THE BEGINNING OF THE END OF INTERNET FREEDOM

DAWN C. NUNZIATO*

ABSTRACT

Although the Internet was initially viewed as a medium for expression in which censorship would be impossible to implement, recent developments suggest exactly the opposite. Countries around the world—democracies as well as dictatorships—have implemented nationwide filtering systems that are changing the shape of Internet freedom. In addition to usual suspects like China, liberal democracies such as the United Kingdom and Australia have taken steps to implement nationwide Internet filtering regimes. In 2013, United Kingdom Prime Minister David Cameron announced a plan to require mandatory "family friendly" default filtering of all Internet access by the end of 2014. While such Internet filtering regimes may have laudable goals—like preventing children from accessing harmful content and preventing access to illegal child pornography—they inevitably lead to overblocking of harmless Internet content and present grave dangers of censorship.

International protections for freedom of expression, as well as the United States' protections for First Amendment freedoms, provide not only substantive but also procedural protections for speech. These procedural protections are especially important for countries to observe in the context of nationwide Internet filtering regimes, which embody systems of prior restraint. Prior restraints on speech historically have been viewed with great suspicion by courts and any system of prior restraint bears a strong presumption of unconstitutionality. To mitigate the dangers of censorship inherent in systems of prior restraint such as those embodied in nationwide filtering systems, any country adopting such a system should provide the requisite procedural safeguards identified in international and U.S. law, including (1) by providing affected Internet users with the ability to challenge the decision to filter before an independent judicial body, (2) by providing meaningful notice to affected Internet users that content was filtered, and (3) by clearly, precisely, and narrowly defining the categories of speech subject to filtering.

---

* Professor of Law, George Washington University Law School, B.A., M.A. (Philosophy), J.D., University of Virginia. I am grateful for the financial support of Interim Dean Greg Maggs, as well as for the superlative research assistance of Gregory W. Kubarych.

I. INTRODUCTION

It was long assumed that the Internet would bring about greater opportunities for free expression than any other medium. In recent years, however, the Internet has increasingly become a tool of censorship, as scores of countries around the world have imposed nationwide filtering regimes to block their citizens' access to various types of Internet speech that they deem harmful. Instead of trending toward greater freedom, the Internet is now trending toward greater censorship and control, as many countries—including democracies such as the United Kingdom—are seeking to exercise greater and greater control over this medium.[1] Today, more than forty countries—in addition to the usual suspects like China, Saudi Arabia, and North Korea—have implemented nationwide filtering of speech on the Internet, and this number is growing.[2] Among democracies, the United Kingdom and Australia are leading the way in implementing filters to attempt to control their citizens' access to harmful content. Indeed, British Prime Minister David Cameron announced in July

1. *See* Jonathan Zittrain & John Palfrey, *Introduction*, *in* ACCESS DENIED: THE PRACTICE AND POLICY OF GLOBAL INTERNET FILTERING 2, 3 (Ronald Deibert et al. eds., 2008) (noting that "more than three dozen states around the world now filter the Internet," including some in Europe).

2. According to the Open Net Initiative, over three dozen countries around the world now impose state-mandated technical filtering of speech on the Internet. *See id.* at 2 (setting forth its systematic global study of all known state-mandated Internet filtering practices). For a general discussion of how nation-by-nation Internet filtering is implemented, including specifics of the technical details of such filtering, see Steven J. Murdoch & Ross Anderson, *Tools and Technology of Internet Filtering*, *in* ACCESS DENIED: THE PRACTICE AND POLICY OF GLOBAL INTERNET FILTERING 57, 57-72 (Ronald Deibert et al. eds. 2008).

that mandatory "family-friendly" filters will be imposed by default on all new computers by the end of 2013.[3] By the end of 2014, such filters will be imposed by default on all existing computers as well, creating a virtual tyranny of the default.[4] Such extensive mandatory filtering builds on the solid foundation of the United Kingdom's comprehensive system for filtering and blocking harmful Internet content through the mechanisms set in motion by an entity known as the Internet Watch Foundation. The United Kingdom's experience is paralleled by similar events in Australia. For much of 2012, the Australian government attempted to introduce mandatory internet service provide (ISP)-level filtering of certain content, requiring all Australian ISPs to block content dealing with "matters of sex, drug misuse or addiction, crime, cruelty, violence or revolting or abhorrent phenomena in such a way that they offend against the standards of morality, decency and propriety generally accepted by reasonable adults."[5] Although political considerations caused this plan to be formally withdrawn, it is likely that Australia will remain in the business of Internet filtering for the foreseeable future.[6] In short, liberal democratic regimes, as well as authoritarian regimes, are now implementing unprecedented restrictions on Internet content and changing the face of Internet freedom.

Nationwide Internet filtering has become a powerful tool for many governments to control the content that their citizens are able to access. Given the extent and increasing effectiveness of efforts to censor Internet speech throughout the world, protectors of Internet free speech can no longer rest comfortably on the assurance given by Internet pioneer John Gilmore two decades ago that "the Net inter-

3. *See Online Pornography to Be Blocked by Default, PM Announces*, BBC NEWS, July 22, 2013, http://www.bbc.co.uk/news/uk-23401076; Anthony Faiola, *Britain's Harsh Crackdown on Internet Porn Prompts Free-Speech Debate*, WASH. POST, Sept. 28, 2013, http://www.washingtonpost.com/world/europe/britains-harsh-crackdown-on-internet-porn-prompts-free-speech-debate/2013/09/28/d1f5caf8-2781-11e3-9372-92606241ae9c_story.html.

4. *See* Faiola, *supra* note 3.

5. For further detail on each of these examples of state-mandated Internet filtering, see *Action Alert-Australian Internet Censorship*, ELECTRONIC FRONTIERS AUSTRALIA, Mar. 31, 1999, https://www.efa.org.au/Campaigns/alert99i.txt (detailing Australia's state-mandated filtering plans); ACCESS CONTROLLED: THE SHAPING OF POWER, RIGHTS, AND RULE IN CYBERSPACE (Ronald Deibert et al. eds., 2010); OPENNET INITIATIVE, http://opennet.net/ (last visited Dec. 8, 2013).

6. Successive Australian governments have proposed nationwide Internet filtering and blocking systems similar to those in the United Kingdom. However, for a variety of reasons—logistical difficulties in designing a working filter and broader public opinion problems—Australia in the end chose not to implement a filtering system.

prets censorship as damage and routes around it."[7] Although free speech advocates broadly denounce such censorship, it is likely that many countries—having seized upon these powerful tools of control—will continue to restrict Internet content to prohibit their citizens from accessing speech that they deem to be harmful.

It is commonly understood—and understandable—that different countries around the world adopt different definitions of what speech is protected and what speech is unprotected, online as well as offline. Given, for example, European countries' horrific experiences with the Holocaust, it is not surprising that some of these countries consider racial and religious hate speech to be unprotected. While there is substantial divergence on the *substantive* contours of free speech protections[8]—which categories of speech are protected and which are not—there is some convergence among nations regarding *procedural* protections for speech.[9] These procedural protections are inherent in and flow from widely-shared concepts of fundamental due process, and have been embodied in the widely-adopted International Covenant on Civil and Political Rights (the ICCPR), as recently construed by the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression to provide procedural protections for Internet speech.[10] A second important source of procedural protections for Internet speech is the Anglo-American tradition's hostility toward prior restraints on speech and (relative) preference for subsequent punishment as a means of restricting expression. Over the past four hundred years, Anglo-American jurisprudence has developed a presumption against the legality of any prior censorship or prior restraints on expression and has imposed a set of procedural safeguards that must be in place before any system of prior restraint can be

---

7. Philip Elmer-Dewitt, *First Nation in Cyberspace*, TIME, Dec. 6, 1993, http://content.time.com/time/magazine/article/0,9171,979768,00.html.

8. *See, e.g.*, RONALD J. KROTOSZYNSKI, JR., THE FIRST AMENDMENT IN CROSS-CULTURAL PERSPECTIVE: A COMPARATIVE LEGAL ANALYSIS OF THE FREEDOM OF SPEECH (2006); Robert A. Sedler, *Freedom of Speech: The United States Versus the Rest of the World*, 2006 MICH. ST. L. REV. 377 (2006); Stephanie Farrior, *Molding the Matrix: The Historical and Theoretical Foundations of International Law Concerning Hate Speech*, 14 BERKELEY J. INT'L L. 1 (1996).

9. *See* International Covenant on Civil and Political Rights, G.A. Res. 2200, U.N. GAOR, 21st Sess., Supp. No. 16, at 52, U.N.Doc A/6316 (1966) [hereinafter ICCPR]; Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, U.N. Doc. A/HRC/17/27 (May 16, 2011) (by Frank La Rue) [hereinafter *Special Rapporteur's Report*].

10. *See* ICCPR, *supra* note 9; *Special Rapporteur's Report*, *supra* note 9.

legally imposed.[11] The nationwide filtering systems that have become increasingly pervasive in the past few years embody *prior restraints* on speech—restrictions on speech imposed prior to a judicial determination of the speech's illegality[12]—and fail to accord these important procedural protections for speech embodied in the International Covenant on Civil and Political Rights and in Anglo-American free speech jurisprudence. Nationwide filtering systems should either be jettisoned or revised so as to accord these fundamental procedural protections on speech.

In Part II, this Article focuses on the nationwide filtering systems espoused by the United Kingdom. Part III analyzes the procedural protections on free speech that have been articulated both under the International Covenant on Civil and Political Rights and within the U.S. legal tradition, especially with respect to prior restraints on speech. In Part IV, the Article compares the procedural protections provided under the United Kingdom's Internet filtering system with the protections required within U.S. jurisprudence and the ICCPR, and finds this nationwide filtering system to be lacking. I propose modifications to this system and suggest that if nationwide filtering systems are to be imposed, certain procedural safeguards must be implemented within any such system—especially those that are imposed within liberal democracies.

A. *The United Kingdom's Expanding Nationwide Filtering System*

> Journey with us to a state where an unaccountable panel of censors vets 95 per cent of citizens' domestic internet connections. The content coming into each home is checked against a mysterious blacklist by a group overseen by nobody, which keeps secret the list of censored URLs not just from citizens, but from internet service providers themselves. And until recently, few in that country even knew the body existed. Are we in China? Iran? Saudi Arabia? No—the United Kingdom. . . .[13]

11. *See, e.g.*, ERWIN CHEMERINSKY, CONSTITUTIONAL LAW: PRINCIPLES AND POLICIES 949, 964-65 (3d ed. 2006).

12. *See infra* text accompanying notes 74-100.

13. C.J. Davies, *The Hidden Censors of the Internet*, WIRED UK, May 20, 2009, http://www.wired.co.uk/magazine/archive/2009/06/features/the-hidden-censors-of-the-internet?page=1.

To understand what is at stake in a nationwide filtering system like that adopted—and soon to be expanded—in the United Kingdom, and how the system implicates the rights of Internet users, consider the operation of a filtering scheme translated to the real space context. Imagine a vast real space forum for authors and readers in which millions of authors bring their books to be made available for billions of potentially interested readers. The authors place their books on the bookshelves of the forum and then depart. Billions of readers also come to the forum to search for books of potential interest to them. Unbeknownst to either the authors or the readers, before the content of any book is made available to the readers—or at some point after the books are placed on the bookshelves—the books are scrutinized by unseen and unknown censors to determine whether the content is "permissible," according to some criteria that are unstated and undiscoverable. If these censors determine that a book or some of its content is impermissible, it is placed on a blacklist and removed from circulation. When the readers enter the forum to search for books of potential interest to them, they do not know which books have been removed, nor do the authors of the banned books ever learn whether (or why) their books have been removed. This scenario replicates in real space what occurs in cyberspace under a nationwide filtering system like that in operation in the United Kingdom when websites are placed on blacklists and the country's Internet users are prohibited from accessing such content.

In July 2013, U.K. Prime Minister David Cameron surprised the world by announcing that "family-friendly" filters would soon be imposed by default on all computers in the nation.[14] Such filters would affect both wireline and wireless Internet access.[15] Cameron offered no details as to how the filters would work or exactly which categories of speech they would attempt to block—other than that ISPs themselves would administer the program.[16] If, as is likely, the new nationwide filtering system replicates the system already in place affecting mobile Internet access, it is likely that the following categories of content will be blocked by default: pornography, violent material, extremist and terrorist related content, anorexia and eating disorder websites, suicide related websites, alcohol, smoking, web forums, esoteric material, and,

14. David Cameron, U.K. Prime Minister, The Internet and Pornography: Prime Minister Calls for Action (July 22, 2013), *available at* https://www.gov.uk/government/speeches/the-internet-and-pornography-prime-minister-calls-for-action.

15. *See id.*

16. *Id.*

of course, web blocking circumvention tools.[17] Cameron did not offer any details about what remedies could be taken by citizens if legal material is mistakenly blocked by such filters, about which entities would be responsible for creating such filters and determining which websites fell under the prohibited categories, or about whether such entities would be accountable to the public in any way.[18]

This new and comprehensive initiative to impose extensive filters on the nation's Internet access builds upon the United Kingdom's pre-existing nationwide filtering system, which already affects 99% of its residential Internet subscribers.[19] This system, implemented in 1996, blocks access to websites containing certain types of content that are deemed "potentially illegal" by an unaccountable organization known as the Internet Watch Foundation (IWF).[20] Below, this Article describes the evolution of the IWF, the development of the United Kingdom's initial nationwide filtering system, and some of the difficulties posed by this system, and by the proposed expansion of this system.

In 1996, as concerns arose about illegal content being hosted and facilitated by ISPs in the United Kingdom, the U.K. Department of Trade and Industry facilitated discussions between the Metropolitan Police, the Home Office, and a group of ISPs with an eye toward remedying these concerns.[21] These discussions resulted in an agreement whereby a private, charitable organization, the Internet Watch Foundation, was formed and charged with policing the Internet for child pornography.[22] In 1999, after three years of the IWF's operation, the U.K. government and Department of Trade and Industry evaluated

17. *See* Jim Killock, *Sleepwalking into Censorship*, OPEN RTS. GROUP (July 25, 2013), https://www.openrightsgroup.org/blog/2013/sleepwalking-into-censorship.

18. *See* Cameron, *supra* note 14.

19. *See FAQs Regarding the IWF's Facilitation of the Blocking Initiative*, INTERNET WATCH FOUND., https://www.iwf.org.uk/members/member-policies/url-list/blocking-faqs (last visited Dec. 1, 2013) ("The government announced in October 2009 that it is pleased with the support from providers which has led to 98.6 per cent of UK consumer broadband lines being covered by the blocking of child sexual abuse content as identified by the Internet Watch Foundation.").

20. *See IWF Funding Council Code of Practice for Notice and Takedown of UK Hosted Content Within IWF Remit*, INTERNET WATCH FOUND., https://www.iwf.org.uk/members/member-policies/funding-council/code-of-practice#F4 (last visited Nov. 20, 2013) ("The IWF operates a notice and takedown service to issue Notices alerting hosting service providers . . . to content hosted on their servers in the UK that it has assessed as potentially illegal.").

21. See *IWF History*, INTERNET WATCH FOUND., https://www.iwf.org.uk/about-iwf/iwf-history (last visited Jan. 7, 2014).

22. *Id.*

and ultimately endorsed the operations of the IWF.[23] The IWF is currently responsible for policing and facilitating the filtering or blocking of at least two types of potentially illegal content: (1) child sexual abuse content hosted anywhere in the world (child pornography); and (2) criminally obscene content ("extreme pornography") hosted in the United Kingdom.[24] The IWF monitors and polices the Internet's content by operating a hotline reporting system, through which U.K. Internet users alert the IWF to such potentially illegal content that they have come across on the Internet.[25] The IWF then employs a handful of analysts trained by police to review flagged websites to analyze whether they contain "potentially illegal content."[26] If an IWF analyst determines that a website is "potentially illegal," she will include the URL for that website on the IWF blacklist.[27] The blacklist, which contains between 500 and 800 sites at any given time and is updated twice daily, is maintained in secret.[28] Approximately 99% of all domestic broadband connections are filtered in compliance with the IWF blacklist, and the U.K. government is actively working to secure 100% compliance.[29]

In the words of commentator Lilian Edwards, the United Kingdom's

---

23. *See Chairing the Internet Watch Foundation, Introduction to the IWF*, ROGER DARLINGTON'S WORLD, http://www.rogerdarlington.me.uk/iwf.html#Introduction%20To%20The%20IWF (last visited Jan. 10, 2014).

24. *See Laws Relating to the IWF's Remit*, INTERNET WATCH FOUND., https://www.iwf.org.uk/hotline/the-laws (last visited Nov. 20, 2013).

25. *See IWF Funding Council Code of Practice for Notice and Takedown of UK Hosted Content Within IWF Remit*, *supra* note 20 ("The IWF operates a notice and takedown service to issue Notices alerting hosting service providers . . . to content hosted on their servers in the UK that it has assessed as potentially illegal."); *FAQs Regarding the IWF's Facilitation of the Blocking Initiative*, *supra* note 19 ("Q: Are the images definitely criminal? A: We refer to content as *potentially* criminal because a definitive legal judgment is a matter for the Courts.").

26. *See IWF Funding Council Code of Practice for Notice and Takedown of UK Hosted Content Within IWF Remit*, *supra* note 20 ("The IWF operates a notice and takedown service to issue Notices alerting hosting service providers . . . to content hosted on their servers in the UK that it has assessed as potentially illegal.").

27. *See id.* ("The IWF operates a notice and takedown service to issue Notices alerting hosting service providers . . . to content hosted on their servers in the UK that it has assessed as potentially illegal.").

28. *See FAQs Regarding the IWF's Facilitation of the Blocking Initiative*, *supra* note 19.

29. *See id.* ("The government announced in October 2009 that it is pleased with the support from providers which has led to 98.6 per cent of UK consumer broadband lines being covered by the blocking of child sexual abuse content as identified by the Internet Watch Foundation."); Weixiao Wei, *Online Child Sexual Abuse Content: The Development of a Comprehensive, Transferable International Internet Notice and Takedown System*, INTERNET WATCH FOUND., https://www.iwf.org.uk/assets/media/resources/IWF%20Research%20Report_%20Development%20of%20an%20

implementation of the IWF blacklist "could be the most perfectly invisible censorship mechanism ever invented."[30] First, the system does not provide meaningful notice on a consistent basis of instances of blocking to either the end-user or the content provider whose content is filtered or blocked. Depending upon the method of implementation employed by an ISP, the end-user may or may not receive any indication that the website he or she has requested has been blocked in compliance with the IWF blacklist or the reason for such blocking.[31] Some ISPs, such as British Telecom and Virgin Media, simply return a generic HTTP "404 Not Found" error message when a user requests access to a site that is on the IWF blacklist.[32] This error message does not give the requesting user any indication that the requested page has been blocked or the reason why the page has been blocked.[33] In contrast, other ISPs, such as Demon, return a HTTP "403 Forbidden" error message, which at least provides some indication to the requesting Internet user that the requested site has been blocked because it is "forbidden."[34] The IWF itself does not require that any type of notice be provided to the content provider that its site has been blocked or to the end-user requesting such content.[35]

Second, the filtering system does not provide for a method of appeal

---

international%20internet%20notice%20and%20takedown%20system.pdf (last visited Dec. 1, 2013).

30. Lilian Edwards, *From Child Porn to China, in One Cleanfeed*, SCRIPT-ED, Sept. 2006, http://www.law.ed.ac.uk/ahrc/script-ed/vol3-3/editorial.asp.

31. *See id.*; *Wikimedia, IWF Respond to Block of Wikipedia Over Child Pornography Allegations*, WIKINEWS, Dec. 8, 2008, http://en.wikinews.org/wiki/Wikimedia,_IWF_respond_to_block_of_Wikipedia_over_child_pornography_allegations (noting an instance in which Wikipedia page was blocked, IWF provided no notice to Wikimedia).

32. *See, e.g.*, Becky Hogge, *Lessons and Questions for the IWF*, OPEN RTS. GROUP (DEC. 15, 2008), https://www.openrightsgroup.org/blog/2008/lessons-and-questions-for-the-iwf.

33. See *id.*; *Wikimedia, IWF Respond to Block of Wikipedia Over Child Pornography Allegations*, *supra* note 31.

34. Demon returns the following error message for those attempting to access images on the IWF blacklist: "We have blocked this page because, according to the Internet Watch Foundation (IWF), it contains indecent images of children or pointers to them; you could be breaking UK law if you viewed the page." *See* DEMON BLOCK PAGE, http://www.convergence.cx/demon403.txt (last visited Dec. 28, 2013).

35. The IWF does not require notice, or assume responsibility for notifying, individuals that their content has been required to be blocked under the IWF system. In the "FAQs Regarding the IWF's Facilitation of the Blocking Initiative," the IWF website explains: "Q. Are site 'owners' notified that they have been added to this list? A. Notifying the website owner of any blocked URL is the responsibility of the Hotline or relevant law enforcement agency in the country believed to be hosting the content." *See FAQs Regarding the IWF's Facilitation of the Blocking Initiative*, *supra* note 19.

for an independent judicial determination of whether the blocked content is in fact illegal. The IWF website indicates that "any party with a legitimate association with the [blacklisted] content . . . who believes they are being prevented from accessing legal content may appeal against the accuracy of an assessment."[36] The appeal procedure provided by the IWF, however, does not contemplate judicial review. Rather, the contemplated appeal merely involves a second look by the IWF itself with no input from or representation of the affected content provider or end user, and following that, a review by a police agency, whose assessment is final.[37] Further, as discussed above, it is unclear in many cases how a party would learn that the content she was seeking, or seeking to make available, was subject to the IWF's blacklist, since the IWF does not require that notice be provided to the affected parties.

In summary, the United Kingdom's nationwide filtering system based on the IWF blacklist operates as an opaque, non-transparent system that does not provide end-users or content providers with meaningful notice that a potentially illegal website has been blocked. The system operates so as to place ultimate authority over Internet content in an unaccountable, nontransparent body. Not surprisingly, this has led to instances of overblocking.[38] Although it is difficult to secure meaningful information regarding how many sites on the IWF's blacklist are actually illegal, some clear instances of IWF overblocking have become subject to public scrutiny, as I discuss below.[39]

In December 2008, acting on a hotline notification from an Internet user, the IWF placed a Wikipedia article discussing an album by the popular German rock band The Scorpions (of "Rock You Like a Hurricane" fame) on its blacklist.[40] The cover of The Scorpions' 1976 album *Virgin Killer* depicted a pre-pubescent girl unclothed, but with her genitalia obscured from view. The IWF had determined that the cover art was "potentially illegal," and placed the Wikipedia article

36. *See id.*

37. *See Content Assessment Appeal Process*, INTERNET WATCH FOUND., http://www.iwf.org.uk/accountability/complaints/content-assessment-appeal-process (last visited Dec. 1, 2013).

38. *See infra* text accompanying notes 40-48.

39. *Id.*

40. *See, e.g.*, Cory Doctorow, *How to Make Child-Porn Blocks Safe for the Internet*, GUARDIAN, Dec. 16, 2008, http://www.theguardian.com/technology/2008/dec/16/cory-doctorow-wikipedia; Frank Fisher, *A Nasty Sting in the Censor's Tail*, GUARDIAN, Dec. 9, 2008, http://www.theguardian.com/commentisfree/2008/dec/09/scorpions-virgin-killer-censorship; Cade Metz, *IWF Pulls Wikipedia from Child Porn Blacklist*, REGISTER, Dec. 10, 2008, http://www.theregister.co.uk/2008/12/10/iwf_reverses_wikiban/; Jane Fae Ozimek, *Scorpions Tale Leaves IWF Exposed*, REGISTER (Dec. 9 2008), http://www.theregister.co.uk/2008/12/09/iwf/?page=1.

discussing the album on its blacklist, without notifying the content provider of its decision to blacklist the website.[41] As a result, the vast majority of U.K. residential users was unable to access this content and was not informed as to why the content was blocked.[42]

As a result of the general public awareness and outcry regarding the consequences of the IWF's blocking of the Virgin Killer page, the IWF removed the page from its blacklist but nonetheless maintained that the image of the album cover was potentially illegal.[43] As Mike Godwin, General Counsel for Wikimedia, explained, "When we first protested the block, [the IWF's] response was, 'We've now conducted an appeals process on your behalf and you've lost the appeal.' When I asked who exactly represented the Wikimedia Foundation's side in that appeals process, they were silent. It was only after the fact of their blacklist and its effect on U.K. citizens were publicised that the IWF appears to have felt compelled to relent."[44]

In January 2009, in another incident of massive overblocking, U.K. Internet users learned that all 85 billion pages of the Wayback Machine—the application that archives the Internet's content—had been blocked, apparently because the archive contained one or more URLs that were on the IWF's blacklist.[45]

Most recently, in November 2011, the implementation of the IWF blacklist blocked U.K. subscribers of the ISP Virgin Media from accessing any files from the popular Fileserve file-hosting service—one of the ten most popular file-sharing sites on the Internet—which allows users to store and share files in the cloud.[46] IWF apparently intended to blacklist only certain Fileserve URLs, but the effect of its placing these URLs on its blacklist (as the blacklist was implemented by Virgin Media) was to block any and all files from being downloaded from Fileserve.[47] After this latest instance of overblocking by ISPs pursuant to IWF's blacklist, Jim Killock—head of online rights activists the Open

41. *See* Doctorow, *supra* note 40; Metz, *supra* note 40.

42. Davies, *supra* note 13.

43. *Id.*

44. *Id.*

45. Cade Metz, *Brit Porn Filter Censors 13 Years of Net History*, REGISTER, Jan. 14, 2009, http://www.theregister.co.uk/2009/01/14/demon_muzzles_wayback_machine/ ("Britain's child porn blacklist has led at least one ISP to muzzle the Internet Archive's Wayback Machine—an 85 billion page web history dating back to 1996.").

46. David Meyer, *Blacklist Hitch Causes Virgin Media FileServe Block*, ZDNET, Nov. 21, 2011, http://www.zdnet.com/blacklist-hitch-causes-virgin-media-fileserve-block-3040094466/.

47. Tom Jowitt, *IWF Blacklist Snafu Blocks Access to FileServe*, TECHWEEKEUROPE, Nov. 21, 2011, http://www.techweekeurope.co.uk/news/iwf-blacklist-snafu-blocks-access-to-fileserve-46702.

Rights Group—noted that the accidental blocking of legitimate content has become "quite a common occurrence."[48]

Under Prime Minister Cameron's plan for expanded "family friendly" Internet filtering, a far wider range of websites will be blocked than the categories of child pornography and extreme pornography that are currently blocked under the IWF filtering system.[49] Cameron's filtering plan fundamentally shifts the default of Internet usage towards censorship, presenting serious problems for freedom of expression in the U.K., as discussed in greater detail in Part III.

II. Procedural Protections for Speech

A. *International Procedural Protections for Speech*

International treaties and documents of international law provide protections for speech that must be respected in the context of nationwide Internet filtering systems.[50] The International Covenant on Civil and Political Rights (ICCPR),[51] which has been adopted by 167 parties (including the United Kingdom), and is considered a binding international law treaty, provides in its Article 19 that:

> 1. Everyone shall have the right to hold opinions without interference.
> 2. Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice.[52]
> 3. [These rights] may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary:
>   (a) For respect of the rights or reputations of others;
>   (b) For the protection of national security or of

48. *See* Meyer, *supra* note 46.

49. To view the breadth of sites affected by current mobile carrier filters and which will presumably serve as a baseline for the Cameron program, see Jim Killock, *Sleepwalking into Censorship*, Open Rights Group (July 25, 2013), https://www.openrightsgroup.org/blog/2013/sleepwalking-into-censorship.

50. *See, e.g.*, ICCPR, *supra* note 9.

51. *Id.*

52. *Id.* art. 19.

> public order (ordre public), or of public health or morals.[53]

The right to freedom of expression is also protected under the European Convention for the Protection of Human Rights, which has been signed by 47 nations (including the United Kingdom), is considered binding, and is enforced by the European Court of Human Rights.[54] Article 10 of the European Convention provides:

> Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers . . . .
>
> The exercise of these freedoms . . . may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary.[55]

These international protections for free speech not only have a *substantive* dimension of which categories of speech to protect and which to restrict—which differ from country to country[56]—but such protections also have important *procedural* dimensions, which require that "sensitive tools" be implemented to distinguish between protected and unprotected speech.[57] As free speech theorist Henry Monaghan explains, "procedural guarantees play an equally large role in protecting freedom of speech; indeed, they 'assume an importance fully as great as

53. *Id.* art. 19. The ICCPR provides further, in Article 20, that any propaganda for war or advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence, is prohibited by law. *Id.* art. 20.

54. European Convention for the Protection of Human Rights and Fundamental Freedoms (ETS 5), 213 U.N.T.S. 222 (entered into force Sept. 3, 1953, as amended by Protocol 11 (ETS 155) which entered into force May 11, 1994) [hereinafter European Convention].

55. *See id.*

56. *See, e.g.*, KROTOSZYNSKI, *supra* note 8; Sedler, *supra* note 8; Farrior, *supra* note 8.

57. Bantam Books, Inc. v. Sullivan, 372 U.S. 58, 66 (1963).

the validity of the substantive rule of law to be applied.'"[58]

While there is substantial variation in the substantive protections for speech, there is more widespread agreement regarding the procedures that are essential to ensure protection for the categories of speech that are deemed legal. These procedural requirements were recently articulated in the Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/HRC/17/27,[59] and were expanded upon in a recent decision of the European Court of Human Rights. While recognizing that countries enjoy some discretion to restrict speech that constitutes child pornography, hate speech, defamation, incitement to genocide, discrimination, hostility or violence, the Special Rapporteur's Report explains that:

> Any [such] limitation to the right to freedom of expression must pass the following [multi]-part, cumulative test:
>
> (a) It must be provided by law, which is clear and accessible to everyone (principles of predictability and transparency); and
>
> (b) It must pursue one of the purposes set out in article 19, paragraph 3, of the [International Covenant on Civil and Political Rights], namely (i) to protect the rights or reputations of others, or (ii) to protect national security or of public order, or of public health or morals (principle of legitimacy); and
>
> (c) It must be proven as necessary and the least restrictive means required to achieve the purported aim (principles of necessity and proportionality) [;]
>
> [It] must be applied by a body which is independent of any political, commercial, or other unwarranted influences in a manner that is neither arbitrary nor discriminatory, and with adequate safeguards against abuse, including the possibility of challenge and remedy against its abusive application.[60]

The case of *Yildirim v. Turkey*,[61] decided by the European Court of Human Rights in December 2012, was the first to consider the issue of Internet censorship through the lens of the procedural protections for

58. Henry Monaghan, *First Amendment "Due Process"*, 83 HARV. L. REV. 518, 518 (1970) (quoting Speiser v. Randall, 357 U.S. 513, 520 (1958)).

59. *See Special Rapporteur's Report, supra* note 9.

60. *Id.* ¶ 24.

61. Yildirim v. Turkey (No. 3111/10), Eur. Ct. H.R. (2012).

speech articulated in the Special Rapporteur's report. In that case, Mr. Yildirim, an Turkish graduate student and blogger, challenged the Turkish government's blocking of his academic website on Google Sites (a platform for creation and hosting of websites).[62] The Turkish government had ordered the entire platform of Google Sites to be blocked within Turkey during the prosecution of another Google Sites webpage containing content allegedly insulting to modern Turkey's founder Mustafa Kemal Ataturk.[63] Even after the case against the offending Google Site had been dropped, Mr. Yildirim's content on Google Sites—as well as all other Google Sites content—remained blocked indefinitely in Turkey.[64] After having his application dismissed by Turkish courts, Mr. Yildirim petitioned the European Court of Human Rights, which unanimously found that Turkey had violated Article 10 of the European Convention on Human Rights in ordering the indefinite blocking of all content available on Google Sites.[65] In his concurring opinion, Judge Paulo Pinto de Albuquerque explained that "blocking orders imposed on sites and platforms . . . [are] pure censorship."[66] Building upon the procedural protections for speech set forth in the Special Rapporteur's Report, Judge Paulo Pinto de Albuquerque set forth minimum procedural protections that any nationwide blocking or filtering of Internet content must embody:

> The *minimum* criteria for Convention-compatible legislation on Internet blocking measures are:
>
> (1) a definition of the categories of persons and institutions liable to have their publications blocked . . . ;
>
> (2) a definition of the categories of blocking orders, such as blocking of entire websites, IP addresses, ports, network protocols or types of use, like social networking;
>
> (3) a provision on the territorial ambit of the blocking order, which may have region-wide, nationwide, or even worldwide effect;
>
> (4) a limit on the duration of the blocking order;
>
> (5) an indication of the "interests" . . . that may justify the blocking order;

62. *Id.* at 2.
63. *Id.* at 22.
64. *Id.* at 3.
65. *Id.* at 20-21.
66. *Id.* at 29.

> (6) observance of the criterion of proportionality, which provides for a fair balancing of freedom of expression and the competing "interests" pursued . . . ;
>
> (7) compliance with the principle of necessity, which enables an assessment to be made as to whether the interference with freedom of expression adequately advances the "interests" pursued and goes no further than is necessary to meet the said "social need";
>
> (8) definition of the authorities competent to issue a reasoned blocking order;
>
> (9) a procedure to be followed for the issuance of that order, which includes the examination by the competent authority of the case file supporting the request for a blocking order and the hearing of evidence from the affected person or institution . . . ;
>
> (10) notification of the blocking order and the grounds for it to the person or institution affected; and
>
> (11) a judicial appeal procedure against the blocking order.[67]

Because none of these procedural protections for speech were provided in the *Yildirim* case, Judge Paulo Pinto de Albuquerque concluded that the Convention's protections for freedom of expression were violated.[68]

B. *United States Procedural Protections for Speech*

Within the context of United States First Amendment jurisprudence, courts have also constructed a powerful "body of procedural law which defines the manner in which they and other bodies must evaluate and resolve [free speech] claims—a [free speech] 'due process,' if you will."[69] In particular, courts have applied stringent procedural safeguards in scrutinizing prior restraints on speech—restraints on speech that are imposed prior to a judicial determination of the speech's illegality.[70] In so doing, courts developed "a comprehensive

---

67. *Id.* at 27-28.

68. *Id.* at 29.

69. *See* Monaghan, *supra* note 58, at 518.

70. This strong presumption against the legality of prior restraints is also shared by the Latin American countries that have ratified the American Convention on Human Rights, which provides that "the right to freedom of thought and expression . . . shall not be subject to prior censorship . . . ." American Convention on Human Rights art. 13, Nov. 22, 1969, 1144 U.N.T.S. 123.

system of 'procedural safeguards designed to obviate the dangers of a censorship system.'"[71]

The prohibition against prior restraints has been a foundational principle of freedom of expression in the Anglo-American tradition.[72] In William Blackstone's *Commentaries*, "Freedom of the Press" is defined simply as the right to be free from prior restraints.[73] In his *Commentaries*, Blackstone explained that "the liberty of the press is indeed essential to the nature of a free state; but this consists in laying no previous restraints upon publications, and not in freedom from censure for criminal matter when published."[74] Indeed, following Blackstone, some have argued that the sole purpose of the First Amendment was to foreclose in the United States any system of prior restraint such as was embodied in the English censorship system.[75]

In order to restrict harmful speech, governments might choose the methods of prior restraints, such as those imposed via the types of filtering systems discussed above, or subsequent punishment, such as by criminally prosecuting content providers who make available harmful speech, for example, under obscenity or child pornography statutes. Regulations that proceed via subsequent punishment provide vastly greater procedural protections for speech and are likely to be implemented in ways that are far more speech-protective than regulations embodying the method of prior restraints. As pre-eminent First Amendment theorist Thomas Emerson explains, systems of prior restraint are likely to operate in a manner that is much more hostile toward free speech than systems of subsequent punishment, for a number of reasons.[76] First, Emerson explains, systems of prior restraint are prone to adverse decisions.[77] Such systems are constructed as to make it easier, and more likely, that the censor will rule adversely to free expression.[78] As Emerson argues, "[an] official thinks longer and harder before deciding to undertake the serious task of subsequent punishment . . . . Under a system of prior restraint, he can reach the result by a simple stroke of the pen. Thus, [in the case of subsequent

71. *See* Monaghan, *supra* note 58, at 518 (quoting Freedman v. Maryland, 380 U.S. 51, 58 (1965)).

72. *See* 4 WILLIAM BLACKSTONE, COMMENTARIES *150, *150-53.

73. *Id.*

74. *Id.*

75. *See, e.g.*, Thomas I. Emerson, *The Doctrine of Prior Restraint*, 20 LAW & CONTEMP. PROBS. 648 (1955).

76. *Id.*

77. *Id.* at 649.

78. *Id.*

punishment], the burden of initial action falls upon the government; in the other, on the citizen . . . . For these and similar reasons, a decision to suppress in advance is usually more readily reached, on the same facts, than a decision to punish after the event."[79]

Second, under a system of prior restraint, the issue whether to suppress expression is determined by an administrative procedure, instead of via criminal procedure.[80] Accordingly, "the procedural protections built around the criminal prosecution—many of which are constitutional guarantees—are not applicable to prior restraint. The presumption of innocence, the heavier burden of proof borne by the government, the stricter rules of evidence, the stronger objection to vagueness, the immeasurably tighter and more technical procedure—all these are not on the side of free expression when its fate is decided."[81]

Third, within a system of prior restraints, the decision to restrict speech rests with a single administrator or functionary instead of with a judge and/or jury.[82] Both judge and jury function to provide important safeguards against abuses of power and are designed to secure independent and objective judgments. Such safeguards are not necessarily present within an administrative system implementing prior restraints.[83]

Fourth, systems of prior restraint are more likely to operate out of the public view and in such a manner that they are hidden from public scrutiny, appraisal, and accountability.[84] In contrast, subsequent punishments take place in the glare of public scrutiny and accountability. As Emerson explains,

> [In systems of prior restraint,] decisions are less likely to be made in the glare of publicity that accompanies a subsequent punishment. The policies and actions of the licensing official do not as often come to public notice; the reasons for his action are less likely to be known or publicly debated; material for study and criticism are less readily available; and the whole apparatus of public scrutiny fails to play the role it normally does under a system of subsequent punishment . . . . [T]he

79. *Id.* at 657.
80. *Id.*
81. *Id.*
82. *Id.* at 657-68.
83. *Id.*
84. *Id.* at 658.

> preservation of civil liberties must rest upon an informed and active public opinion. Any device that draws a cloak over restrictions on free expression seriously undermines the democratic process."[85]

Finally, and perhaps most importantly, the institutional framework in which systems of prior restraint operate are such that they favor suppression of expression. As Emerson argues,

> The function of the censor is to censor . . . . He is often acutely responsive to interests which demand suppression . . . and not so well attuned to the forces that support free expression . . . . The long history of prior restraints reveals over and over again that the personal and institutional forces inherent in the system [of prior restraints] nearly always end in . . . unnecessary, and extreme suppression.[86]

These considerations make clear that the historical opposition to systems of prior restraint are not arbitrary historical accidents, but follow clearly from the importance of according meaningful protections for free expression.

Even though the restraints on speech implemented under a nationwide filtering plan like the IWF's notice and take-down system are not imposed before the speech is made available in the first instance, such restraints still embody the dangers of prior restraints.[87] Prior restraints are typically imposed ex ante, via pre-publication licensing schemes, as occur in the context of motion picture censorship boards.[88] Yet, prior restraints can also be imposed midstream, after initial circulation but sometime before a judicial determination that the speech is illegal has been made, such as with nationwide internet filtering systems.[89] Indeed, Internet filtering systems work from evolving blacklists of websites that are maintained in response to tips or complaints from Internet users.[90] Because midstream prior restraints are imposed prior to a judicial determination of the content's illegality, they are also constitutionally suspect.

---

85. *Id.*
86. *Id.* at 659.
87. *See infra* text accompanying notes 89-97.
88. *See, e.g.*, Freedman v. Maryland, 380 U.S. 51 (1965).
89. *See infra* text accompanying notes 91-97.
90. *See supra* text accompanying notes 27-28.

The U.S. Supreme Court considered and struck down an example of midstream prior restraints in the case of *Bantam Books, Inc. v. Sullivan*.[91] In *Bantam Books*, the Rhode Island Commission to Encourage Morality in Youth was charged with investigating and recommending prosecution of booksellers for the distribution of printed works that were obscene or indecent.[92] The Commission reviewed books and magazines after they were already in circulation, and notified distributors in cases in which a book or magazine had been distributed that the Commission found objectionable.[93] In reviewing the constitutionality of this scheme, the Supreme Court held that, even though the restrictions on publication were imposed after initial circulation and distribution, the Commission's actions nonetheless effectuated an unconstitutional prior restraint.[94] The Court explained that "the separation of legitimate from illegitimate speech calls for . . . . sensitive tools" and reiterated its insistence that regulations of speech "scrupulously embody the most rigorous procedural safeguards."[95] The Court observed that, under the scheme at issue, "the publisher or distributor is not even entitled to *notice and hearing* before his publications are listed by the Commission as objectionable" and that there was "no provision whatever for *judicial superintendence* before notices issue or even for *judicial review* of the Commission's determinations of objectionableness."[96] The Court concluded that, in the context of this system of midstream prior restraint, the "procedures of the Commission are radically deficient" and unconstitutional.[97]

III. The Procedural Dimensions of Current Internet Filtering Systems

The procedural framework set forth in the Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/HRC/17/27, as espoused in the recent European Court of Human Rights decision, and by the United States Supreme Court for assessing the legality of prior restraints provides a helpful starting point for assessing state-imposed filtering of Internet

91. 372 U.S. 58, 73 (1963).
92. *Id.*
93. *Id.*
94. *Id.*
95. *Id.* at 66.
96. *Id.* at 71.
97. *Id.*

content.[98] Translated into the context of nationwide filtering or blocking of Internet speech, these safeguards require, at a minimum (1) that the filtering system provide Internet users with the *opportunity to appeal any such blocking decisions*, to a *judicial body and in an expeditious manner*; (2) that the filtering scheme *operate in an open and transparent manner*, such that affected Internet users and content providers are provided with *notice* that the content was filtered and the reason for such filtering; and (3) that any filtering be imposed subject to *clear and precise definitions of the speech to be regulated*.[99] These procedural requirements are imposed in addition to the general requirements under both the ICCPR and the First Amendment that any content-based restrictions on speech (which Internet filters generally embody) satisfy strict judicial scrutiny and embody the least restrictive means of achieving a compelling government interest.[100] These procedures do not themselves dictate what categories of speech are deemed harmful or dangerous and as such do not interfere with the prerogative of each state to make such substantive determinations. Rather, they impose meaningful, process-based safeguards on the implementation of restrictions of whatever categories of speech are deemed harmful or dangerous by each nation. Such procedures and sensitive tools for protecting free speech are as important as the substantive protections themselves, as "[t]he history of freedom is, in no small part, the history of procedure."[101] In protecting our free speech rights, courts have constructed a powerful body of procedural law which defines the manner in which they and other bodies must evaluate and resolve free speech claims.

A. *Ability to Challenge Decision to Filter Content*

Both the International Covenant on Civil and Political Rights, as construed in the Special Rapporteur's Report, and U.S. prior restraint jurisprudence require that any filtering decision be subject to meaningful challenge before an impartial decision-maker.[102] In construing the ICCPR, the Special Rapporteur explains that:

---

98. *See Special Rapporteur's Report, supra* note 9.

99. *Id.* ¶¶ 47-48.

100. *See, e.g.*, Reno v. ACLU, 521 U.S. 844, 849 (1997) (under First Amendment law); *Special Rapporteur's Report, supra* note 9, ¶ 24 ("[L]imitation[s] to the right [of] freedom of expression . . . must be proven as necessary and the least restrictive means required to achieve the purported aim.")

101. Malinski v. New York, 324 U.S. 401, 414 (1945) (Frankfurter, J., concurring).

102. *See* ICCPR, *supra* note 9; *Special Rapporteur's Report, supra* note 9, ¶ 69.

> [A]ny legislation restricting the right to freedom of expression must be applied . . . with adequate safeguards against abuse, including the possibility of challenge [before an independent body] and remedy against its abusive application.[103]

Building upon this requirement, the European Court of Human Rights in its recent *Yildirim* decision set forth the requirement that any legislation mandating Internet blocking or filtering embody, at a minimum, a judicial appeal procedure.[104]

Similarly, U.S. courts have consistently emphasized the importance of the availability of *prompt judicial review* affording parties notice and an opportunity to be heard regarding censorship determinations in the prior restraint context.[105] As the Supreme Court explained, "because only a judicial determination in an adversary proceeding ensures the necessary sensitivity to freedom of expression, only a procedure requiring a judicial determination suffices to impose a valid final [prior] restraint."[106] In order for a nationwide filtering system to effectuate a legal prior restraint, such a system must provide for an opportunity to secure judicial review of a censorship decision.

Attempts within the U.S. to filter Internet speech at the ISP level have been deemed unconstitutional because they have failed to provide for prompt judicial review of the decision to censor.[107] In the *Center for Democracy and Technology v. Pappert*,[108] for example, the Commonwealth of Pennsylvania sought to combat online child pornography by enacting the Internet Child Pornography Act, which required ISPs serving Pennsylvanians to block access to websites believed to be associated with child pornography. The Act permitted the Pennsylvania Attorney General or Pennsylvania district attorneys to seek an ex parte court order requiring an ISP to remove or disable access to items accessible through the ISP's service, upon a showing of probable cause that the item constituted child pornography.[109] The Act did not require an actual, final judicial determination that the material to be

---

103. *Special Rapporteur's Report*, *supra* note 9, ¶ 24.

104. Yildirim v. Turkey (No. 3111/10), Eur. Ct. H.R. at 27-28 (2012).

105. *See* United States v. Thirty-Seven Photographs, 402 U.S. 363, 372-74; Kingsley Books, Inc. v. Brown, 354 U.S. 436, 440 (1957); Interstate Circuit, Inc. v. City of Dall., 390 U.S. 676, 679 (1968); Bantam Books, Inc. v. Sullivan, 372 U.S. 58 (1963).

106. *See* United States v. Pryba, 502 F.2d 391, 405 (D.C. Cir. 1974).

107. Ctr. for Democracy and Tech. v. Pappert, 337 F. Supp. 2d 606 (E.D. Pa. 2004).

108. *Id.* at 610.

109. *Id.* at 619.

removed actually constituted child pornography before it was placed on the blacklist.[110] In consultation with the affected ISPs, the Attorney General's office decided to implement the Act by proceeding without even securing ex parte court orders.[111] Instead, it provided "Informal Notices of Child Pornography" to ISPs that hosted websites, initially reported by an agent or a citizen, that the Office of the Attorney General found to contain suspected child pornography.[112] The Informal Notice directed the ISP to remove or disable Pennsylvania citizens' access to the suspected material within five days of receipt of the Notice.[113]

The statute was challenged, inter alia, as an unconstitutional prior restraint lacking the requisite procedural safeguards.[114] In defense of the statute, the attorney general explained that only material that its office had probable cause to believe constituted child pornography was requested to be removed.[115] The court found that the probable cause showing did not save the statute (nor did the fact that the attorney general only issued "Informal Notices" not court orders, and that the process was therefore "voluntary" not coercive[116]). First, the court explained that in order to comply with the Supreme Court's exacting procedural requirements for prior restraints, to be constitutional, a prior restraint must be imposed by a judicial determination in an adversary proceeding.[117] The attorney general's determination that there was probable cause that the material was illegal was insufficient.[118] Further, even an ex parte judicial determination that the material was illegal would not suffice to impose a constitutional final prior restraint because it did not result from an adversarial proceeding.[119] As the Supreme Court explained in *Freedman*, "only a judicial determination *in an adversary proceeding* ensures the necessary sensitivity

110. *Id.* at 622.

111. *Id.* at 625.

112. *Id.* at 622.

113. *Id.*

114. *Id.* at 611.

115. *See id.* at 664.

116. On this point, the court explained that the informal and technically non-coercive nature of the attorney general's removal requests did not insulate them from constitutional scrutiny. The court explained that removal requests issued by law enforcement officials were not interpreted by the recipient ISPs as being voluntary, even if technically they did not have the force of law. *Id.* at 660.

117. *Id.* at 656.

118. *Id.* at 657.

119. *Id.*

to freedom of expression."[120] Ex parte judicial determinations that are made in the absence of notice and an opportunity to be heard on the part of the adversely affected speaker are constitutionally deficient, and ex parte *nonjudicial* determinations are constitutionally deficient by an even greater measure.[121]

Under filtering systems implemented in countries like the United Kingdom, provisions do exist for some sort of appeal of the censorship decision.[122] However, such provisions for appeal generally do not provide for *judicial* determination and instead merely provide for a second look by the administrative body that made the censorship determination in the first place.[123] In the United Kingdom, for example, the IWF website indicates that "any party with a legitimate association with the [blacklisted] content . . . who believes they are being prevented from accessing legal content may appeal against the accuracy of an assessment."[124] The appeal procedure provided by the IWF, however, does not contemplate judicial review. Rather, the procedure for appeal involves a second look by the IWF itself and, following that, a review by a police agency whose assessment is final.[125] Further, such appeal procedures do not provide for the opportunity to be heard by the adversely affected parties.[126] As the General Counsel for Wikimedia explained regarding the IWF's review of its decision to blacklist the Scorpion's Wikipedia page, "When we first protested the block, [the IWF's] response was, 'We've now conducted an appeals process on your behalf and you've lost the appeal.' When I asked who exactly represented the Wikimedia Foundation's side in that appeals process,

120. Freedman v. Maryland, 380 U.S. 51, 58 (1965).

121. Because of the impartiality and independence of the judiciary, courts have held that review of censorship decisions by an independent judicial body is a necessary check on a censor's discretion. *See, e.g.*, Freedman v. Maryland, 380 U.S. 51, 58 (1965) (holding that "only a judicial determination [of the decision to censor speech] . . . ensures the necessary sensitivity to freedom of expression."). Any review or appeal of a decision to censor speech that does not allow for such review by an independent judicial body is therefore deficient.

122. *See, e.g.*, INTERNET WATCH FOUNDATION CODE OF PRACTICE, https://www.iwf.org.uk/members/member-policies/funding-council/code-of-practice (last visited Jan. 15, 2013) (setting forth the procedure for appealing IWF's decision to blacklist a website, which procedure does not provide for judicial review); OPENNET INITIATIVE: INTERNET FILTERING IN THE US AND CANADA 2006-2007, https://opennet.net/studies/namerica2007 (last visited Jan. 15, 2013) (explaining that Canada's filtering system does not provide for judicial review of the decision to blacklist a website).

123. *See supra* note 122.

124. *See FAQs Regarding the IWF's Facilitation of the Blocking Initiative*, *supra* note 19.

125. *See Content Assessment Appeal Process*, *supra* note 37.

126. *Id.*

they were silent."[127] The IWF's provisions for appeal—because they do not provide for a judicial determination of the affected parties' rights—fail to accord the requisite protections for freedom of expression.

B. *Meaningful Notice to Affected Internet Users*

The International Covenant on Civil and Political Rights as construed by Special Rapporteur, as well as U.S. prior restraint jurisprudence, require that individuals affected by nationwide filtering systems must at a minimum be made aware of such a decision to filter so that they can effectively challenge such actions.[128] The right to meaningfully challenge a filtering decision, discussed above, presupposes that affected individuals have *notice* of any such censorship so that they can be secured a *meaningful opportunity to challenge* the decision to censor in a judicial forum. As the Special Rapporteur's Report explains:

> [A]ny limitation to the right to freedom of expression must . . . be provided by law, which is clear and accessible to everyone (principles of predictability and transparency) . . . .[129]

Building upon this requirement, the European Court of Human Rights in its recent *Yildirim* decision set forth the requirement that any legislation mandating Internet blocking or filtering provide, at a minimum, notification of the blocking order and the grounds for it to the person or institution affected.[130] Filtering systems that do not require that the affected parties be made aware of filtering decisions—such as that employed by the IWF[131]—fail this threshold requirement.

Countries implementing nationwide filtering systems to restrict their citizens' access to content that they deem harmful should at the very least operate these systems in an open and transparent manner, in which the restrictions on speech are provided by law and are clear and accessible to everyone, to adhere to the principles of predictability and transparency articulated in the ICCPR and in prior restraint jurisprudence. These systems should operate in a manner such that (1) Internet users are made aware of the operation of such filtering

127. Davies, *supra* note 13.

128. *See* ICCPR, *supra* note 9; *Special Rapporteur's Report*, *supra* note 9, ¶ 47.

129. *See Special Rapporteur's Report*, *supra* note 9, ¶ 24.

130. Yildirim v. Turkey (No. 3111/10), Eur. Ct. H.R. at 27-28 (2012).

131. *See supra* text accompanying notes 30-35.

systems generally, and (2) affected users—both content providers and would-be end users—are specifically informed of instances in which the filters operate to block access to a particular website. Only then can affected content providers and end users have the meaningful notice necessary to challenge the decision to censor.

C. *Categories of Prohibited Speech Should Be Clearly and Precisely Defined*

A third procedural requirement for nationwide Internet filtering systems is that the censor's discretion be meaningfully constrained by clearly defined and precise criteria specifying what content is illegal.[132] This requirement serves to cabin and constrain the discretion of the initial censors and require that the censors adhere to the legal determination of what content is proscribable.[133] While countries may reasonably differ in their determinations of what categories of speech are illegal content—pornography, hate speech, Holocaust denial, etc.—it is important that, within each country, the definitions of illegal speech subject to prior restraint be carefully and precisely defined so as to constrain the initial censor's discretion. The Special Rapporteur has articulated the requirement that any restrictions on Internet speech be set forth by laws that are precise and clear, so that Internet users can reasonably predict how the law will apply to their content.[134] Similarly, the Supreme Court has strictly cabined the discretion of censors in systems of prior restraint and has rejected as unconstitutional any systems that reposit unbounded discretion to determine whether or not speech is protected.[135] For example, in *Shuttlesworth v. Birmingham*, the Court evaluated the constitutionality of a parade permitting system that vested the City Commission with the broad discretion to deny parade permits in cases where "in [the Commission's] judgment the public welfare, peace, safety, health, decency, good order, morals or convenience require that [the parade permit] be refused."[136] In ruling on a challenge to the statute, the Court held that, because the permitting scheme constituted a prior restraint on expression that conferred "virtually unbridled and absolute power" on the Commission, it failed to comport with the essential due process requirement that any law subjecting the exercise of First Amendment freedoms to the prior

132. *See, e.g.*, CHEMERINSKY, *supra* note 11, at 966-67.

133. *Id.*

134. *See Special Rapporteur's Report*, *supra* note 9, ¶¶ 47-48.

135. *See* Shuttlesworth v. Birmingham, 394 U.S. 147, 151 (1969).

136. *Id.* at 149-50.

restraint of a license must embody "narrow, objective, and definite standards to guide the licensing authority."[137]

Furthermore, requiring that the criteria by which the censoring authority makes the decision to censor be set forth with precision helps to cabin administrative discretion and also helps to limit "mission creep" within the censoring body. Without a precise and detailed specification of the criteria for censorship, the censor can exercise unbridled and expansive discretion to restrict speech.

The IWF fails to provide meaningful constraints on its decision to blacklist a website. While its website purports to specify the criteria used by the IWF in blacklisting websites, the links to such criteria are broken.[138] Further, the current IWF mandate seems to be susceptible to the problem of mission creep. Although its initial mission was solely to reduce the availability of images of child sexual abuse, the target of its censorship has now been extended beyond child pornography to encompass adult content that is "criminally obscene," and has at times been extended to encompass hate speech and "incitement to racial hatred content."[139] Further, it has yet to be specified precisely what content will fall within the definition of "family-unfriendly" content under Prime Minister Cameron's recently-announced expansion of the United Kingdom's nationwide filtering system. It is unclear how

---

137. *Id.* at 150-51.

138. The IWF website provides the following information:

What are the criterion for a URL to be added to the list?

> The URLs are assessed according to UK law, specifically the Protection of Children Act 1978, and in accordance with the UK Sentencing Guidelines Council. All URLs added to the list depict indecent images of children, advertisements for or links to such content. This content is likely to be an offence to download, distribute, or possess in the UK . . . . The policy by which a decision is made to include a URL on the list can be seen here: http://www.iwf.org.uk/corporate/page.49.626.htm.

However, the policy page referenced above provides a broken link. Accordingly, no information is provided about the policy by which a decision is made to include a URL on the IWF blacklist.

139. *See Hotline*, INTERNET WATCH FOUND., http://web.archive.org/web/20110101125637/https://www.iwf.org.uk/hotline (last visited Dec. 28, 2013) (referencing the January 2011 version of the IWF webpage). The January 2011 version of IWF webpage formerly indicated that an Internet user can report online content that falls within the following categories: (1) child sexual abuse images; (2) nonphotographic child sexual abuse images; (3) criminally obscene adult content; (4) incitement to racial hatred content. *Id.* As of April 2011, incitement to racial hatred content has been removed from IWF's remit. *See Incitement to Racial Hatred Removed from IWF's Remit*, INTERNET WATCH FOUND., https://www.iwf.org.uk/about-iwf/news/post/302-incitement-to-racial-hatred-removed-from-iwfs-remit (last visited Jan. 25, 2014). Responsibility for addressing hate crime content on the Internet is now vested in an organization called True Vision. *See, e.g.*, *Internet Hate Crime/Incident Reporting Form*, TRUE VISION, https://secure.met.police.uk/hatecrime_internet/ (last visited Dec. 28, 2013).

the definitions of "family-friendly" and "family-unfriendly" content will be formulated and who will get to make the decisions regarding what content will be allowed and what content will be blocked. Such categories of family-unfriendly speech—which will apparently include such vague and ill-defined subcategories as "esoteric material"[140]—present the danger of enabling the censors to block whatever speech they choose, without meaningful constraints on their discretion.

IV. Conclusion

In summary, state-mandated Internet filtering systems—the likes of which are now being imposed by over forty countries worldwide, including liberal democracies like the United Kingdom—embody prior restraints on speech, which violate the due process requirements inherent in the free speech guarantee absent the inclusion of fundamental process-based safeguards embodied in both U.S. First Amendment jurisprudence and in the developing free speech jurisprudence under the International Covenant on Civil and Political Rights and the European Convention on Human Rights. These free speech due process requirements mandate that such prior restraints implemented by filtering systems be implemented *in an open and transparent manner*, such that affected Internet users and content providers are provided with information that content was blocked and the reason for such blocking; that any restraints on speech are imposed subject to *clear and precise definitions of the speech to be regulated*; and such that affected Internet users are provided with the *opportunity to appeal any such blocking decisions*, to an *independent judicial body*. Only such "sensitive tools" for distinguishing between protected and unprotected speech can adequately protect individuals' free speech rights.

140. *See supra* text accompanying notes 14-18.