first place. Based on anecdotal evidence such as the Pennsylvania child pornography example and dropping of all of Usenet and the incentives of ISPs, overblocking seems more likely. Lessig expresses the view that with regulation of speech through private action, a greater amount of content will be blocked than if "government acted wisely and efficiently".[152] Overblocking is a concern because it catches legitimate content.

Voluntary filtering agreements can also be subject to "mission creep", particularly as lack of openness and transparency mean that the public is largely unaware of filtering in the first place, so there is no reason why the public would be made aware of an extension to filtering goals. The IWF 'Cleanfeed' system was initially implemented to cover child pornography, however, it has been extended to "criminally obscene" content, hate speech and incitement to racial hatred.[153]

Approved Agency

I am unable to assess whether content that is taken down as a result of the Agency's requests matches the mischief that the draft Bill is aimed at as the Agency does not yet officially exist and has therefore not made any requests. However, the same point applies to the Agency that if the Agency's censorship is not transparent, there is no way to tell if what is censored in reality matches what the State says it blocks.

4 Accountability

The test for public accountability was whether there is any way for public involvement in decisions of what to block, and whether there is any aspect of recourse or appeal after content has been blocked. Pressure and Persuasion also fails here. Due to the nature of this censorship method, there is no way for public involvement in decisions of what

[152] Lessig, above n 11, at 255.

[153] Nunziato, above n 38, at 1148.

to block. When employing this method, government censorship efforts are shielded from judicial review because a private party makes the formal censorship decision. Perhaps the intermediary can become involved in the decision of what content to block and accountability would therefore flow between the State and the intermediary. However, as Bambauer notes, "measures that are voluntary for intermediaries become mandatory for users."[154] Any recourse after removal will only be through the intermediary, who does not have to follow natural justice. The problem with this is that when intermediaries are pressured into blocking or removing content, their actions can become a "de facto exercise of authority",[155] yet there is no way to get recourse.[156] However, with respect to the Cleanfeed example, the public can recommend URLs to be added to the blocklist. There is nothing innate about voluntary agreements that mean that the public cannot be involved in blocking decisions. However, with a government mandated direct censorship scheme there might be an expectation of public accountability, whereas with a voluntary scheme it is up to the intermediary to allow for public involvement and the possibility of appeal. As a private actor, an intermediary is not bound by laws upholding freedom of expression. For example, in a New Zealand context, NZBORA s 14 (affirming the right to freedom of expression) would prima facie not apply to an ISP per s 3, unless a court held that an ISP was performing a public function conferred pursuant to law. Individuals are less likely to challenge intermediary censorship in court, making Pressure and Persuasion a method which is more able to be "consciously manipulated to suppress protected speech."[157]

Users can make intermediaries accountable through taking their custom elsewhere. However, in order to make this choice, users need to know that censorship is occurring in the

[154] Bambaeur "Orwell's Armchair", above n 7, at 32.

[155] Deibert, Palfrey, Rohozinksi and Zittrain, above n 43, at 17.

[156] One avenue for recourse would be to apply the state action doctrine from United States jurisprudence, where United States courts look past the private entity who is implementing speech restrictions, and attribute these actions to the State when the State has "authorised or encouraged or become entangled with private unconstitutional conduct": Nunziato, above n 38. However, the concern of this dissertation is with the legitimacy of censorship mechanisms and therefore suggesting ways to make the governmental censor accountable are beyond the scope of my topic.

[157] Kreimer, above n 6, at 31.

first place. As Pressure and Persuasion is not open or transparent, this makes this option of market accountability less attainable, and the true target of accountability is the government anyway.

Approved Agency

The Agency's use of Pressure and Persuasion would not fail the accountability requirement with respect to deciding what to block, because the decision of what to block would initially come from a member of the public contacting the Agency. The functions of the Agency do not anticipate the Agency proactively seeking out harmful content. However, the second aspect of accountability, whether there is any ability to appeal or recourse after content has been blocked, would not be satisfied by the Agency's use of pressure and persuasion. If the Agency employs Pressure and Persuasion to encourage an intermediary to remove content, this is a "private" agreement. The resolution of the complaint will prevent it reaching the Tribunal. The draft Bill includes a right of appeal for a complainant who has had their complaint considered by the Tribunal.[158] It does not include a right of appeal for a defendant, though Steven Price considers this to be a mistake in drafting and not a deliberate choice.[159] As any agreement reached between the Agency and an intermediary will be beyond the law, there is no realistic ability for a content creator or user trying to access blocked/removed content to appeal the "free" choice made by the intermediary.

D Conclusion

Pressure and Persuasion is not open, transparent, narrow or publicly accountable as a general censorship method. It is therefore illegitimate, yet is increasingly employed to allow governments to remove content without suffering the consequences that doing so through hard, direct censorship methods would cause. The fact that Pressure and Persuasion is an illegitimate censorship method when assessed using the process-oriented framework is what

[158] Communications (New Media) Bill 2012 cl 18.

[159] Steven Price "Critiquing the Law Commission" (2 September 2012) Media Law Journal <www.medialawjournal.co.nz>.

allows it to be used to censor legitimate content. The examples given of use of the method demonstrate that it has been used in this way.

I have argued that the NZLC's proposal for an Approved Agency draws on the Pressure and Persuasion method in order to be effective in resolving complaints. However, the NZLC does not assess the problems with employing this method and do not explicitly state that they are drawing on this method. When assessing the Approved Agency's use of Pressure and Persuasion, I have found that this use will be open, not transparent and partially accountable. Assessment of narrowness is not possible. The Agency's use of pressure and persuasion therefore fails for overall legitimacy using Bambauer's framework.

VII The Law Commission's Proposal for a Takedown Power and the "Pretext" Censorship Method

A The "Pretext" censorship method

1 Definition

This is the justification of censorship by referring to an unrelated law. This is also called "mission creep" in the context of government mandated filtering systems, where systems adopted to deal with one content category (typically pornography) are then employed to deal with other public policy issues.[160] I would also include within this method the use of a law to remove content for a purpose of that law, but the fact that that purpose is being met is a side effect of the motivation to remove the content for another reason. For example, an individual could use copyright law to attain removal of a website which is critical of their work. That site may actually be infringing their copyright. By doing this, they are therefore protecting their intellectual property, which is a purpose of copyright law. However, the fact of protecting their intellectual property is irrelevant or secondary to the motivation of removing the critical content. In this scenario, the person is using copyright law advantageously to attain content removal.

2 Soft/hard classification

The Pressure and Persuasion method is different from Pretext because it involves non-legal measures (although it works as a method because of the threat of the law). Pressure and Persuasion is therefore a true "soft" method of censorship. The Pretext method works within the law, not outside it. It is therefore harder to classify with Bambauer's soft/hard distinction. However, Bambauer himself classifies it as a "soft" method. I would also classify it as a "soft" method. However, the problem I have with this classification is that I would include malicious or advantageous use of the DMCA notice and takedown scheme as Pretext censorship, and therefore a soft method. Bambauer takes the opposite view, saying "I have argued that the provisions of the DMCA that press intermediaries to censor in return for

[160] Deibert, Palfrey, Rohozinski, and Zittrain, above n 43, at 9.

immunity from copyright liability should be viewed as justified", thereby seeing this as hard censorship.[161]

3 Examples

I will be using the Pretext method to reflect on the NZLC's proposal that the Communications Tribunal be able to order takedowns. Therefore the following examples of Pretext censorship are divided into non-takedown examples and takedown examples.

3.1 Non "takedown" examples

Kentucky used gambling regulations stating that any illegal gambling device could be forfeited to the state to attain a seizure notice for 141 gambling domain names (registered and operated outside the United States).[162] The Governor justified the action by arguing that illegal gambling is a threat to Kentucky citizens.[163] However, the actual justification for the attempted seizure was a concern that internet gambling would decrease revenue from offline gambling,[164] as the Governor was affiliated with offline gambling interests and had promised as part of his campaign to bring more offline casinos to Kentucky.[165]

Another Pretext censorship example was the ordering of a domain name registrar to disable sites owned by a company that arranges travel to Cuba. [166] Most of the sites were not travel related.[167, 168] In Kazakhstan, internet security was the stated justification for blocking

---

[161] Bambauer "Orwell's Armchair", above n 7, at 10.

[162] The relevant law defined "gambling device" as a "slot machine" or "any other machine […] designed and manufactured primarily for use in connection with gambling.": Ky Rev Stat § 528.010(4).

[163] Bambauer "Orwell's Armchair", above n 7, at 19.

[164] At 19.

[165] Mike Masnick "Kentucky's Gambling Domain Name Grab Sets A Terrible Precedent" (10 October 2008) Techdirt <www.techdirt.com>.

[166] Travel to Cuba is not illegal, but is restricted as the Cuban Assets Control Regulations require US citizens to be licensed in order to "engage in any travel-related transactions pursuant to travel to, from, and within Cuba.": "Cuba Country Specific Information" (30 April 2012) International Travel <http://travel.state.gov>.

[167] Bambauer "Orwell's Armchair", above n 7, at 21.

20 sites said to promote terrorism and religious extremism.[169] However, one of the blocked sites was livejournal.com, a blogging platform. LiveJournal had previously been blocked in Kazakhstan from 2008-2011 with no official explanation, though Kazakh bloggers believed it was because of the posting of recordings of Government phone conversations.[170] Only one journal was specified as extremist, islamunveiled.livejournal.com, but the entire platform was blocked. There are 17,483 registered users of LiveJournal in Kazakhstan, however registration is not required to use the site.[171] In Turkey, Law 5651 allows a list of keywords to be banned from Turkish domains (aimed at pornography), and includes "free" and "pic".[172] While the inclusion of these words may just be lacking in common sense, the potential for abuse is high should someone want to employ Pretext as a censorship method. Regardless of intent, untargeted (in the sense of being non-pornographic) content will be censored. Law 5651 has been used on a Pretext basis for at least 197 politically motivated blocks.[173]

3.2 "Takedown" examples

Under the DMCA, a copyright owner or their representative can send a notice in the prescribed form to a site hosting allegedly copyright infringing content. The sender must state that they have a good faith belief that the use of the content specified in the notice is not authorised by the copyright owner or by the law. As discussed above, the intermediary

---

[168] See also Adler, above n 40, who describes a webhost that shut down an entire website containing a blog for the Belarussian American Association, as Belarus is subject to American trade sanctions; and shut down Zimbabwean human rights activism sites when American trade sanctions only apply to sites "undermining democratic institutions" in Zimbabwe.

[169] Reporters Without Borders, above n 25.

[170] Sergey Park "Казахстан: Блогеры обсуждают блокировку Живого Журнала" (2 September 2011) (translated ed: Adil Nurmakov (translator) Sergey Park "Kazakhstan: Bloggers Denounce Repeated Block of LiveJournal") Global Voices <http://globalvoicesonline.org>.

[171] "LiveJournal.com Statistics" (14 July 2012) LiveJournal Press Area: Statistics <www.livejournal.com>.

[172] Reporters Without Borders, above n 25.

[173] Blockable content categories are "encouragement of and incitement to suicide, sexual exploitation and abuse of children, facilitation of the use of drugs, provision of substances dangerous to health, obscenity, gambling, and crimes committed against Atatürk": Dr Yaman Akdeniz "Report of the OSCE Representative on Freedom of the Media on Turkey and Internet Censorship" (2009) Organisation for Security and Co-operation in Europe <www.osce.org> at 35.

receiving the takedown notice then has to take the content down for immunity. If they do take the content down, s 512(g)(1) holds that they will not be liable, even if the content is later shown to have not been infringing.[174] This clearly incentivises intermediaries toward content removal without assessing the merits of a complaint.

Google admits that it receives removal requests under the DMCA for legitimate, non-infringing content. Google's Transparency Report was updated in May 2012 to include data about requests to remove content from Search sent to Google. The report highlights the attempted use of DMCA takedowns as a censorship mechanism. In making this data available, Google draws attention to the potential abuses of content removal requests and the issues faced by intermediaries who receive them. In April 2012, Google processed over 1.2 million takedown requests for Google Search, targeting approximately 24,000 websites.[175] In the month June 6 – July the removal of 6,138,294 URLs from Google Search was requested.[176] Between July and December 2011, Google removed 97% of Search results specified in removal requests.[177] In the Transparency Report FAQ, Google lists examples of clearly invalid requests that it has received and states that it did not comply with any of the enumerated requests. However, the FAQ does not go as far as to say that Google never complies with invalid requests, just that it did not comply with those specifically listed invalid requests. With such a large volume of requests, Google is unlikely to assess the merits of each one.

In 2006, an independent report analysed takedown notices sent to Google and found that 57% of notices requesting removal of Search results were sent by businesses targeting competitors' appearance in search listings.[178] This does not mean that this 57% were

---

[174] See Appendix 1 for relevant provisions of the DMCA.

[175] Fred von Lohmann "Transparency for copyright removals in search" (25 May 2012) Google Official Blog <http://googleblog.blogspot.co.nz>.

[176] Google Inc "Google Transparency Report" <www.google.com/transparencyreport>.

[177] At FAQ.

[178] Jennifer Urban and Laura Quilter "*Efficient Process or "Chilling Effects"? Takedown Notices Under Section 512 of the Digital Millenium Copyright Act*" (2006) 22 Santa Clara Computer & High Tech LJ 6212.

illegitimate, as clearly some of the competitors may have been infringing the notice senders' copyrights. However, the sending of a takedown notice for a competitor's site does carry an implication of misusing copyright law to block unrelated content. An example of this is Universal Music Group who used the DMCA to takedown a YouTube video uploaded by their competitor Megaupload, advertising Megaupload's services, without Universal owning any copyright in the uploaded video.[179] However, YouTube (owned by Google) admits that it has no way of assessing videos for fair use so just removes everything requested, which obviously presents a huge risk of abuse.

3.2.1 DMCA takedowns used to block criticism

DMCA takedown notices are frequently used to stifle criticism. For example, a medical research firm used a takedown notice to disable the entire website of an animal rights organization, without specifying what was infringing their copyright.[180] The Church of Scientology famously used a DMCA takedown in 2002 to attain removal of Google Search results linking to a site critical of scientology.[181] In 2008 hundreds of YouTube videos critical of Scientology were the subject of a takedown notice, although they were not infringing of the Church's copyright.[182] Ten years after the first reported use of the DMCA to silence criticism by the Church of Scientology, the Church still advises its members to complain about negative commentary about the Church, whether through the use of a website's code of conduct (claiming that comments critical of scientology degrade a group of individuals by reason of religion), through claims of defamation, or through the use of the DMCA, claiming

[179] "Megaupload to Universal: You've Got Some Explaining To Do" (28 December 2011) Torrent Freak <www.torrentfreak.com>.

[180] Jason Mazzone *Copyfraud and Other Abuses of Intellectual Property Law* (Stanford Law Books, Palo Alto, 2011) at chapter 4.

[181] At chapter 4.

[182] It should be noted that YouTube realised that the overabundance of Scientology related takedown notices received over two days was suspicious and investigated the situation and restored most of the videos and reactivated suspended accounts: Eva Galperin "YouTube Anti-Scientology Takedowns: Good News, Bad News" (25 September 2008) Electronic Frontier Foundation <www.eff.org>.

that reproduction of Scientology materials infringes copyright.[183] Rush Limbaugh used a DMCA takedown in order to attain removal of a YouTube video criticizing his treatment of Sandra Fluke.[184] The video was a montage of controversial comments Limbaugh made about her (seven minutes out of nine hours of his talk show).[185] Universal Music used a DMCA takedown in order to attain removal of a video blog criticizing Universal artist Akon for his use of misogynistic lyrics.[186] In an article about the impact blogging can have on companies' reputations, Forbes magazine advocated using the DMCA notice-and-takedown regime and/or defamation laws to silence criticism. The magazine advised: [187]

> Find some copyrighted text that a blogger has lifted from your Web site and threaten to sue his Internet service provider under the [DMCA]. That may prompt the ISP to shut him down. Or threaten to drag the host into a defamation suit against the blogger. The host isn't liable but may skip the hassle and cut off the blogger's access anyway. Also: Subpoena the host company, demanding the blogger's name or Internet address.

As already discussed, the pattern of intermediary behaviour is to remove content. Not all intermediaries notify the blogger that the content has been removed. An average user may not be aware of the need to file a counternotice, or might not do so in the face of a lawsuit. Even if the blogger does file a counternotice, the content will still be down for 10-14 days. Although there may in reality be some copyrighted material on the blog or the comments made may actually be defamatory, this advice is aimed solely at "fighting back" and discrediting critical bloggers. Because the motive in following this advice is not actually

---

[183] Cory Doctorow "Scientology memo asks members to censor critical web comments with trumped up 'code of conduct' complaints" (7 July 2012) Boing Boing <http://boingboing.net>.

[184] Rush Limbaugh is a radio talk show host. Sandra Fluke is a law student who supported mandating insurance cover for contraceptives at a House of Democrats hearing. Limbaugh called her a "slut" and a "prostitute": Jack Mirkinson "Rush Limbaugh: Sandra Fluke, Woman Denied Right To Speak At Contraception Hearing, A 'Slut'" (29 Feburary 2012) The Huffington Post <www.huffingtonpost.com>.

[185] Matt Zimmerman "[Updated] Limbaugh Copies Michael Savage's Bogus Copyright Theory, Sends DMCA Takedown to Silence Critics" (24 April 2012) Electronic Frontier Foundation <www.eff.org>.

[186] At bulletpoint 3.

[187] Daniel Lyons "Attack of the Blogs" *Forbes Magazine* (New York City, online ed, 14 November 2005).

protecting copyright but censoring criticism, I would class this as use of the Pretext method.[188]

3.2.2 Overblocking and the DMCA

The Pretext method involves the use of laws that are already in force. Overblocking is a subset of the Pretext method because it too involves hard/direct laws already in place. Overblocking results in censorship of legitimate content because the laws that are in place for censoring content that public policy has decided should be censored are not narrow enough, such as the DMCA. Intellectual property law provides a basis for individuals to claim rights beyond what they actually possess, called "copyfraud".[189] However, overblocking is not always deliberate as it is a side effect of other methods of censorship. A blog post by techdirt.com discussing the negatives of SOPA was subject to a DMCA takedown sent on behalf of a pornography company. Nothing in the post infringed the company's copyright. In the same takedown notice, removal of search listings for a TorrentFreak article about the illegal seizure of the domain mooo.com was also requested. A link to an Independent Newspaper article about a cruise ship disaster was also listed in the notice as infringing.[190] The takedown notice also listed the pornography company's own website as infringing its own copyright.[191] Google complied with the takedown noticed and removed the Techdirt blog post from the Search index. Techdirt's post was later reinstated to the index after complaint, which illustrates clearly that the takedown was completely without merit, yet was initially successful.

---

[188] See also "Sleeping uneasy: Call-A-Mattress seeks removal of unfavorable review" (1 October 2012) Chilling Effects Clearing House <http://chillingeffects.org/N/618966> for a copy of a letter sent to a content creator who posted a negative review; and "letter re: defamatory comments on car enthusiast website" (27 August 2012) Chilling Effects Clearinghouse <http://chillingeffects.org/N/574178>.

[189] Mazzone, above n 180.

[190] "Media DMCA (Copyright) Complaint to Google" (20 January 2012) Chilling Effects Clearinghouse <http://chillingeffects.org/notice.cgi?sID=189468>.

[191] Mike Masnick "Key Techdirt SOPA/PIPA Post Censored By Bogus DMCA Takedown Notice" (27 February 2012) Techdirt <www.techdirt.com>.

B The Law Commission takedown power

Takedown powers have been used in a pretext manner to remove legitimate content. The particular takedown power discussed above was that contained in the DMCA.[192] I argued earlier that the NZLC was drawing on the Pressure and Persuasion method in advocating for an Approved Agency to informally resolve complaints. The connection between Pressure and Persuasion and the Agency was clear. The Tribunal takedown power is discussed in the context of the Pretext method because takedown powers have been used in Pretext censorship. However, unlike the link between Pressure and Persuasion and the Agency, the Tribunal takedown power does not draw on the Pretext method. How the Tribunal takedown power is set up actually prevents its use in a pretext manner.[193]

The Communications Tribunal will have the power to order that "material specified in the order be taken down from any electronic media."[194] Such an order may be made against the defendant, an ISP, a website host, or any other person seen as encouraging offensive communication toward the complainant.[195] If the Tribunal orders a takedown, failure to comply will be an offence against cl 22 of the draft Bill and could result in a fine not exceeding $5000. The briefing paper is clear that an intermediary receiving a takedown order will only be accountable where they have received clear notice of the harmful material and do not do all that is reasonable to remove the material. The draft Bill does not specify these considerations.

The Tribunal set up is what prevents an individual using the takedown power in a pretext manner. This is because the power to issue a takedown notice is not vested in the individual complainant. The DMCA takedown involves an aggrieved individual sending a notice in the prescribed form that some specified content is infringing their copyright. The

---

[192] Directive 2000/31 on Electronic Commerce [2000] OJ L178/1 art 13(1)(e) also contains a notice and takedown regime.

[193] See Appendix 3 for relevant provisions of the Communications (New Media) Bill relating to the Communications Tribunal.

[194] Communications (New Media) Bill 2012 cl 16(1)(a).

[195] Communications (New Media) Bill 2012 cl 16(2).

notice sender is therefore the judge and jury over whether that content is infringing. As already discussed, an intermediary receiving a properly filled out notice must acquiesce to the takedown notice in order to be protected from liability. The DMCA allows a takedown notice to be sent before any judicial process occurs and before notification to the content creator. The Tribunal takedown power is only exercisable through a judicial process, and will only be granted ex parte in exceptional circumstances. A DMCA takedown is always ex parte.

The NZLC explicitly states that the takedown power will have to be exercised in accordance with NZBORA. This means that as the takedown power will limit freedom of expression, it must be demonstrably justified in a free and democratic society.[196] In order to receive an order, the complainant must show significant harm. The Tribunal will take into account the public interest in the subject matter under consideration before issuing a takedown order. The controls on the effective use of the power (effective in the sense that sending a takedown notice results in a takedown) are far more stringent than under the DMCA. Under the DMCA, the only control is that the notice must be in the prescribed form. As long as it is in the prescribed form, a takedown will inevitably result (unless the intermediary is so sure that the notice is being used in a pretext manner that they risk their immunity from liability by not taking the content down). A takedown under the Tribunal will be far harder to attain than a takedown under the DMCA.

Additionally, only a natural person can complain to the Tribunal, and in ordinary circumstances this person must be the victim of the communication. A notice sender under the DMCA does not have to be a natural person. Companies can therefore use the DMCA advantageously.

C Assessing the pretext method using Bambauer's legitimacy framework

Using already existing censorship mechanisms in a pretext manner allows for censorship of legitimate content. Censorship mechanisms are able to be advantageously/maliciously misused in a pretext manner because the laws containing them

[196] Bill of Rights Act 1990, s 5.

contain insufficient controls. I will now assess the legitimacy of Pretext censorship using the legitimacy framework and find that it is illegitimate. It should be clear from the above that I consider the Tribunal takedown power is adequately controlled so that under the framework it will be legitimate.

1 Openness

The test for openness was whether a State admits it censors and why. Pretext based censorship satisfies this requirement. A government employing pretext-based censorship will admit that it is censoring content and give a justification for doing so. For example, Kentucky's domain name seizure was open as the State admitted that it was blocking gambling websites, and gave the justification for doing so as protecting Kentucky citizens. If using the Pretext method, the justification given will not be the true justification for censorship, or may just be a side effect of the true justification, but nonetheless a reason for censorship is given. The fact that a justification is given, such as enforcing property rights, is the hallmark of Pretext censorship.

Tribunal takedown

The Tribunal takedown will be open. The fact that the power will be used is admitted in the publicly available draft Bill. When the power is used, the Tribunal will publish its reasons for ordering takedown, therefore the reasons for the censorship will be easily accessible.

2 Transparency

The test for transparency was whether the State is transparent about what is being blocked, and specific about criteria used to determine blockage. Use of Pretext cannot be transparent because what is stated as being blocked in legislation that is used in a pretext manner is not what is actually blocked. If, for example, a State drafts legislation to block "obscenity", you could say that this is being transparent, because the fact that obscenity is targeted is publicly available. However, "obscenity" is broad and could be used to block a vast number of things. Broadness could be either intentional or unintentional. Regardless, transparency is not satisfied.

Tribunal takedown

The Tribunal takedown power will be transparent because the Tribunal must publish reasons for its decisions which one would imagine would specify what is being blocked in a takedown order. The draft Bill sets out specific criteria for the Tribunal to follow in making an order.

3 Narrowness

The test for narrowness was how closely does what the State says it blocks match what is blocked in reality. Pretext works as a method of censorship because the laws containing the censorship mechanisms that are used in a pretext manner are not sufficiently narrow, or contain exploitable loopholes. Another consideration for testing legitimacy under narrowness was whether the State over or underblocks. Pretext censorship clearly results in overblocking.

Tribunal takedown

The Tribunal's takedown power is narrowly circumscribed by the draft Bill. It will satisfy the narrowness requirement as long as the commitment that intermediaries will only be required to take "reasonable steps" in taking content down is kept to.

4 Accountability

The test for accountability was whether there is any way for public involvement in decisions of what to censor, and any aspect of appeal/recourse after content is censored. Pretext based censorship is not accountable because it involves using laws advantageously for purposes that are not purposes of those laws. In a democracy the public may be involved in the drafting of legislation through public submissions. However, when that legislation is used for purposes beyond what was originally intended, any possible public involvement is negated. Bambauer explains the problems with pretext censorship related to accountability:[197]

[197] Bambauer "Orwell's Armchair", above n 7, at 21.

> Laws regulating speech necessarily include safeguards to prevent flaws such as vagueness, overbreadth, or content discrimination. Regulations unrelated to speech usually lack these protections, and concomitantly confer greater power upon government censors and impose greater costs on society. They present a heightened risk of arbitrary enforcement, since they are employed not to address the societal interest that is the laws' initial purpose, but for an orthogonal one.

Tribunal takedown

The Tribunal's use of its takedown power satisfies accountability. The public is involved in the Tribunal's exercise of its takedown power, as it will only do so based on a complaint from a member of the public. The Tribunal has no power to order a takedown without a complaint activating its jurisdiction. The draft Bill includes a right of appeal, but only for the complainant.

D Conclusion

The Pretext method of censorship makes use of already existing censorship mechanisms to censor legitimate content. Censorship is justified using those laws when the true reason for censorship is separate. This is able to occur because the censorship mechanisms that are maliciously/advantageously used are not open, transparent or sufficiently narrow. The single most important aspect allowing censorship of legitimate content is insufficient narrowness of censorship mechanisms. This is clearly demonstrated by the broad requirements of the DMCA. Not only are the censorship mechanisms which are misused not open, transparent or sufficiently narrow, but Pretext as a censorship method is not open, transparent, narrow or publicly accountable. It is therefore an illegitimate censorship method. In comparison, the Tribunal's takedown power is open, transparent and narrowly circumscribed.

VIII Conclusion

Indirect censorship is obscured by concerns such as intellectual property enforcement and national security. It is a threat to freedom of expression because by its nature people do not know it is occurring. I have argued that the insidious nature of indirect censorship has meant that it has often been overlooked in favour of condemning overt censorship. However, a failure to address indirect censorship makes criticism of any type of internet censorship incomplete, and often hypocritical. In a perfect world, countries exercising overt internet censorship would listen to the vocal criticism of their practices and change their ways. This is of course unrealistic, and would regardless be treating the surface problem without addressing the hidden root of indirect censorship. Without addressing the aspects of censorship mechanisms that allow indirect censorship to occur, such as broad opaque laws, "treating" overt censorship will be hollow, as governments can achieve their censorship objectives through indirect mechanisms while disclaiming overt censorship.

The evolutionary course of the internet was discussed as a background factor in establishing indirect censorship as a problem. This was because the growth of the internet from the "Open Commons" phase through to the current "Access Contested" phase has shown a pattern of increasingly sophisticated regulation, and conversely, increasingly sophisticated censorship. The current contested nature of the internet has "opened the lid" on the internet to allow consideration of all aspects of the space, including that which has been obscured by other concerns, namely the use of indirect censorship mechanisms to attain removal/blockage of legitimate online content.

The New Zealand Law Commission's proposals to deal with "harmful digital communications" were discussed in the context of Lessig's argument that the "architecture" of the internet embeds values which make up an "unwritten constitution of cyberspace". This argument was that as the nature of the internet is entirely built, we cannot allow it to be built in one particular way and then complain that we are constrained by how it has been built. By discussing the incentives of intermediaries relating to content removal, I argued that the

"constitution" of the internet has become one where content removal and overbroad filtering are the norm, through conscious legislative choices which have embedded these values and conscious manipulating of the incentives of intermediaries. The NZLC's proposals are conscious choices which, when implemented, will embed values in the internet.

An aim of this dissertation was to explain how methods of indirect censorship have been used to censor legitimate content. The methods of censorship focused on were Pressure and Persuasion and the Pretext method. Applying Bambauer's process oriented legitimacy framework, I argued that neither of these methods is legitimate as they are not open, transparent, narrow, or publicly accountable. As they lack characteristics of legitimacy, they have been able to be misused to block legitimate content. The NZLC's proposal for an Agency with persuasive power draws directly on Pressure and Persuasion. The controls on the Agency are too few, and the Agency's proposed use of Pressure and Persuasion would not be legitimate applying the framework. In this respect, the NZLC's proposals further embed the norm of content removal rather than content protection and further incentivise intermediaries to quickly remove content when asked. In comparison, the NZLC's proposal for a Tribunal takedown power is legitimate as it is far more tightly controlled and prevents Pretext method misuse. In the interests of preventing censorship of legitimate content through misuse of censorship mechanisms, the Agency proposal requires more careful consideration than it has received.

Bibliography

A Cases

1 New Zealand

Nationwide News Pty Ltd v University of Newlands [2005] NZCA 317.

O'Brien v Brown [2001] DCR 1065.

Sadiq v Baycorp (NZ) Ltd HC Auckland CIV 2007-404-6421, 31 March 2008.

Solicitor-General for New Zealand v Siemer HC Auckland CIV-2008-404-472, 8 July 2008.

2 United Kingdom

Al Amoudi v Brisard [2006] 3 All ER 294 (QB)

Byrne v Deane [1937] 1 KB 818.

Godfrey v Demon Internet [1999] EMLR 542.

B Legislation

1 New Zealand

Bill of Rights Act 1990.

Communications (New Media) Bill 2012.

2 European Union

Directive 2000/31 on Electronic Commerce [2000] OJ L178/1.

3 United Kingdom

Defamation Act 1996.

4 United States of America

Digital Millennium Copyright Act 17 USC.

C Treaties

International Covenant on Civil and Political Rights 999 UNTS 407 (opened for signature 19 December 1966, entered into force 23 March 1976).

UN General Assembly *Universal Declaration of Human Rights* GA Res 217 A, III (1948).

D Books and Chapters in Books

Danielle Keats Citron "Civil Rights in Our Information Age" in Saul Levmore and Martha C Nussbaum (ed) *The Offensive Internet: Speech, Privacy, and Reputation*, (Harvard University Press, Cambridge, Massachusetts, 2010) 31.

Ronald Deibert , John Palfrey, Rafal Rohozinski, and Jonathan Zittrain (ed) *Access Denied: The Practice and Policy of Global Internet Filtering* (MIT Press, United States of America, 2008).

Ronald Deibert , John Palfrey, Rafal Rohozinski, and Jonathan Zittrain (ed) *Access Controlled: The Shaping of Power, Rights and Rule in Cyberspace* (MIT Press, United States of America, 2010).

Ronald Deibert , John Palfrey, Rafal Rohozinski, and Jonathan Zittrain "Toward the Fourth Phase of Cyberspace Controls" in John Palfrey Ronald Deibert, Rafal Rohozinski, and Jonathan Zittrain (ed) *Access Contested: Security, Identity and Resistance in Asian Cyberspace*, (MIT Press, Cambridge, 2011) 3.

Lilian Edwards and Charlotte Waelde (ed) *Law and the Internet* (3rd ed, Hart Publishing Ltd, Oxford, 2009).

Robert Faris and Nart Villeneuve "Measuring Global Internet Filtering" in Ronald Deibert, John Palfrey, Rafal Rohozinski, and Jonathan Zittrain (ed) *Access Denied: The Practice and Policy of Global Internet Filtering* (MIT Press, United States of America, 2008) 1.

Judge David Harvey *internet.law.nz* (3rd ed, Lexis Nexis, Wellington, 2011).

Lawrence Lessig *Code and Other Laws of Cyberspace* (Basic Books, United States of America, 1999).

Lawrence Lessig *Code: And Other Laws of Cyberspace, Version 2.0* (Basic Books, United States of America, 2006).

David Mamet *Writing In Restaurants* (Penguin Books, New York, 1987).

Jason Mazzone *Copyfraud and Other Abuses of Intellectual Property Law* (Stanford Law Books, Palo Alto, 2011).

E Journal Articles

Julie Adler "The Public's Burden in a Digital Age: Pressues on Intermediaries and the Privatization of Internet Censorship" (2011) 20 JL & Pol'y 231.

Christian Ahlert, Chris Marsden and Chester Yung "How 'Liberty' Disappeared from Cyberspace: The Mystery Shopper Tests Internet Content Self-Regulation" (2004).

Derek Bambauer "Guiding the Censor's Scissors: a Framework to Assess Internet Filtering" (August 2008) Selected Works of Derek Bambauer <http://works.bepress.com/derek_bambauer>.

Derek E Bambauer "Cybersieves" (2009) 59(3) Duke LJ 377.

Derek E Bambauer "Orwell's Armchair" (September 2011) Social Science Research Network <http.ssrn.com>.

Jessica Bauml "It's a mad, mad Internet: globalization and the challenges presented by Internet censorship" (2011) 63(3) Fed Comm LJ 697.

Kathy Bowrey "Can a public-minded copyright deliver a more democratic Internet?" (2007) 56 University of New Brunswick Law Journal 26.

Annemarie Bridy "Graduated Response and the Turn to Private Ordering in Online Copyright Enforcement" (2010) 89 Or LR 81.

Lilian Edwards "From child porn to China, in one Cleanfeed" (2006) 3(3) SCRIPT-ed.

Seth F Kreimer "Censorship by Proxy: the First Amendment, Internet Intermediaries, and the Problem of the Weakest Link" (2006) 155(11) U Pa L Rev.

Tyler Moore and Richard Clayton "The Impact of Incentives on Notice and Take-down" (13 June 2008) Computer Laboratory, University of Cambridge <www.cl.cam.ac.uk>.

Dawn C Nunziato "How (Not) to Censor: Procedural First Amendment Values and Internet Censorship Worldwide" (2011) 42 Georgetown Journal of International Law 1123.

Jacob Rowbottom "To Rant, Vent and Converse: Protecting Low Level Digital Speech" (2012) 71(2) CLJ.

Wendy Seltzer "Free speech unmoored in copyright's safe harbor: chilling effects of the DMCA on the First Amendment" (2010) 24(1) Harv J L & Tech 171.

Frederick Schauer "The Exceptional First Amendment" (February 2005) Social Science Research Network <www.ssrn.com>.

Jennifer Urban and Laura Quilter "*Efficient Process or "Chilling Effects"? Takedown Notices Under Section 512 of the Digital Millenium Copyright Act*" (2006) 22 Santa Clara Computer & High Tech LJ 621.

Felix T Wu "Collateral Censorship and the Limits of Intermediary Immunity" (2011) (1) Notre Dame L Rev 293.

F Law Commission Materials

Law Commission *The News Media Meets 'New Media': Rights, Responsibilities and Regulation in the Digital Age* (NZLC IP27, 2011).

Law Commission *Harmful Digital Communications: The adequacy of the current sanctions and remedies* (August 2012).

G Reports

Dr Yaman Akdeniz "Report of the OSCE Representative on Freedom of the Media on Turkey and Internet Censorship" (2009) Organisation for Security and Co-operation in Europe <www.osce.org>.

Frank La Rue *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression* A/HRC/17/27 (2011).

H Internet Resources

"BPI DMCA (Copyright) Complaint to Google" (21 September 2012) Chilling Effects Clearinghouse <http://chillingeffects.org/N/593158>.

"Declaration of Internet Freedom" (2 July 2012) Internet Declaration <www.internetdeclaration.org>.

"Google agrees $1.65 billion deal for YouTube" *Tech*, *New Scientist* (onlined ed, London, 10 October 2006).

"letter re: defamatory comments on car enthusiast website" (27 August 2012) Chilling Effects Clearinghouse <http://chillingeffects.org/N/574178>.

"LiveJournal.com Statistics" (14 July 2012) LiveJournal Press Area: Statistics <www.livejournal.com>.

"Media DMCA (Copyright) Complaint to Google" (20 January 2012) Chilling Effects Clearinghouse <http://chillingeffects.org/notice.cgi?sID=189468>.

"Megaupload to Universal: You've Got Some Explaining To Do" (28 December 2011) Torrent Freak <www.torrentfreak.com>.

"Sleeping uneasy: Call-A-Mattress seeks removal of unfavorable review" (1 October 2012) Chilling Effects Clearing House <http://chillingeffects.org/N/618966>.

John Perry Barlow "Declaration of the Independence of Cyberspace" (8 February 1996) Electronic Frontier Foundation Projects <http://projects.eff.org>.

Samuel Blackstone "Iran Plans To Stop Using The Internet By 2013" (9 August 2012) Business Insider <www.businessinsider.com>.

Anne Broache "RIAA: No need to force ISPs by law to monitor piracy" (30 January 2008) CNET News <http://news.cnet.com>.

Nicholas Carlson "BREAKING: Google Pulls Search Engine Out Of China" (22 March 2010) Business Insider <http://articles.businessinsider.com>.

Dorothy Chou "More transparency into government requests" (18 June 2012) Google Official Blog <http://googleblog.blogspot.com>.

Piotr Czerski "We, the Web Kids." (15 Feburary 2012) Pastebin <http://pastebin.com/0xXV8k7k>.

Cory Doctorow "Scientology memo asks members to censor critical web comments with trumped up 'code of conduct' complaints" (7 July 2012) Boing Boing <http://boingboing.net>.

Eva Galperin "YouTube Anti-Scientology Takedowns: Good News, Bad News" (25 September 2008) Electronic Frontier Foundation <www.eff.org>.

Google Inc "Google Transparency Report" <www.google.com/transparencyreport>.

David Kravets "U.S. Copyright Czar Cozied Up to Content Industry, E-Mails Show" (14 October 2011) Wired Threat Level <www.wired.com/threatlevel>.

Fred von Lohmann "Transparency for copyright removals in search" (25 May 2012) Google Official Blog <http://googleblog.blogspot.co.nz>.

Matthew Lasar "Big Content, ISPs nearing agreement on piracy crackdown system" (24 June 2011) Ars Technica Law & Disorder <http://arstechnica.com/tech-policy>.

Mike Masnick "Can't We All Get Along: Principles Over Policy; Ideas Over Ideology" (6 July 2012) Techdirt <www.techdirt.com>.

Mike Masnick "Kentucky's Gambling Domain Name Grab Sets A Terrible Precedent" (10 October 2008) Techdirt <www.techdirt.com>.

Mike Masnick "Key Techdirt SOPA/PIPA Post Censored By Bogus DMCA Takedown Notice" (27 February 2012) Techdirt <www.techdirt.com>.

Mike Masnick "White House's New Report on Intellectual Property Enforcement Should Get A Copyright As A Creative Work Of Fiction" (30 March 2012) Techdirt <www.techdirt.com>.

Mike Masnick "Worst Kept Secret Now Confirmed: Government Was Very Involved Helping RIAA/MPAA Negotiate Six Strikes" (14 October 2011) Techdirt <www.techdirt.com>.

Jason Mick "Obama Conscripts ISPs as 'Copyright Cops', Unveils 'Six Strikes' Plan" (8 July 2011) Daily Tech <www.dailytech.com>.

Jack Mirkinson "Rush Limbaugh: Sandra Fluke, Woman Denied Right To Speak At Contraception Hearing, A 'Slut'" (29 Feburary 2012) The Huffington Post <www.huffingtonpost.com>.

Glyn Moody "UK Government Pressuring Search Engines To Censor Results In Favor Of Copyright Industries" (2 March 2012) Techdirt <www.techdirt.com>.

Steven Musil "Iran expected to permanently cut off Internet by August" (9 April 2012) CNET News <http://news.cnet.com>.

OpenNet Initiative "Burma (Myanmar)" (6 August 2012) OpenNet Initiative Research Profiles <http://opennet.net>.

OpenNet Initiative "China" (9 August 2012) OpenNet Initiative Research Profiles <http://opennet.net>.

OpenNet Initiative "North Korea" (10 May 2007) OpenNet Initiative Research Profiles <http://opennet.net>.

OpenNet Initiative "ONI Methodology, Tools, and Data FAQ" OpenNet Initiative <http://opennet.net/oni-faq>.

OpenNet Initiative "Tunisia" (7 August 2009) OpenNet Initiative Research Profiles <http://opennet.net>.

OpenNet Initiative "Filtering Data" (8 November 2011) OpenNet Initiative <http://opennet.net/research/data>.

Sergey Park "Казахстан: Блогеры обсуждают блокировку Живого Журнала" (2 September 2011) (translated ed: Adil Nurmakov (translator) Sergey Park "Kazakhstan: Bloggers Denounce Repeated Block of LiveJournal") Global Voices <http://globalvoicesonline.org>.

Steven Price "Critiquing the Law Commission" (2 September 2012) Media Law Journal <www.medialawjournal.co.nz>.

Reporters Without Borders "Internet Enemies Report 2012" (12 March 2012) Reporters Sans Frontieres <http://march12.rsf.org./en/>.

TC Sottek "The Declaration of Internet Freedom: how the net's minutemen plan to protect the future" (2 July 2012) The Verge <www.theverge.com>.

United Nations "Status International Covenant on Civil and Political Rights" (8 July 2012) United Nations Treaty Collection <http://treaties.un.org>.

Mu Xuequan "China issues report on human rights in the US" (25 May 2012) Xinhuanet English News <http://news.xinhuanet.com>.

Matt Zimmerman "[Updated] Limbaugh Copies Michael Savage's Bogus Copyright Theory, Sends DMCA Takedown to Silence Critics" (24 April 2012) Electronic Frontier Foundation <www.eff.org>.

I Other Resources

Genevieve Carbery "Governments 'filtering, censoring' content" *Irish Times* (online ed, Ireland, June 18 2012).

Richard Clayton "Judge and Jury? how "Notice & Take Down" gives ISPs an unwanted role in applying the Law to the Internet" (26 July 2000).

Hilary Clinton, Secretary of State of the United States of America "Remarks on Internet Freedom" (speech to The Newseum, Washington DC, 21 January 2010).

Laws of New Zealand Defamation (online ed).

Daniel Lyons "Attack of the Blogs" *Forbes Magazine* (New York City, online ed, 14 November 2005).

Rebecca MacKinnon "Let's take back the Internet!" (speech to TedGlobal, Edinburgh (13 July 2011).

New Zealand Law Foundation Centre for Law and Policy in Emerging Technologies “Cyberbullying: Do we need new law for cyber-bullies?” Public Discussion (University of Otago, 6 September 2012).

Appendix 1: Relevant Provisions of the Digital Millennium Copyright Act

S 512(c) Information Residing on Systems or Networks At Direction of Users.—

(1) In general.— A service provider shall not be liable for monetary relief, or, […] for injunctive or other equitable relief, for infringement of copyright by reason of the storage at the direction of a user of material that resides on a system or network controlled or operated by or for the service provider, if the service provider—

(A)

(i) does not have actual knowledge that the material or an activity using the material on the system or network is infringing;

(ii) in the absence of such actual knowledge, is not aware of facts or circumstances from which infringing activity is apparent; or

(iii) upon obtaining such knowledge or awareness, acts expeditiously to remove, or disable access to, the material;

(B) does not receive a financial benefit directly attributable to the infringing activity, in a case in which the service provider has the right and ability to control such activity; and

(C) upon notification of claimed infringement as described in paragraph (3), responds expeditiously to remove, or disable access to, the material that is claimed to be infringing or to be the subject of infringing activity.

(2) Designated agent.— The limitations on liability established in this subsection apply to a service provider only if the service provider has designated an agent to receive notifications of claimed infringement described in paragraph (3), by making available through its service, including on its website in a location accessible to the public, and by providing to the Copyright Office, substantially the following information:

(A) the name, address, phone number, and electronic mail address of the agent.

(B) other contact information which the Register of Copyrights may deem appropriate.

(3) Elements of notification.—

(A) To be effective under this subsection, a notification of claimed infringement must be a written communication provided to the designated agent of a service provider that includes substantially the following:

(i) A physical or electronic signature of a person authorized to act on behalf of the owner of an exclusive right that is allegedly infringed.

(ii) Identification of the copyrighted work claimed to have been infringed, or, if multiple copyrighted works at a single online site are covered by a single notification, a representative list of such works at that site.

(iii) Identification of the material that is claimed to be infringing or to be the subject of infringing activity and that is to be removed or access to which is to be disabled, and information reasonably sufficient to permit the service provider to locate the material.

(iv) Information reasonably sufficient to permit the service provider to contact the complaining party, such as an address, telephone number, and, if available, an electronic mail address at which the complaining party may be contacted.

(v) A statement that the complaining party has a good faith belief that use of the material in the manner complained of is not authorized by the copyright owner, its agent, or the law.

(vi) A statement that the information in the notification is accurate, and under penalty of perjury, that the complaining party is authorized to act on behalf of the owner of an exclusive right that is allegedly infringed.

(B)

(i) Subject to clause (ii), a notification from a copyright owner or from a person authorized to act on behalf of the copyright owner that fails to comply substantially with the provisions of subparagraph (A) shall not be considered under paragraph (1)(A) in determining whether a service provider has actual knowledge or is aware of facts or circumstances from which infringing activity is apparent.

(ii) In a case in which the notification that is provided to the service provider's designated agent fails to comply substantially with all the provisions of subparagraph (A) but substantially complies with clauses

(ii), (iii), and (iv) of subparagraph (A), clause (i) of this subparagraph applies only if the service provider promptly attempts to contact the person making the notification or takes other reasonable steps to assist in the receipt of notification that substantially complies with all the provisions of subparagraph (A).

s 512(f) Misrepresentations.— Any person who knowingly materially misrepresents under this section—

(1) that material or activity is infringing, or

(2) that material or activity was removed or disabled by mistake or misidentification,

shall be liable for any damages, including costs and attorneys' fees, incurred by the alleged infringer, by any copyright owner or copyright owner's authorized licensee, or by a service provider, who is injured by such misrepresentation, as the result of the service provider relying upon such misrepresentation in removing or disabling access to the material or activity claimed to be infringing, or in replacing the removed material or ceasing to disable access to it.

S 512(g) Replacement of Removed or Disabled Material and Limitation on Other Liability.—

(1) No liability for taking down generally.— Subject to paragraph (2), a service provider shall not be liable to any person for any claim based on the service provider's good faith disabling of access to, or removal of, material or activity claimed to be infringing or based on facts or circumstances from which infringing activity is apparent, regardless of whether the material or activity is ultimately determined to be infringing.

(2) Exception.— Paragraph (1) shall not apply with respect to material residing at the direction of a subscriber of the service provider on a system or network controlled or operated by or for the service provider that is removed, or to which access is disabled by the service provider, pursuant to a notice provided under subsection (c)(1)(C), unless the service provider—

(A) takes reasonable steps promptly to notify the subscriber that it has removed or disabled access to the material;

(B) upon receipt of a counter notification described in paragraph (3), promptly provides the person who provided the notification under subsection (c)(1)(C) with a copy of the counter notification, and informs that person that it will

replace the removed material or cease disabling access to it in 10 business days; and

(C) replaces the removed material and ceases disabling access to it not less than 10, nor more than 14, business days following receipt of the counter notice, unless its designated agent first receives notice from the person who submitted the notification under subsection (c)(1)(C) that such person has filed an action seeking a court order to restrain the subscriber from engaging in infringing activity relating to the material on the service provider's system or network.

(3) Contents of counter notification.— To be effective under this subsection, a counter notification must be a written communication provided to the service provider's designated agent that includes substantially the following:

(A) A physical or electronic signature of the subscriber.

(B) Identification of the material that has been removed or to which access has been disabled and the location at which the material appeared before it was removed or access to it was disabled.

(C) A statement under penalty of perjury that the subscriber has a good faith belief that the material was removed or disabled as a result of mistake or misidentification of the material to be removed or disabled.

(D) The subscriber's name, address, and telephone number, and a statement that the subscriber consents to the jurisdiction of Federal District Court for the judicial district in which the address is located, or if the subscriber's address is outside of the United States, for any judicial district in which the service provider may be found, and that the subscriber will accept service of process from the person who provided notification under subsection (c)(1)(C) or an agent of such person.

(4) Limitation on other liability.— A service provider's compliance with paragraph (2) shall not subject the service provider to liability for copyright infringement with respect to the material identified in the notice provided under subsection (c)(1)(C).

Appendix 2: DMCA Complaint Template

Adapted from "BPI DMCA (Copyright) Complaint to Google" (21 September 2012) Chilling Effects Clearinghouse <http://chillingeffects.org/N/593158>.

Sender Information:
Name:
Address:
Email:
Phone number:

Recipient Information:
Name:
Company Name:
Address:

Sent via: online form
Re: Websearch Infringement Notification via Online Form Complaint
Google DMCA Form: Infringement Notification for Web Search
Contact Information
Name:
Company Name:
Copyright holder:
Country/Region:
YOUR COPYRIGHTED WORK
Copyright claim #1:
Original work URL(s):
Allegedly infringing URLs:

Copyright claim #2:
Original work URL(s):
Allegedly infringing URLs:

SWORN STATEMENTS
I have a good faith belief that use of the copyrighted materials described in this notification as allegedly infringing is not authorized by the copyright owner, its agent, or the law.
The information I have submitted is accurate, and I swear, under penalty of perjury, that with respect to this notification, I am the copyright owner or am authorized to act on behalf of the owner of an exclusive right that is allegedly infringed.

[signed]

Appendix 3: Relevant Provisions of the Communications (New Media) Bill

Cl 3 Purpose

The purpose of this Act is to mitigate harm caused to individuals by electronic communications.

Cl 7 Communication principles

(1) Every Agency or Tribunal performing functions or exercising powers under this Act must take account of the following communication principles:

Principle 1

A communication should not disclose sensitive personal facts about an individual.

Principle 2

A communication should not be threatening, intimidating, or menacing.

Principle 3

A communication should not be grossly offensive to a reasonable person in the complainant's position.

Principle 4

A communication should not be indecent or obscene.

Principle 5

A communication should not be part of a pattern of conduct that constitutes harassment.

Principle 6

A communication should not make a false allegation.

Principle 7

A communication should not contain a matter that is published in breach of confidence.

Principle 8

A communication should not incite or encourage anyone to send a message to a person with the intention of causing that person harm.

Principle 9

A communication should not incite or encourage another person to commit suicide.

Principle 10

A communication should not denigrate a person by reason of his or her colour, race, ethnic or national origins, religion, ethical belief, gender, sexual orientation, or disability.

(2) Section 13(2) states how the communication principles affect the exercise of functions or powers of the Tribunal.

Cl 9 Functions of Agency

The functions that may be conferred on an Agency by notice under section 8 are—

(a) to help people to resolve problems caused to them by electronic communications:

(b) to receive and assess complaints about electronic communications:

(c) to investigate complaints, unless the Agency considers the subject-matter of the complaint is unlikely to cause harm or the complaint is otherwise inappropriate for investigation:

(d) to liaise with website hosts and internet service providers and, if appropriate, to request them to take down or amend posts that are clearly offensive:

(e) […]

(f) to advise a complainant in an appropriate case to apply to a Tribunal for an order under section 16 requiring a website host, internet service provider, or telecommunications provider to identify the author of an offensive communication:

(g) to advise a complainant to refer a complaint to a Tribunal if the Agency is satisfied that—

(i) the complaint meets the appropriate level of seriousness and has proved incapable of resolution by other means; or

(ii) the complaint is so serious, and the resolution of it is so urgent, that it should be referred directly to the Tribunal without mediation:

(h) to certify that is has recommended the referral of a complaint to a Tribunal:

[…]

Cl 10 Powers of Agency

(1) An Agency has all the powers necessary for carrying out the Agency's functions.

Cl 13 Functions, duties, and powers of Tribunals

(1) The functions of a Tribunal are—

(a) to consider and determine applications for any order under section 16:

(b) to exercise and perform any other functions, powers, and duties that are conferred or imposed on it by or under this Act or any other enactment:

(c) to do any other thing necessary for performing, or reasonably incidental to, the Tribunal's functions.

(2) The Tribunal must not consider or determine any application for any order under section 16 unless it is satisfied that—

(a) a communication principle has been breached; and

(b) that breach has caused or is likely to cause significant harm to an individual.

(3) The Tribunal has all the powers that are reasonably necessary to enable it to perform its functions

Cl 15 Consideration and determination of complaints by Tribunal

(1) […]

(2) A Tribunal must consider and determine a complaint with as little formality and technicality, and as speedily, as is permitted by—

(a) the requirements of this Act; and

(b) a proper consideration of the complaint; and

(c) the principles of natural justice.

Cl 16 Orders that may be made by Tribunal

(1) A Tribunal may, on a complaint or an application, make 1 or more of the following orders:

(a) an order requiring that material specified in the order be taken down from any electronic media:

[…]

(2) A Tribunal may apply an order or part of an order under this section to all or any of the following:

(a) the defendant:

(b) an internet service provider:

(c) a website host:

(d) […]

(3) In deciding whether or not to make an order, and the form of an order, a Tribunal must take into account the following:

(a) the content of the communication, its offensive nature, and the level of harm caused by it:

(b) the purpose of the communicator in communicating it:

(c) the occasion, context, and subject-matter of the communication:

(d) the extent to which the communication has spread beyond the original communicator and recipient:

(e) the age and vulnerability of the complainant:
(f) the truth or falsity of the statement:
(g) the extent to which the communication is of public interest:
(h) the conduct of the defendant, including any attempt by the defendant to minimise the harm caused:
(i) the conduct of the complainant, including the extent to which that conduct has contributed to the harm suffered.

(4) In exercising its functions, the tribunal must have regard to the importance of freedom of expression.

(5) A Tribunal must give reasons for its decisions and those reasons must be published.