Internet Filtering: Rhetoric, Legitimacy, Accountability and Responsibility[*]

TJ McIntyre[1] and Colin Scott[2]

> "I do intend to carry out a clear exploring exercise with the private sector … on how it is possible to use technology to prevent people from using or searching dangerous words like bomb, kill, genocide or terrorism" – EU Justice and Security Commissioner Franco Frattini, 10 September 2007[3]

1. Introduction

In the Internet context, filtering and blocking refer to technologies which provide an automatic means of preventing access to or restricting distribution of particular information. There is, of course, nothing new about seeking to control access to media and other resources. Governments have long had lists of banned books, sought to control access to newspapers, or sought cuts to films prior to their general exhibition. But we argue that qualitative differences between contemporary internet content filtering practices and traditional censorship raise new problems of regulatory accountability and legitimacy.

Consider the following recent examples. Many states, such as Saudi Arabia and China, have deployed filtering at a national level to censor political or pornographic material, in effect creating "borders in cyberspace".[4] Google's Chinese language site, at the behest of the Chinese Government, has introduced censorship of searches such as "Tiananmen Square".[5] The UK's dominant incumbent telecommunications operator, British Telecom, (in consultation with the Home Office) has put in place a "Cleanfeed" system which automatically blocks customer requests for websites alleged to be hosting child pornography[6] and the Government has indicated its intention to ensure that all UK internet service providers (ISPs) should adopt a similar system, whether by voluntary cooperation or otherwise.[7] In Belgium, the courts have ordered an ISP to implement technical measures to prevent user access to filesharing websites and to stop users from distributing certain music files.[8] In Canada the ISP Telus blocked its subscribers from seeing a website supporting a strike by its

[*] This paper appeared in Brownsword, R. and Yeung, K., *Regulating Technologies* (Oxford: Hart Publishing, 2008) and is reproduced here with the kind permission of Hart Publishing.

[1] Lecturer in Law, UCD School of Law. Contact: tjmcintyre@ucd.ie

[2] Professor of EU Regulation and Governance, UCD School of Law.

[3] I. Melander, "Web search for bomb recipes should be blocked: EU" *Reuters* 10 September 2007. Available at http://www.reuters.com/article/internetNews/idUSL1055133420070910

[4] N. Villeneuve, "The Filtering Matrix: Integrated Mechanisms of Information Control and the Demarcation of Borders in Cyberspace" (2006) 11(1) *First Monday*. Available at http://firstmonday.org/issues/issue11_1/villeneuve/index.html

[5] H. Bray, "Google China censorship fuels calls for US boycott" *The Boston Globe* 28 January 2006. Available at http://www.boston.com/news/world/articles/2006/01/28/google_china_censorship_fuels_calls_for_us_boycott/

[6] M. Bright, "BT puts block on child porn sites" *The Observer* 6 June 2004. Available at http://www.guardian.co.uk/technology/2004/jun/06/childrensservices.childprotection. See also Hunter, "BT Siteblock" (2004) 9 *Computer Fraud and Security* 4.

[7] W. Grossman, "The Great Firewall of Britain" *net.wars* 24 November 2006, quoting Vernon Coaker, Parliamentary Under-Secretary for the Home Department to Parliament. Available at http://www.pelicancrossing.net/netwars/2006/11/the_great_firewall_of_britain.html

[8] *Sabam v. Scarlet,* Decision of the Court of First Instance in Brussels of 29 June 2007, discussed in *OUT-LAW News*, 6 July 2007. Available at http://www.out-law.com//default.aspx?page=8239

employees, inadvertently blocking many unrelated sites also.[9] Meanwhile throughout the world ISPs and users deploy spam filters and sender blacklists with varying degrees of success.[10]

These examples differ greatly from each other. But in each case the blocking shares some common features. First, it is *automatic and self-enforcing* in its nature. Once the technology is developed and deployed no further human intervention is required, unless and until users find ways to circumvent the intended controls. Secondly, it is often *opaque*. Some filtering mechanisms may be transparent to the affected user, as with some email filtering systems which send to users summaries of email which has been blocked as spam. But in many cases filtering is, of necessity, opaque, in at least some dimensions, as a condition of its effectiveness. Thirdly, filtering generally involves *intermediaries*. Again, this is not always the case. A user may run a spam filter locally on their own machine. But since much filtering involves denying the end user access to certain material it is more common for filtering to be directed to other internet points of control.[11]

These three features are not unique to filtering. Lessig has pointed out the automatic and often opaque nature of code as a modality of regulation[12] while theorists such as Boyle[13] and Swire[14] have noted that the decentralised and international nature of the internet will encourage regulators to focus on indirect enforcement, targeting intermediaries rather than end users, “elephants” rather than “mice”. But we will suggest that in the particular context of filtering they interact to raise some important issues.

By way of introduction we will examine the *rhetoric* underlying the use of the term “filtering”. We suggest that this term, convenient though it is as shorthand for this technology, is loaded and that it may be preferable to talk in more neutral terms of “blocking” or even of “censorware”.

We will then go on to explore where filtering fits into our modalities of governance and the resulting issues of *legitimacy* and *accountability.* As regards legitimacy we argue in particular that the use of technology to exert control over internet users frequently challenges tenets associated with the rule of law concerning both the process for and content of norms governing behaviour. These challenges emerge, in particular, where technology is linked to compliance with voluntary codes or soft law instruments by non-state actors. Whilst it may be suggested that the voluntary character of compliance with such instruments reduces or removes the requirements suggested by rule of law concerns, the consequences of compliance will often accrue to third parties who do not experience compliance as voluntary and in situations

[9] CBC News 24 July 2005. Available at http://www.cbc.ca/story/canada/national/2005/07/24/telus-sites050724.html

[10] See for example L. Lessig, “The Spam Wars” *The Industry Standard* 31 December 1998. Available at http://www.lessig.org/content/standard/0,1902,3006,00.html

[11] See for example J. Zittrain, “Internet Points of Control”. (2003) 43 *Boston College Law Review* 1 discussing how and why regulators target ISPs rather than users.

[12] L. Lessig, *Code and Other Laws of Cyberspace* (Cambridge: MA, 1999).

[13] J. Boyle, “Foucault in Cyberspace: Surveillance, Sovereignty, and Hardwired Censors” (1997) 66 *University of Cincinnati Law Review* 177.

[14] P. Swire, “Of Elephants, Mice, and Privacy: International Choice of Law and the Internet” (August 1998). Available at SSRN: http://ssrn.com/abstract=121277

where many of the elements of the regime of control are determined by non-state actors outside of the normal public policy process.

Following on from that, we will argue that the combination of automatic enforcement, opaque systems and rules directed at intermediaries may leave affected users unaware that their behaviour is being controlled, so that the opaque nature of filtering may result in a loss of *accountability*. Where, as is often the case, it is not clear what is being blocked, why, or by whom, the operation of mechanisms of accountability – whether by way of judicial review, media scrutiny, or otherwise – is greatly reduced.

Finally we will consider the argument that, as compared with control through legal instruments, filtering may rob users of moral agency or *responsibility* in their use of the internet, with the implication that they may freely do whatever it is technically possible to do, with no necessity of moral engagement in their activities. If such consequences were to follow through into wider patterns of social interaction, the consequences for responsibility, and for social ordering generally, of such low trust mechanisms of control might be troubling.

We do not reject the use of filtering in the Internet context. Without filtering our email inboxes would rapidly become unusable. It is through the technology of filtering rather than legal controls that spam has, to a greater or lesser extent, been effectively tackled. The development of commercial websites which accredit and testify to the relevance of material, or the trustworthiness of others, has given many firms great success and is clearly meeting a demand.[15] The efficiency which is promised is seductive. However, we do suggest that the legitimacy of filtering in any particular context requires close examination by reference to issues of transparency, responsibility and accountability in respect of the devising and administering of controls, the purposes for which such controls are deployed, and the consent (or absence of consent) of those whose behaviour is controlled as a result.

2. *Rhetoric*

The term "filtering" is widely used – even by critics – as shorthand for bundles of practices through which technology is used to exert control over users of the Internet.[16] Other terms for filtering – such as the British Telecom "Cleanfeed" project – have also sought to capture the rhetorical allure of cleanliness and purity.

Others, however, have challenged this terminology. The term "filtering", it has been argued, implies an element of choice on the part of the affected user, with "censorware" being a more appropriate term for blocking which is beyond user control. [17] The term carries an illusion of precision:

[15] Y. Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (New Haven, 2006), 12, 75. Blogs, search engines, online bookstores, journals, online encyclopaedias, and buying/selling intermediaries such as eBay each engage in different forms of filtering.

[16] For example, Y. Akdeniz, "Who Watches the Watchmen? The role of filtering software in Internet content regulation" in C.Möller and A. Amouroux (eds.), *The Media Freedom Internet Cookbook* (Vienna, 2004); B. Esler, "Filtering, Blocking and Rating: Chaperones or Censorship?" in M. Klang and A. Murray (eds.), *Human Rights in the Digital Age* (Glasshouse Books, London, 2005); R.P. Wagner, "Filters and the First Amendment" (1999) 83 *Minnesota Law Review* 755.

[17] Oral testimony before the Library of Congress Copyright Office Hearing on anti-circumvention mechanisms under the Digital Millennium Copyright Act, 11 April 2003. Transcript available at http://

> "The word 'filter' is much too kind to these programs. It conjures up inaccurate, gee-whiz images of sophisticated, discerning choice … When these products are examined in detail, they usually turn out to be the crudest of blacklists, long tables of hapless material which has run afoul of a stupid computer program or person, perhaps offended by the word 'breast' (as in possibly 'breast cancer')." [18]

We would agree that the metaphorical deployment of the term filtering is loaded with meanings which imply virtue and thereby resists challenge through rhetoric. In particular (by analogy with the filtering of drinking water) the term may reinforce a view of the Internet as something that is piped into one's home where it is passively consumed. This view – building on the pervasiveness doctrine in broadcasting law – has already been deployed[19] to justify greater regulation of the internet, often coupled with an explicit comparison of objectionable material on the Internet to sewage in the domestic water supply.[20] The more interactive character of Web 2.0 technologies, such as social networking sites, removes them further from a parallel with broadcasting.

In addition, to say that we are filtering something implies that we are treating that something as an undifferentiated mass, and as noted by Finkelstein, that we are doing so in a relatively straightforward and scientific way. This may reflect a popular conception of the Internet as a single entity, but it is at odds with the reality of the Internet as being a network of networks – an architecture which links together a disparate collection of protocols, applications, sites, and users. If we wish to block certain content online then we may do so in a variety of different ways, in a number of different locations, and to a number of different users – for example at national boundaries or at the organisational level, on the server side or the user side, over all protocols or merely HTTP.[21] The loose use of the term internet filtering tends to undermine this diversity and may suggest that a one size fits all solution is appropriate.

Of course, alternative terms could equally be objected to. For example, to frame the discussion as one about "automated censorship" or "censorware" might draw the riposte that many aspects of the practice are distinct from censorship as it is

www.copyright.gov/1201/2003/hearings/transcript-apr11.pdf

[18] Congressional evidence of S. Finkelstein, quoted in B. Miner, "Internet Filtering: Beware the Cyber Censors" 12(4) *Rethinking Schools Online* (Summer 1998). Available at http://www.rethinkingschools.org/archive/12_04/net.shtml

[19] J.D. Wallace, "The Specter of Pervasiveness: Pacifica, New Media, and Freedom of Speech" *CATO Briefing Paper 35* (12 February 1998). Available at http://www.cato.org/pubs/briefs/bp-035.html

[20] A comparison notably made by the United States Department of Justice in its opening statement in *ACLU v. Reno*, 23 October 1996, transcript available at http://www.aclu.org/pdfs/freespeech/copatranscript_20061023.pdf. "If a water source was mixed with a sewer system, and you had a filter that screened out but 6.6 percent of it, would that be a solution to the problem? Would that cure the problem of the drinking water." Similarly (though speaking of viruses and other malware rather than pornography) technology site ZDNet recently editorialised that "But when we attach a PC to the Internet, we might as well be wading through open sewers. Currently, many ISPs are allowing Internet traffic to flow through their systems completely unfiltered, which is akin to a water authority pumping out raw sewage to its customers to clean for themselves." "Time to filter out the Internet effluent", *ZDNet* 18 August 2004. Available at http://news.zdnet.co.uk/leader/0,1000002982,39163885,00.htm

[21] R. Deibert and N. Villeneuve, "Firewalls and Power: An Overview of Global State Censorship of the Internet" in M. Klang and A. Murray (eds.), *Human Rights in the Digital Age* (Glasshouse Books, London, 2005) p. 114.

traditionally practiced, while it might also be said that the term "blocking" doesn't adequate convey the precision and selectivity which technology may make possible. Nonetheless, we would suggest that the term be used with caution.[22]

3. Implications of Filtering as a Method of Governance: Legitimacy and Accountability

Control of Internet use through filtering is part of a broader pattern of governance in which technology forms only a part. The technologies and practices associated with filtering, and the associated fragmentation in both the actors and modalities engaged in control of social and economic practices provide a key example of the difficulties of adapting traditional narratives of legitimacy and accountability to contemporary governance. In this section of the paper we first address the nature of governance practices associated with filtering and then address some of the normative implications.

Lawrence Lessig's celebrated claim "code is law"[23] dramatically highlighted the potential of software architecture to substitute for law in the control of behaviour. Elaborating on Lessig's four-way analysis, we recognise hierarchy (or law in Lessig's terms), competition (or markets), community (or norms) and design (or architecture) as four basic modalities of governance (or control).[24] Working with these four modalities of governance it appears mistaken to think of architecture as displacing other modalities. Design has long had a key role in controlling behaviour not separate from, but allied to, other modalities of governance, in particular the hierarchical exercise of legal power. Famously, Jeremy Bentham's Panopticon, a design for a prison in which a smaller number of guards are able to keep an eye on all the prison corridors from a central tower,[25] is dependent for success on the exercise of legal authority to detain prisoners and apply discipline to those who are observed breaching prison rules. Thus surveillance was used to support the exercise of legal power.[26] More recent work on crime control has emphasised the role of architecture and design in inhibiting criminal conduct, but again against a background of legal enforcement.[27]

There is potential also for linking control through design to the other governance modalities. Thus competition and design may operate together in the voluntary provision by car manufacturers of control mechanisms which enhance safety, such as inhibitors to prevent driving while under the influence of alcohol. Physical controls over the use of space in parks or pubs, which inhibit certain forms of behaviour, may be used to give expression to community norms rather than legal rules.

[22] We should also be conscious that the term "blocking" can be used in a more technical manner to refer to means of denying access to particular IP addresses or services on particular port numbers. See R. Deibert and N. Villeneuve, "Firewalls and Power: An Overview of Global State Censorship of the Internet" in M.Klang and A. Murray (eds.), *Human Rights in the Digital Age* (Glasshouse Books, London, 2005) at 112.

[23] L. Lessig, *Code and Other Laws of Cyberspace* (2nd ed Basic Books, Cambridge: MA, 2006), p6.

[24] A. Murray and C. Scott, "Controlling the New Media: Hybrid Responses to New Forms of Power" (2002) 65 *Modern Law Review* 491.

[25] J. Bentham, *Panopticon or the Inspection House* (Dublin, 1791).

[26] M. Foucault, *Discipline and Punish: The Birth of the Prison* (Harmondsworth, 1977); J. Scott, *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed* (New Haven, 1998).

[27] E.g. O. Newman, *Defensible Space: Crime Prevention Through Urban Design* (New York, 1972).

With filtering we can readily see that the technology may be linked to legal authority, as where ISPs are directed to block access to certain websites. Filtering may also be an aspect of market-based control, for example where businesses market filtering software for email and the test of the take-up and success of the product lies not with compliance with legal rules, but rather with the extent of sales in the market. A third possibility is that filtering is part of community-based system of control, for example where norms governing website access are reflected in shared software for inhibiting access to blacklisted sites. Frequently two or more modalities may be in play.[28]

Observing these governance modalities in play with filtering raises some important normative issues, particularly concerning the legitimacy of certain aspects of governance In its most general sense legitimacy refers to that general acceptance of governance arrangements which sustains the capacity for governance, even through times when the content of what is being done may be controversial. Internet governance raises problems because of the manner in which a traditional understanding of the separation between the role of governments, markets and communities is challenged by practices such as those deployed in respect of internet filtering. This challenge is reflected in anxieties that standard tenets of public accountability for governance decisions and compliance with values for the rule of law may be undermined.

Both accountability and rule of law aspects are reflected in the extent to which the implementation of filtering, though it may be mandated or permitted in public legislation, moves us away from public actors and legal rules for its implementation. Thus where a government legislates to prohibit access to certain books, there is a public legislative process involving elected representatives in the making of rules for enforcement by public officials. Much of the policy making and implementation in respect of internet filtering occurs through other mechanisms involving different actors with risks both to the transparency and accountability dimensions which, through conceptions of the rule of law, underpin legitimacy in governance.

Automatic Enforcement

A key feature (some would call it a virtue) of technological control is that it is applied automatically without human intervention. There is no scope for argument, no exercise of discretion and (depending on the code) all users are treated alike. At first glance this may seem to be virtuous from a rule of law perspective, since it reduces the scope for discretionary or biased enforcement, and thus for users to be treated differently without legitimate cause. But there are some troubling aspects of automatic enforcement.

First, by ruling out any element of discretion we may end up with an all or nothing approach to governance which may not comply with principles of proportionality –

[28] The analysis of multi-modal governance, in relation to the Internet and other social and economic activities, begs the question whether design is modality of control at all. See C. Scott, "Spontaneous Accountability" in M. Dowdle (ed.), *Public Accountability: Designs, Dilemmas and Experiences* (Cambridge University Press, Cambridge, 2006). The particular quarrel with design as a modality of control is that, in contrast with the other three modalities, it has no obvious 'accountability template' mirroring its control function.

we may, for example, see an entire website or domain blocked due to offending material on a single page. Villeneuve has pointed out that this form of over-blocking is common[29] – filtering systems tend not to be sufficiently granular to restrict themselves to the targeted material. Indeed, even sites which have no affiliation with the offending material may find themselves blocked if the common but crude approach of IP address filtering is used.

While in some cases overblocking may result from clumsy or lazy technical implementations, there is a deeper problem which may not lend itself to a technical solution. Software is a very efficient mechanism for implementing rules, but not so good when it comes to standards.[30] This presents a particular problem in relation to the filtering of material alleged to be distributed in breach of copyright.[31] Here, filtering software may be very efficient when identifying whether excerpts of copyright material are being used – but will fall down when faced with the standards-based assessment of whether that amounts to a "fair use" or "fair dealing" with the copyright work. The result may be to upset the balance struck by copyright law, resulting in hyper-efficient enforcement of copyright claims but systematic neglect of the situations where the law recognises that unauthorised use of copyright material is socially desirable. Whilst blocking may be automatic, "the process by which [users] protest their innocence and get the right to communicate back will be slow, bureaucratic, and manual."[32]

Consider for example the way in which students used the Internet to reveal serious security flaws in electronic voting machines produced by Diebold Electronics.[33] A particularly important aspect of that campaign was the use of internal emails of Diebold which had been leaked. Unsurprisingly, Diebold claimed copyright in the emails and threatened sites hosting them with legal action unless they were removed. The response on behalf of the students was what they described as "electronic civil disobedience" – disseminating the emails widely throughout the internet while simultaneously seeking a judicial declaration that this use was privileged. They were ultimately successful in the latter endeavour – the court accepted that portions of the email archives which were used to inform the public about concerns as to the legitimacy of elections were clearly subject to the fair use exception under US law.[34] However, had a filtering system been in place restricting the distribution of that material, discussion of an important matter of public concern would have been silenced in the meantime – notwithstanding the formal position which US law takes against prior restraints.[35]

[29] R. Villeneuve, "The Filtering Matrix: Integrated Mechanisms of Information Control and the Demarcation of Borders in Cyberspace" (2006) 11(1) *First Monday*. Available at http://firstmonday.org/issues/issue11_1/villeneuve/index.html

[30] J. Grimmelman, "Regulation by Software" (2005) 114 *Yale Law Journal* 1719.

[31] For example Google in October 2007 introduced a copyright filtering system for its video sharing site YouTube. See L. Rosencrance, "Google launches video blocking tool for YouTube" *Computerworld* 16 October 2007. Available at http://www.computerworld.com/action/article.do?command=viewArticleBasic&taxonomyId=15&articleId=9042839

[32] C. Doctorow, "French law proposal will force ISPs to spy on users and terminate downloaders without trial" *Boing Boing* 25 November 2007. Available at http://www.boingboing.net/2007/11/25/french-law-will-forc.html.

[33] See, e.g., Y. Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (New Haven, 2006) at 225 *et seq*.

[34] *Online Policy Group v. Diebold* 337 F. Supp. 2d 1195 (2004).

[35] See e.g. *New York Times Co. v. United States* 403 US 713 (1971).

Secondly, the automatic nature of software may eliminate the feedback mechanisms normally associated with good governance. Whereas the hierarchical mechanisms of legislative implementation contain mechanisms for registering concerns about what is or is not fair or effective, and similar feedback loops arise within both market and community governance processes, the automaticity of design-based controls associated with software negates the existence of a feedback loop.[36]

The feedback loop associated with rulemaking and enforcement is illustrated by the 1960 prosecution of Penguin Books for publishing *Lady Chatterley's Lover* – a trial which by drawing public attention to the book not only made it a best seller but also resulted in a substantial relaxation of the test of obscenity in English law.[37] Filtering systems, by doing away with a public enforcement process, may inhibit this evolution of norms.

We have noted the automatic nature of software-based implementation of filtering norms. A key contrast between such automatic enforcement, on the one hand, and bureaucratic enforcement of legal rules on the other, is that bureaucratic enforcers can cater for harsh or unintended effects through the exercise of discretion. Indeed much research on regulatory enforcement suggests that the discretion not to enforce is routinely applied by most enforcement agencies for a variety of reasons, and that formal enforcement occurs only in a minority of cases, frequently linked to perceptions of motivation or persistence in respect of breaches.[38] The observation of discretion in rule enforcement enables rule makers to set rules which might be considered harsh if systematically applied, but where the degree of stringency in the rule itself is considered necessary to address the objectives behind the rule. Automaticity within enforcement of filtering norms has no mechanism to deal with the problem of excess stringency in application and the related problem of over-inclusiveness in application of the norm. This is a problem not only for the achievement of objectives, but also for the legitimacy of norms which, in their application, inhibit conduct beyond what was intended.

Opaque Nature of Filtering

Traditional forms of censorship generally require that if items – such as particular books, videos or periodicals – are prohibited, then a list of those items must be made publicly available. After all, without such a list how is the citizen to know that they are breaking the law by importing or possessing such an item? In addition, traditional censorship mechanisms will generally give persons affected by the designation of items an opportunity to be heard prior to designation or to challenge a designation.[39] Also, in traditional censorship mechanisms we expect the publication of criteria which

[36] J. Grimmelman, "Regulation by Software" (2005) 114 *Yale Law Journal* 1719; L. Tien, "Architectural Regulation and the Evolution of Social Norms" (2003-2004) *Yale Journal of Law and Technology* 1

[37] See, for example, C.H. Rolph, *The Trial of Lady Chatterley: Regina v. Penguin Books* (London, 1961) for an (edited) transcript of the trial and explanation of the context in which it took place.

[38] P.Grabosky J. Braithwaite. *Of Manners Gentle: Enforcement Strategies of Australian Business Regulatory Agencies*. (Melbourne, Oxford University Press, 1986).

[39] The process in Irish law in respect of film and video is described in Rockett, *Irish Film Censorship: A Cultural Journey from Silent Cinema to Internet Pornography* (Dublin, 2004).

will be applied in determining whether particular material is objectionable. These factors can be lacking in the case of filtering.

At one level, the end-user may not be made aware that filtering is in operation[40], or that access to a particular site has been blocked by filtering. Nor will the site owner necessarily be aware unless they spot and can diagnose a fall-off in traffic. In some states websites deemed unacceptable by governments (for example those of opposition political groupings, media and human rights organisations) are routinely blocked, with feedback to the user suggesting that the website is not available ('file not found') or that access has been inhibited by some technical problem (eg 'connection timeout'). [41] The more transparent and accurate message, 'access blocked by government order' is less commonly given. The use of error pages has been described as "an attempt to deflect criticism, allowing the authorities to claim that they are not censoring Internet content".[42]

Alternatively, the end user may be actively misled – Uzbekistan, for example, informs users that sites banned for political reasons are blocked for supposed pornographic content.[43] This appears to neatly combines two layers of deception – simultaneously justifying the block and smearing political opposition. It has been observed that governments [u]nable to justify the reason for blocking political content…choose to obscure or deny the fact that such content is in fact targeted."[44] Even if a user is aware of the fact of filtering, they may not know who is responsible for it: it may be any entity upstream of the user.[45] We may not know, for example, whether it is the Chinese government blocking material, or some commercial entity which finds it expedient to cooperate.

There are also commercial imperatives at work. Manufacturers of filtering software guard their lists of blocked sites, seeing them as trade secrets. Those lists are generally encrypted, and the manufacturers have sued or threatened to sue those who would make them public.[46] Consequently the lists may not be subject to independent scrutiny or analysis. Villeneuve illustrates this with an interesting example:

> "Saudi Arabia was condemned by human rights organisations for blocking access to non-pornographic gay and lesbian sites. After learning about the blocked sites, the Saudi authorities promptly removed the blocking. Saudi Arabia never intended to block access to those sites. These sites were likely misclassified by the commercial filtering product, SmartFilter, that Saudi Arabia implemented at the national level. In effect, US corporations are in a position to determine what millions of citizens can and cannot view on the Internet.

[40] A point made by L. Lessig, *Code and Other Laws of Cyberspace* (Cambridge: MA, 1999) where he refers to "truth in blocking" as a desirable characteristic.

[41] R. Deibert and N. Villeneuve, "Firewalls and Power: An Overview of Global State Censorship of the Internet" in M. Klang and A. Murray (eds.), *Human Rights in the Digital Age* (Glasshouse Books, London, 2005) at 119.

[42] N. Villeneuve, "The Filtering Matrix: Integrated Mechanisms of Information Control and the Demarcation of Borders in Cyberspace" (2006) 11(1) *First Monday*. Available at http://firstmonday.org/issues/issue11_1/villeneuve/index.html

[43] *Ibid.*

[44] *Ibid.*

[45] L. Lessig, *Code and Other Laws of Cyberspace* (2nd Ed. Basic Books, Cambridge: MA, 2006), 257.

[46] For an example see B. Fitzgerald, "Note: Edelman v. N2H2 – At the Crossroads of Copyright and Filtering Technology" (2004) 69 *Brooklyn Law Review* 1471.

> Even the countries implementing filtering products do not know for certain what is in fact being blocked."[47]

Indeed, in numerous cases, manufacturers have taken advantage of this fact to blacklist and thereby silenced their critics.[48]

At least in some situations, it may be the case that transparency would destroy the effectiveness of filtering. For example, there is understandable concern that revealing the list of blocked child pornography sites censored by British Telecom's Cleanfeed system would simply advertise them further. The filtering of spam has also been marked by a battle of wits between spammers and filters – and some spam filters therefore keep their internal workings secret for fear that their effectiveness would be lost if spammers could tailor their offerings to circumvent the filters. This may be a general problem with any filters which engage in content analysis.

On the other hand, some jurisdictions have implemented elements of transparency. In Saudi Arabia, for example:

> "users are presented with a blockpage which states that the requested Web site has been blocked but it also contains a link to a Web form through which users can petition to have the site unblocked … The acknowledgement of blocked content allows users to petition to have sites unblocked if there has been a mis-classification. It also requires governments to justify why a specific site is blocked."[49]

However, such transparency might itself give rise to concern. This blunt statement – that the requested site has been blocked – will also serve to remind the user that their online activities are of some interest to the state, thus possibly having a chilling effect on further internet use.

Consequently, the opaque nature of many internet filtering processes serve to challenge key requirements in both public and market governance relating to feedback on the operations of the process. From public governance perspective the problem relates to the inability of those affected to know about and challenge decisions on filtering. From a market governance perspective such opacity removes the possibility of feedback processes through which errors can be detected and corrected.

Role of Intermediaries

Traditional forms of censorship and control of information have generally focused on either the person making available certain information (such as prohibiting the

[47] Villeneuve, "The Filtering Matrix: Integrated Mechanisms of Information Control and the Demarcation of Borders in Cyberspace" (2006) 11(1) *First Monday*. Available at http://firstmonday.org/issues/issue11_1/villeneuve/index.html

[48] For examples see The Free Expression Policy Project, *Internet Filters – a Public Policy Report* (New York, 2006), available at http://www.fepproject.org/policyreports/filters2.pdf ; a Electronic Frontiers Australia press release, "Government approved net filters attempt to silence critics" available at http://www.efa.org.au/Publish/PR000629.html; TIME Digital Magazine, "Cybersitter decides to take a time out" 8 August 1997 available at http://web.archive.org/web/20000830022313/http://www.time.com/time/digital/daily/0,2822,12392,00.html

[49] N. Villeneuve, "The Filtering Matrix: Integrated Mechanisms of Information Control and the Demarcation of Borders in Cyberspace" (2006) 11(1) *First Monday*. Available at http://firstmonday.org/issues/issue11_1/villeneuve/index.html

publication of certain material) or, less often, the recipient (as where a person is punished for possession of child pornography).[50] Addressing regulation to intermediaries is not unprecedented (consider, for example, the liability of printers and distributors in defamation) but has been less common. The growth of filtering, with its focus on intermediaries, threatens to alter this in a way which might change the dynamic of regulation.

At the outset, filtering which is implemented by intermediaries is inherently more opaque, lacking as it does any necessity that speaker or recipient be notified. We have already noted that in many existing systems site owners and users alike may not be aware either that filtering is in operation or that particular sites are blocked. This is not a necessary characteristic of filtering – for example, libraries in the United States have been active in informing their patrons that legally required filtering systems are in place.[51] However, not all intermediaries may share the ideological commitment to transparency or freedom of expression which would lead them to do this.

Filtering by intermediaries also increases our concerns about the application of the rule of law. Decisions to require filtering are often made by public authorities, even though others are responsible for their implementation. Compliance with some version of the rule of law is a key part of the legitimating apparatus for public authority decision makers, but may be lacking in the case of filtering by intermediary.

In some instances, such as the Australian Interactive Gambling Act 2001, there is specific legal authority for a public body to investigate particular content, make determinations and issue notices requiring ISPs to block access to that content.[52] But more problematic is the situation where government uses its inherent steering capacity, without legislation, to encourage ISPs or other intermediaries to engage in content filtering.

For example, in the UK the Government has encouraged ISPs to engage in filtering as part of self-regulation. This was initially done by way of consultation and cooperation with the incumbent and dominant operator, British Telecom, which developed a "Cleanfeed" system to automatically block customer access to URLs alleged to host child pornography, the list of blocked URLs being maintained by the Internet Watch

[50] See e.g. S. Kreimer, "Censorship by Proxy: The First Amendment, Internet Intermediaries and the Problem of the Weakest Link" (2006) 155 *University of Pennsylvania Law Review* 11 at 13: "The archetypal actors in the First Amendment drama appear on stage in dyads: in free speech narratives, a speaker exhorts a listener; in free press accounts, a publisher distributes literature to readers. In the usual plot, the government seeks to disrupt this dyad (for legitimate or illegitimate reasons) by focusing sanctions on the source of the speech... On occasion, the government turns its efforts to the listener, seeking to punish receipt of illicit messages or possession of illicit materials preparatory to reading them, and the courts proceed to evaluate the constitutionality of those proposed sanctions."

[51] See American Library Association, "Access to Electronic Information, Services and Networks", 19 January 2005, available at http://www.ala.org/Template.cfm?Section=interpretations&Template=/ContentManagement/ContentDisplay.cfm&ContentID=133993. This provides that "Users' access should not be restricted or denied for expressing or receiving constitutionally protected speech. If access is restricted or denied for behavioral or other reasons, users should be provided due process, including, but not limited to, formal notice and a means of appeal." The ALA has been active in opposing federally required filtering systems, notably in *United States v. American Library Association* 539 U.S. 194 (2003).

[52] Interactive Gambling Act (Cwlth) 2001, s.24.

Foundation.[53] Now, however, the Government has indicated its intention to ensure that all UK ISPs should adopt either "Cleanfeed" or a similar system, with the threat of legislation should ISPs fail to do so "voluntarily".[54]

This presents a number of challenges for the rule of law. Even if an individual ISP's actions can be described as voluntary, the effect is to subject users without their consent to a state mandated regime of internet filtering of which they may be unaware. The Internet Watch Foundation (IWF), which determines which URLs should be blocked, also has a curious legal status, being a charitable incorporated body, funded by the EU and the internet industry, but working closely with the Home Office, the Ministry of Justice, the Association of Chief Police Officers and the Crown Prosecution Service.[55] There is no provision for site owners to be notified that their sites have been blocked.[56] While there is an internal system of appeal against the designation of a URL to be blocked, that mechanism does not provide for any appeal to a court – instead, the IWF will make a final determination on the legality of material in consultation with a specialist unit of the Metropolitan Police.[57]

Consequently the effect of the UK policy is to put in place a system of censorship of internet content, without any legislative underpinning, which would appear (by virtue of the private nature of the actors) to be effectively insulated from judicial review.[58] Though the take-up of the regime may be attributable to the steering actions of government, the way in which the regime is implemented and administered complies neither with the process or transparency expectations which would attach to legal instruments.

There is also cause for concern about the incentives which delegating filtering to intermediaries might create. From the point of view of the regulator, requiring intermediaries to filter may allow them to externalise the costs associated with monitoring and blocking, perhaps resulting in undesirably high levels of censorship.[59]

[53] M. Bright, "BT puts block on child porn sites" *The Observer* 6 June 2004. Available at http://www.guardian.co.uk/technology/2004/jun/06/childrensservices.childprotection. See also Hunter, "BT Siteblock" (2004) 9 *Computer Fraud and Security* 4.

[54] W. Grossman, "The Great Firewall of Britain" *net.wars* 24 November 2006, quoting Vernon Coaker, Parliamentary Under-Secretary for the Home Department: "We believe that working with the industry offers us the best way forward, but we will keep that under review if it looks likely that the targets will not be met." Available at http://www.pelicancrossing.net/netwars/2006/11/the_great_firewall_of_britain.html.

[55] See, for example, the "Memorandum of Understanding Between Crown Prosecution Service (CPS) and the Association of Chief Police Officers (ACPO) concerning Section 46 Sexual Offences Act 2003" dated 6 October 2004, available at http://www.iwf.org.uk/documents/20041015_mou_final_oct_2004.pdf, which gives special recognition to the role of the IWF. See generally Internet Watch Foundation, "About the Internet Watch Foundation" available at http://www.iwf.org.uk/public/page.103.htm.

[56] Internet Watch Foundation, "Child Sexual Abuse Content URL List" available at http://www.iwf.org.uk/public/page.148.437.htm.

[57] Internet Watch Foundation, "Child Sexual Abuse Content URL Service: Complaints, Appeals and Correction Procedures" available at: http://www.iwf.org.uk/public/page.148.341.htm.

[58] As Akdeniz puts it "When censorship is implemented by government threat in the background, but run by private parties, legal action is nearly impossible, accountability difficult, and the system is not open or democratic." Y. Akdeniz, "Who Watches the Watchmen? The role of filtering software in Internet content regulation" in C. Moller and A. Amouroux (eds.), *The Media Freedom Internet Cookbook* (Vienna, 2004), at 111.

[59] S. Kreimer, "Censorship by Proxy: The First Amendment, Internet Intermediaries and the Problem of the Weakest Link" (2006) 155 *University of Pennsylvania Law Review* 11 at 27.

But perhaps more worrying are the incentives which filtering creates for intermediaries. Kreimer has argued that by targeting online intermediaries regulators can recruit "proxy censors", whose "dominant incentive is to protect themselves from sanctions, rather than to protect the target from censorship".[60] As a result, there may be little incentive for intermediaries to engage in the costly tasks of distinguishing protected speech from illegal speech, or to carefully tailor their filtering to avoid collateral damage to unrelated content. Kreimer cites the US litigation in *Centre for Democracy & Technology v. Pappert*[61] to illustrate this point – in that case more than 1,190,000 innocent web sites were blocked by ISPs even though they had been required to block fewer than 400 child pornography web sites.

4. Responsibility

A central objection to technology as regulator generally is, that to the extent that otherwise available choices for human action are inhibited, there is a loss of responsibility for one's actions. We are accustomed to assuming moral responsibility for actions which are within an acceptable range of possible actions. If actions outside the acceptable range are simply impossible, then we need no longer engage in moral choice, since our actions will, of necessity be acceptable. This effect, Brownsword has suggested, may be corrosive of our moral capacity.[62] Where restrictions on unacceptable conduct are created through technology in some social domains (such as the internet) it creates the risk that our moral capacity to act in other, less restricted, domains will be reduced, with adverse social consequences. Or, as Spinello argues, "code should not be a surrogate for conscience".[63]

Perhaps paradoxically, the converse may also be true – the fact that technology makes certain acts easier to perform may in some contexts reduce the moral or legal responsibility of users for those acts. If something is easy to do it may be less clear that it is illegal. Zittrain[64] for example has argued that:

> The notion that some content is so harmful as to render its transmission, and even reception, actionable – true for certain categories of both intellectual property and pornographic material – means that certain clicks on a mouse can subject a user to intense sanctions. Consumers of information in traditional media are alerted to the potential illegality of particular content by its very rarity; if a magazine or CD is available in a retail store its contents are likely legal to possess. The Internet severs much of that signaling, and the ease with which one can execute an Internet search and encounter illegal content puts users in a vulnerable position. Perhaps the implementation of destination ISP-based filtering, if pressed, could be coupled with immunity for users for most categories of that which they can get to online in the natural course of surfing.

Taken further, Zittrain's argument suggests where technical controls on behaviour are in place users may come to believe, or the law may come to accept, that those online actions which are not blocked by some technical means are permissible. Indeed, a similar viewpoint is already reflected in many national laws which criminalise

[60] *Ibid.* at 28.

[61] 337 F. Supp 2d. 606 (ED Pa 2004)

[62] R. Brownsword, "Code, Control, and Choice: Why East is East and West is West" (2005) 25 *Legal Studies* 1 and "Neither East Nor West, Is Mid-West Best?" (2006) 3(1) *SCRIPT-ed.*

[63] R. Spinello, "Code and Moral Values in Cyberspace" (2001) 3 *Ethics and Information Technology* 137.

[64] J. Zittrain, "Internet Points of Control". (2003) 43 *Boston College Law Review* 1 at 36.

unauthorised access to a computer system only if the user has circumvented some technical security measure protecting that system.[65]

In the case of filtering, these arguments may intersect to suggest that pervasive filtering may reduce the moral accountability of users significantly both by reducing their capacity to make moral choices and by signalling to them that those actions which are not blocked are permissible.

5. Conclusions

Filtering is likely to remain amongst the most important technologies mediating between users and suppliers of content, but presents significant issues relating to the legitimacy and accountability of particular types of blocking.

Where the purposes of the filtering are those of the user it is not difficult to imagine systems for filtering which meet most of the normative requirements discussed in this chapter. Users may opt in where they have a need to do so and the system may have feedback so that users can see corrections or opt out if the filtering mechanism is insufficiently accurate to meet their purposes. Many email filtering systems have these properties. A third feature of such systems – that they are designed and operated by commercial firms – raises few concerns in this context since a user who is dissatisfied with the way the system works is able to migrate to a different provider. Such systems are, in effect, regulated through competition.

An filtering system which is applied by an intermediary (rather than a user) and which lacks transparency, because the user does not know it has been applied, or cannot see which messages are filtered out is weak in two respects – it lacks consent, and it lacks a feedback mechanism to correct for technical weaknesses in the system. For example, a user may be aware of false negatives, because spam email will reach their inbox, but will be unable to detect false positives where email they wanted to receive was discarded.

Much filtering is, of course, directed not at the purposes of the user but rather at broader public purposes, such as the blocking of offensive, controversial or illegal internet content. In some instances parents may be choosing to apply filtering to protect children. There is here an element of consent. However, many such regimes lack transparency and feedback mechanisms such that over-inclusive control, which blocks sites which parents would not have sought to block, is not systematically addressed within the system.

We have noted that some governmental regimes for blocking internet content, while they lack consent from users, nevertheless contain elements of transparency, because users are told that sites are blocked, and elements of feedback, because users are invited to inform operators of the system if they think a site has been blocked in error. Similarly, some regimes may comply with the requirements of the rule of law where there is legislative authorisation for the filtering and the determination of content to be

[65] See the discussion in S.M. Kierkegaard, "Here comes the 'cybernators'!", (2006) 22(5) *Computer Law & Security Report* 381. O. Kerr, "Cybercrime's Scope: Interpreting 'Access' and 'Authorization' in Computer Misuse Statutes" (2003) 78 *New York University Law Review* 1596 suggests that this approach should apply to unauthorised access offences generally.

blocked is made by a public body with appropriate safeguards and mechanisms of appeal in place.

Regimes which lack consent, transparency, and feedback mechanisms are open to two basic objections. First that they are not amenable to correction where they operate in an over- (or under-) inclusive manner and second that they remove responsibility from users. Even where governments maintain control over such regimes these weaknesses are significant and difficult to justify. *A fortiori* the most challenging regimes are those with these properties operated by commercial firms either at the request or command of governments, or for the own purposes of firms.