



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/147870/>

Version: Accepted Version

Book Section:

Erickson, K and Kretschmer, M (2020) Empirical Approaches to Intermediary Liability. In: Frosio, G, (ed.) The Oxford Handbook of Online Intermediary Liability. Oxford University Press, pp. 104-121. ISBN: 9780198837138.

<https://doi.org/10.1093/oxfordhb/9780198837138.013.5>

© Giancarlo Frosio 2020. This is an author produced version of a book chapter published in The Oxford Handbook of Online Intermediary Liability. Uploaded in accordance with the publisher's self-archiving policy.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Chapter 5

Empirical Approaches to Intermediary Liability

Kris Erickson and Martin Kretschmer

Legal theory has failed to offer a convincing framework for the analysis of the responsibilities of online intermediaries. Our co-contributors to this volume have identified a wide range of contested issues, from due process to costs to extra-territorial matters. This chapter considers what empirical evidence may contribute to these debates. What do we need to know in order to frame the liability of intermediaries and, *a fortiori*, what does the relationship between theory and empirics imply for the wider issue of platform regulation?

While the liability of online intermediaries first surfaced as a technical issue with the emergence of Internet Service Providers (ISPs) during the 1990s, recently, the focus has shifted to the dominance of a handful of global internet giants that structure everything we do online. Platform regulation has become the central policy focus. There is an awareness of the increasing importance of communication between users in constituting a digital public sphere, and simultaneously greater pressure to control online harm (be it relating to child protection, security or fake news).

The allocation of liability in this context becomes a key policy tool. But we know very little about the effects of this allocation for the different stakeholders. This is in part due to secrecy about the rules under which platforms operate internally. As online users, we are governed by processes and algorithms that are hidden from us. If the rules are the result of algorithmic or AI decision-making, even service providers themselves may not fully understand why decisions are made. Still, platforms see their rules of decision-making as key to their competitive advantage as firms. We live in what has been catchily labelled a ‘black box society’.¹

So how can we as researchers open the box to let in some empirical air? There is a global trend towards increased reporting requirements for platforms. Most prominently, the German platform law of 2017 *NetzDG* requires all for-profit social media services with at least two million registered users in Germany to remove obviously illegal content within 24 hours of notification.² There are obligations to

¹ See Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard U Press 2015). See also *infra* Chapter 35.

² See 2017 The Network Enforcement Act (*Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken*) (*NetzDG*) (DE), s 3(2)(2). The obligation to remove and block content relates to a specific list of criminal offences (not including intellectual property right infringements), and

report every six months and high sanctions for failures to comply (up to €5m fines that can be multiplied by 10). During the first six months of the law's operation (January-June 2018), Facebook received a total of 1,704 takedown requests and removed 362 posts (21.2%), Google (YouTube) received 214,827 takedown requests and removed 58,297 posts (27.1%), Twitter received 264,818 requests and removed 28,645 (10.8%). These are much lower numbers than were expected.³

There are also a number of antitrust inquiries that have extracted sensitive data from platforms, in particular the European Commission's investigations of Google under EU competition law.⁴ The Australian Competition and Consumer Commission's digital platforms inquiry is considering the establishment of a new platform regulator with wide-ranging information gathering and investigative powers. The regulator would monitor platforms with revenues from digital advertising in Australia of more than AU\$100 million. These firms would be subject to regular reporting requirements with the aim to control whether they are engaging in discriminatory conduct, for example by predatory acquisition of potential competitors or advertising practices that favour their own vertically integrated businesses.⁵

So in the future, there is likely to be much greater public scrutiny and knowledge-gathering about platforms' practices. The paradox is that we will know more only by the time new regulatory regimes, and a new framework for online intermediary liability, have been selected. We appear to be in the midst of a paradigm shift in intermediary liability, moving from an obligation to act once knowledge is obtained to an obligation to prevent harmful content appearing. This imminent shift towards filtering, even general monitoring⁶, is exemplified by the controversial Article 17 (formerly Article 13) of the 2019 EU Directive on Copyright in the Digital Single

there is a carve out, exempting platforms that support communications between individuals (this is designed to exempt professional networks, sales platforms, games and messaging services).

³ See William Echikson and Olivia Knodt 'Germany's NetzDG: A Key Test for Combatting Online Hate' (2018) Centre for European Policy Studies (CEPS) Report no 2018/09 <https://www.ceps.eu/system/files/RR%20No2018-09_Germany%27s%20NetzDG.pdf>.

⁴ See European Commission 'Press Release: Commission fines Google €1.49 billion for abusive practices in online advertising' (Case 40411 Google Search (AdSense)) (20 March 2019) <http://europa.eu/rapid/press-release_IP-19-1770_en.htm>; European Commission 'Press Release: Commission fines Google €4.34 billion for illegal practices regarding Android mobile devices to strengthen dominance of Google's search engine' (Case 40099 Google Android) (18 July 2018) <http://europa.eu/rapid/press-release_IP-18-4581_en.htm>; European Commission 'Press Release: Commission fines Google €2.42 billion for abusing dominance as search engine by giving illegal advantage to own comparison shopping service' (Case 39740 Google Search (Shopping)) (27 June 2017) <http://europa.eu/rapid/press-release_IP-17-1784_en.htm>.

⁵ See Australian Competition and Consumer Commission (ACCC), Digital platforms inquiry (intermediary report, published 10 December 2018) <<https://www.accc.gov.au/focus-areas/inquiries/digital-platforms-inquiry/preliminary-report>>. An independent regulator also has been proposed in the UK to oversee a new statutory 'duty of care' for platforms. See Department for Digital, Culture, Media & Sport and Home Department, *Online Harm* (White Paper, Cp 59, 2019) <<https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper-executive-summary--2>>.

⁶ Expressly prohibited as an obligation on service providers by Art. 15 of the EU E-Commerce Directive and the case law of the Court of Justice of the European Union. See also Chapter 28.

Market that makes certain platforms ('online content sharing services') directly liable under copyright law for the content uploaded by their users.⁷

This chapter evaluates what we already know after almost two decades of operation of one liability regime: the so-called safe harbour introduced in the United States by the Digital Millennium Copyright Act (DMCA), and in a related form in European Union Member States with the E-Commerce Directive.⁸ Immunity for 'Online Service Providers' that act expeditiously to remove infringing material was first introduced in the United States under Section 512 of the U.S. Copyright Act (as amended by the DMCA 1998). Section 512 specifies a formal procedure under which service providers need to respond to requests from copyright owners to remove material. Rightholders who wish to have content removed must provide information 'reasonably sufficient to permit the service provider to locate the material' (such as a URL) and warrant that the notifying party is authorized to act on behalf of the owner of an exclusive right that is allegedly infringed. The practice is known as 'notice-and-takedown'. Importantly, 'counter notice' procedures are also specified under which alleged infringers are notified that material has been removed and can request reinstatement.

Under the EU Directive on Electronic Commerce (2000/31/EC), hosts of content uploaded by users will be liable only upon obtaining knowledge of the content and its illegality. The safe harbour of the E-Commerce Directive applies to all kinds of illegal activity or information, not only copyright materials. But unlike the DMCA, the E-Commerce Directive does not regulate the procedure for receiving the necessary knowledge. This is left to the Member States. The regime is sometimes characterized as 'notice-and-action'.⁹

Even though the notice-and-takedown regime established under the DMCA is narrow (applying to copyright only) and limited to one jurisdiction, it has become the dominant paradigm for organizing liability of online intermediaries. This is for at least two reasons: (1) through Google's practices, notice-and takedown has become a global standard even in jurisdictions that do not have safe harbour laws;¹⁰

⁷ See Directive 2019/790/EU of the European Parliament and the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC [2019] OJ L 130/92, Art 17.

⁸ See Digital Millennium Copyright Act, Pub L no 105-304, 112 Stat 2860; Directive 2000/31/EC of the European Parliament and the Council of 17 July 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market [2000] OJ L178/1.

⁹ Cf Chapter 27.

¹⁰ According to Google's 2019 Transparency Report, in total more than 4bn copyright takedown requests have been received. They are processed under DMCA formalities, regardless of whether the country in which the request was filed prescribed these formalities or had any safe harbor laws: 'It is our policy to respond to clear and specific notices of alleged copyright infringement. The form of notice we specify in our web form is consistent with the Digital Millennium Copyright Act (DMCA) and provides a simple and efficient mechanism for copyright owners from countries around the world. To initiate the process to remove content from Search results, a copyright owner who believes a URL points to infringing content sends us a take-down notice for that allegedly infringing material. When we receive a valid take-down notice, our teams carefully review it for completeness and check for other problems. If the notice is complete and we find no other issues, we remove the URL from Search results.' See Transparency Report Google <<https://transparencyreport.google.com/copyright/overview>>.

(2) copyright liability affects a far wider range of user practices than for example defamation, obscenity and other forms of illegal use. So, the lessons from the operation of the notice-and-takedown regime may have wider application, addressing questions of transparency, due process, cost allocation and freedom of expression.

We now proceed to review the body of empirical studies on copyright intermediary liability during the 20-year period from 1998 (the year that DMCA was passed) to 2018. We use a snowball sampling method enhanced by a seed sample of empirical papers drawn from the *Copyright Evidence Wiki*, an open-access repository of findings related to copyright’s effects.¹¹ Based on the initial sample, we searched forwards and backwards in time among those articles to identify further published research. The sample is focused exclusively on work deemed empirical (containing new data gathered or constructed by the authors). It includes articles, books, reports and published impact assessments.

Based on our survey of this body of research, we identify five key sub-fields of empirical inquiry pursued so far. These relate to: the volume of takedown requests, the accuracy of notices, the potential for over-enforcement or abuse, transparency of the takedown process, and the costs of enforcement borne by different parties (see Table 1). Each of these areas are discussed in further detail in the remainder of the chapter. We conclude by identifying some of the gaps and limitations in this existing body of scholarship on intermediary liability for copyright, and offer some recommendations for future research.

Table 1. Thematic foci of empirical research on notice-and-takedown and key studies

Policy Issue	Key studies
Volume of use of notice-and-takedown	Urban and Quilter (2006); Seng (2014); Karaganis and Urban (2015); Cotropia and Gibson (2016); Strzelecki (2019)
Accuracy of notices	Urban and Quilter (2006); Seng (2015); Bar-Ziv and Elkin-Koren (2017); Urban, Karaganis and Schofield (2017)

¹¹ See The Copyright Evidence Wiki: Empirical Evidence for Copyright Policy, CREATE Centre: University of Glasgow <http://CopyrightEvidence.org>.

Over-enforcement / abuse	Ahlert and others (2004); Nas (2004); Urban, Karaganis and Schofield (2017); Bar-Ziv and Elkin-Koren (2017); Erickson and Kretschmer (2018); Jacques and others (2018)
Transparency / due process	Seng (2014); Perel and Elkin-Koren (2017); Fiala and Husovec (2018)
Balancing innovation / costs	Heald (2014); Schofield and Urban (2016); Urban, Karaganis and Schofield (2017)

1. Volume of Notices

In their 2006 study of notice-and-takedown, Urban and Quilter found that Google had received 734 notices between March 2002 (when the Chilling Effects¹² database started collecting Google reports) and August 2005, the cut-off date of their study.¹³ The majority of takedown requests in their sample related to search engine links, but the overall quantity was relatively small. Since then, a number of follow-up studies have noted an explosion in the quantity of DMCA Section 512 notices sent by rightsholders to online service providers, including Google. This sharp increase in volume did not occur immediately upon introduction of the DMCA and e-Commerce Directive; rather, this uptake took nearly a decade and accelerated from 2010 onwards.¹⁴ There are a range of explanations for this increase, ranging from rightsholder frustration following rejection of the Stop Online Piracy Act (SOPA),¹⁵ to new technical affordances introduced by Google to receive and process requests. A general observation from this research is that

¹² The Chilling Effects (now Lumen) database was founded by Wendy Seltzer in 2001 and is currently maintained by the Berkman Klein Center for Internet and Society at Harvard University. The project website <www.lumendatabase.org> collects and enables analysis of takedown requests received by online intermediaries.

¹³ See Jennifer Urban and Laura Quilter, 'Efficient Process or 'Chilling Effects'? Takedown Notices Under Section 512 of the Digital Millennium Copyright Act' (2006) 22 Santa Clara Computer and High Tech L J 621.

¹⁴ See Daniel Seng 'The state of the discordant union: An empirical analysis of DMCA takedown notices' (2014) 18 Va J L & Tech 369.

¹⁵ See Daniel Seng, 'Who Watches the Watchmen?' An Empirical Analysis of Errors in DMCA Takedown Notices' (2015) SSRN Research Paper no 25632023, 3. <<https://ssrn.com/abstract=25632023>>.

notice-and-takedown processes can, at the very least, be considered successful in terms of uptake and use by rightsholders and Online Service Providers (OSPs), with the safe harbour it provides to internet intermediaries viewed as important for commercial innovation.¹⁶ However, the volume of notices may also produce challenges for OSPs in terms of cost of compliance, with implications for other thematic areas such as due process. Several empirical studies have examined these issues in significant detail.¹⁷

A 2014 study by Seng used data obtained from the Chilling Effects repository and from Google's Transparency report, on takedown notices received by Google and other services which report to the database. The dataset consisted of 501,286 notices and some 56,991,045 individual takedown requests.¹⁸ The dataset covered the period from 2001 – 2012 including the period of significant increase in volume. Seng notes that the increase not only affected Google, but other OSPs as well, lending support to his view that legislation was key to explaining the increase. For example, twitter submitted less than 500 notices to the repository in 2010, but reported more than 4,000 requests the following year in 2011, and more than 6,000 in 2012.¹⁹

Seng found that a large majority of notices were sent by industry associations, collecting societies, and third-party enforcement agencies. The British Phonographic Industry (BPI) was the top issuer of notices in the study, with 191,790 notices sent, or 38% of the total sample. Agents such as Degban and WebSheriff also made up a significant portion of the total volume of notices sent.²⁰ The music industry accounted for the largest share of total notices sent, by sector, with nearly 60% of the total notices.

Seng observed an increasing concentration among the top issuers of notices over time. The top-50 issuers accounted for 23.9% of notices in 2010, but reached 74.7% of notices in the dataset for the year 2012.²¹ In other words, a smaller number of organisations, such as BPI and RIAA generated the bulk of notices, skewing the average. Most providers identified in the dataset issued only one notice. Industry organisations and enforcement agents also crammed more claims and requests (sometimes numbering in the thousands) into each notice, while individual claimants tended to include fewer claims and requests together in the same notice.

Other studies have examined the volume of takedown notices received by non-commercial institutions such as universities and academic libraries. Both groups receive fewer takedown requests than commercial internet companies, but

¹⁶ See Jennifer Urban, Joe Karaganis and Brianna Schofield, 'Notice and Takedown: Online Service Provider and Rightsholder Accounts of Everyday Practice' (2017) 64 J Copyright Soc'y 371.

¹⁷ Recently see Artur Strzelecki, 'Website removal from search engines due to copyright violation' (2019) 71(1) *Aslib J of Information Management* 54-71.

¹⁸ See Seng (n 14) 382.

¹⁹ *ibid* 444.

²⁰ *ibid* 448.

²¹ *ibid* 393.

empirical studies show changing practices over time and concern about possible volume increases in future.

Cotropia and Gibson surveyed from a population of 1,377 four-year colleges and universities in the USA.²² They sent the survey to 680 institutions that had a DMCA agent registered with the US Copyright Office. The presence of registered DMCA agents was skewed towards large, higher-ranked institutions. Among the 532 institutions that had working contact information and received the survey, the authors achieved a response rate of approximately 15% (or 80 responses).²³ The average number of takedown requests received per institution was 200 (std. deviation 329.35). The majority of institutions surveyed (72.5%) received between 0 and 270 notices, with a smaller number of schools receiving larger quantities of notices. Some 12.5% of responding institutions spent more than 500 hours per year dealing with takedown requests, but most institutions (62%) reported spending 50 hours or less.²⁴ The majority of DMCA notices received (67.6%) related to cases in which students used institutions' networks to download copyright infringing material onto personal computers (falling under the transitory communications safe harbour under Section 512(a)).²⁵

Universities were found to be using a number of technical measures to deter copyright infringing behaviour by users, in a similar manner to the DMCA-plus OSPs discussed by Urban and co-authors.²⁶ Techniques used by universities included requiring individual network logins, port banning / firewalls, packet shaping, bandwidth throttling and monitoring of suspicious network traffic. Some institutions reported that these techniques had reduced the number of DMCA notices received.²⁷

The findings were limited by the low response rate and potential for response bias this introduced. Since universities may be wary of reputational damage or increased liability for infringing behaviour, there are disincentives to disclose the number of DMCA requests received or to discuss copyright infringing behaviours in general. However, the data offer original and unique insight on the sophistication of legal representatives within higher education with regards to DMCA provisions, and the specific techniques institutions have developed to confront these challenges.

2. Accuracy of Notices

One important focus for empirical investigation has been the accuracy of notices received by service providers. Here, research is concerned with the incentives structure for various parties in the notice-and-takedown regime, and its

²² See Christopher Cotropia and James Gibson, 'Commentary to the U.S. Copyright Office Regarding the Section 512 Study: Higher Education and the DMCA Safe Harbors' (2016) SSRN Research Paper np 2846107 <<https://ssrn.com/abstract=2846107>>.

²³ *ibid* 4.

²⁴ *ibid* 12.

²⁵ *ibid* 13.

²⁶ See Urban, Karaganis and Schofield (n 16) 382.

²⁷ See Cotropia and Gibson (n 22) 14.

effectiveness in identifying and removing actually infringing content. Since the volume of takedown requests being sent has increased substantially, the problem of accuracy is potentially amplified. Directly studying accuracy of notices is challenging without access to the notices themselves and ideally, the targeted work. The Chilling Effects (now Lumen) database has been used extensively as a source of data for empirical research on accuracy.²⁸

One finding consistent across studies is that a small number of issuers are responsible for a disproportionate amount of takedown requests received by platforms, with consequences for accuracy of notices. Bar-Ziv and Elkin-Koren found that 65% of their sample of takedown notices were sent by a single entity,²⁹ while a study by Urban, Karaganis and Schofield identified a single individual responsible for 52% of the takedown notices in their separately collected sample.³⁰ The concentration of issuers appears related to both the ease of filing using automated web tools, as well as the emergence of third-party services which aggregate and work on behalf of rightsholders. In fact, only 1% of the copyright notices analysed by Bar-Ziv and Elkin-Koren were filed by private individuals, while 82% of the requests were filed by third-party enforcement services.³¹ The authors note this has a potentially negative side effect: increasing the number of steps between actual rightsholders and recipients of takedown notices may contribute to a greater number of inaccuracies observed in bulk requests sent by third-party agencies. As Seng put it, '[i]f the price of each arrow is low or minimal, to improve his chances, the reporter will fire off as many arrows as he could to hit a target, regardless of the accuracy or precision.'³²

In his 2015 study of notice accuracy, Seng observed some improvements in accuracy rates since earlier studies, but worrying issues related to the substantive content of claims remained. He analysed 501,286 notices for the presence or absence of formalities such as rightsholder signature, statement of good faith, statement of accuracy and statement of authorization to act on behalf of a copyright owner. Seng found an extremely low quantity of errors among the notices analysed, decreasing over time: the error rate for formalities measured in 2012 was less than 0.1%.³³ This can be explained by Google's adoption of a structured web form to intake notices, which requires that senders complete fields and provides instructions on how to complete them before sending.

Seng then analysed substantive errors – those related to the nature of the copyright claim itself. In order to do this, he collected notices where the name of the copyright owner had not been redacted in the Chilling Effects dataset, to evaluate whether the claim was legitimate. He identified three specific cases in which an employee of a company (rather than the copyright owner herself) was erroneously identified in the request. Since individual notices can contain

²⁸ See Lumen (n 12).

²⁹ See Sharon Bar-Ziv and Niva Elkin-Koren, 'Behind the Scenes of Online Copyright Enforcement: Empirical Evidence on Notice & Takedown' (2018) 50 Connecticut L Rev 1.

³⁰ See Urban, Karaganis and Schofield (n 16) 99.

³¹ See Bar-Ziv and Elkin-Koren (n 29) 26.

³² Seng (n 15) 48.

³³ *ibid* 19.

thousands of requests, these three systematic errors amounted to a total of 380,379 takedown requests, of which Google complied with over 90%.³⁴ As Seng elaborates, ‘what is alarming is the *magnitude, frequency* and *systematic nature* of these errors, which remained undetected and uncorrected for months on end. While we may excuse these errors on the basis that they arose from programs that are misconfigured with wrong information, automated systems propagated these errors across hundreds and thousands of takedown requests.’³⁵ Seng reports that his findings represent a lower bound in the error rate, since he was unable to test accuracy in other ways, such as by observing removed content directly. Direct observation is rendered difficult by the swiftness with which requests are processed and content taken down.

Overall, it appears on one hand that accuracy has been improved via automated systems, such as Google’s preferential ‘Trusted Copyright Removal Program’, by providing structured web forms, clear instructions and negative consequences (revoked membership in the program) for submitting inaccurate notices. On the other hand, ‘Robo-notices’³⁶ which may be generated in large numbers by third party enforcement agencies not closely tied to rightsholders, can introduce and amplify errors affecting significant quantities of works.

3. Over-enforcement and Abuse

Over-enforcement occurs when non-infringing material is removed, for example because the content has been erroneously identified (e.g. a false positive), or because either the sender or receiver of a notice have not sufficiently considered exceptions such as fair use. Deporter and Walker argue that over-enforcement can be caused by a range of factors, including uncertainties in copyright law, the automation of enforcement, the frequent presence of both infringing and non-infringing uses on the same platform, and the high legal costs of defending one’s right to use copyright material.³⁷

In the first major qualitative study of notice-and-takedown, Urban, Karaganis and Schofield assessed the potential for over-enforcement by interviewing 29 Online Service Providers (OSPs) and 6 senders of high volumes of takedown notices. Respondents included ‘video and music hosting services, search providers, file storage services, e-commerce sites, web hosts, connectivity services, and other user-generated content sites’.³⁸ Data were anonymized before publication. Overall, the authors found that notice-and-takedown procedures were taken seriously by both OSPs and rightsholders. OSP respondents stated that the safe harbour provisions in DMCA were central to their ability to operate, providing a stable framework for managing liability. On the other hand, some OSPs claimed that the

³⁴ *ibid* 36

³⁵ *ibid* 37.

³⁶ Joe Karaganis and Jennifer Urban, ‘The Rise of the RoboNotice’ (2015) 58 *Comm ACM* 28, 28–30.

³⁷ Ben Depoorter and Robert Kirk Walker, ‘Copyright False Positives’ (2013) 89 *Notre Dame L Rev* 319.

³⁸ Urban, Karaganis and Schofield (n 16) 376

fear of liability might lead them to over-enforce, as they struggled to decide what to do about inaccurate or invalid takedown requests.

The authors differentiated OSPs into groups depending on the nature of the takedown notices they received. The group termed ‘DMCA Classic’ received individual takedown requests from single rightsholders and dealt with them using human review.³⁹ However, some service providers received massive amounts of takedown notices (more than 10,000 per year) often sent by electronic means, which the authors deemed ‘DMCA Auto’.⁴⁰ These OSPs developed computerized systems for dealing with the large volume of requests, and made these tools available to rightsholders. These OSPs were not always able to engage in human review of bulk requests. A final group, deemed ‘DMCA Plus’ took further steps to limit the upload of content that might trigger notices, at the same time as it offered more direct automated tools to rightsholders to manage the detection and removal of potentially infringing content.⁴¹ Like the second group, these providers were unable to conduct human review of all requests, leading to issues of transparency, accuracy and over-enforcement.

In addition to the concerns raised by Urban, Karaganis and Schofield, another area of focus for research has been protection of legitimate re-use of material, such as provided by fair use in the USA, and by specific copyright exceptions in the UK and Europe. Due to the high volume of requests observed by researchers, and the seemingly strong incentives for platforms to over-comply with requests, there is potential that limitations and exceptions to copyright could be eroded in this system.

In one study of copyright exceptions, Erickson and Kretschmer longitudinally examined a dataset of user-generated parody videos hosted on YouTube, recording at yearly intervals whether videos had been taken down⁴². The research was conducted in collaboration with Jacques and others, who observed additional time periods. In all, the research covered a dataset of 1,839 videos between 2011 and 2016. The overall takedown rate across the whole 4-year period was 40.8% of videos, with 32.9% of all takedowns attributable to copyright requests.⁴³ Parodies varied in terms of uploader skill, parodic intent, and the nature of material borrowed to make the parody. Underlying musical works varied in terms of genre, territory, size of publisher and commercial success. A hazards model (a statistical technique relating survival over time to one or more covariates) was used to analyse the effect of these variables on the likelihood that a given parody would be taken down. The findings showed that parodist technical skill and

³⁹ *ibid* 379.

⁴⁰ *ibid* 382.

⁴¹ For additional discussion of ‘DMCA-Plus’ behaviours, see Annemarie Bridy, ‘Copyright’s Digital Deputies: DMCA-plus Enforcement by Internet Intermediaries’ in John Rothchild (ed), *Research Handbook on Electronic Commerce Law* (Edward Elgar 2016).

⁴² See Kris Erickson and Martin Kretschmer ‘This Video is Unavailable: Analyzing Copyright Takedown of User-Generated Content on YouTube’ (2018) 9 JIPITE 75.

⁴³ *ibid* 83.

production values were significant in reducing the odds of a takedown. More popular parodies with more views also had lower odds of being removed.⁴⁴

Rightsholder behaviour varied significantly by music genre: rock music rightsholders were significantly more tolerant of parodies than hip hop and pop music rightsholders.⁴⁵ The significance of borrowed sound recordings in predicting a takedown, while controlling for other aspects, suggests that algorithmic detection techniques like ContentID are shaping rightsholder behaviour. The availability of a parody exception to copyright in the UK did not appear to deter rightsholders from issuing takedown requests. Artists originating from the USA were significantly more tolerant of parodies than their UK counterparts.

In a related study, Jacques and others found that the diversity of content is potentially harmed by automated takedown. Using the same dataset of 1,839 parodies, the authors checked to see whether music video parodies had been manually removed or blocked via the ContentID automated system. Differences in the way that YouTube informs would-be viewers of these missing videos allowed the researchers to distinguish between regular takedowns and ContentID blocking.⁴⁶ The researchers attributed 32.1% of the takedowns measured in 2016 to algorithmic takedown, and 6.4% to manual takedown.⁴⁷

The authors then examined the effect of takedown on cultural diversity. They differentiated between ‘supplied’ diversity, which they define as diversity in the array of messages that could be watched on the platform, and ‘consumed’ diversity, which is measured in terms of what viewers actually choose to watch.⁴⁸ They used the Simpson Index of diversity to calculate differences in the ‘effective number of parties’, that is, the concentration of availability of certain expressions, before and after takedowns are detected. The authors find that within the sample there is already a strong difference between ‘supplied’ and ‘consumed’ diversity, related to skewness in the preference by viewers for certain pop songs and specific parodies. This finding mirrors other research on the concentration (‘bottlenecking’) of content consumption online.⁴⁹ Interestingly, the authors find that the application of automated takedown also reduces the effective number of parties in the sample, resulting in an overall loss of diversity of content. Their index of consumed diversity contracted from 27 in 2013 to 20 in 2016 after takedowns had occurred.⁵⁰ However, the effect of automated content blocking on diversity is overwhelmed by the built-in lack of diversity in demand (which occurs due to algorithmic sorting and the limited number of search results offered by YouTube). As the authors put it, ‘if [ContentID] removes the most popular parody, would the next most popular simply replace it, almost [...] as a forest fire which, in taking out the old trees,

⁴⁴ *ibid* 86.

⁴⁵ *ibid* 87.

⁴⁶ See Jacques and others, ‘The Impact on Cultural Diversity of Automated Anti-Piracy Systems as Copyright Enforcement Mechanisms: An Empirical Study of Youtube’s Content ID Digital Fingerprinting Technology’ (2018) 15(2) SCRIPTed 277, 291.

⁴⁷ *ibid* 298.

⁴⁸ *ibid* 292.

⁴⁹ See Matthew Hindman, *The Myth of Digital Democracy* (Princeton U Press 2008).

⁵⁰ See Jacques and others (n 46) 299.

gives breathing space for new growth? If that is the case, then while the take-down [...] may be a personal tragedy for the creator of the most popular parody, this will have little to no effect on diversity or possibly even welfare.⁵¹

Beyond studies of unanticipated effects on freedom of expression, another troubling possibility is that malicious actors could use the copyright claims to remove content they find politically disagreeable, or for other arbitrary reasons unrelated to copyright. In one pair of studies, researchers tested the level of scrutiny that ISPs gave to takedown notices before acting. These studies each created simulated web pages containing non-infringing content, and then sent notices to ISPs asking for it to be removed.⁵² In both cases, the authors used work clearly in the public domain (published pre-20th Century), minimizing doubt that the content could be infringing. The study by Ahlert and others used portions of a chapter of John Stuart Mill's 'On Liberty', while the experiment by Sjoera Nas used a work by Eduard Douwes Dekker dating from 1871. The researchers sent takedown notices, purporting to originate from the non-existent 'John Stuart Mill Heritage Foundation' and the 'E.D. Dekkers Society', from anonymous email addresses.

In the two cases tested by Ahlert and others (one UK and one American ISP), the UK web host acted immediately to block the test webpage, while the American ISP appeared willing to take action, but asked the researchers for further information before removing the content. Specifically, the US ISP asked the researchers to provide a statement 'that the complaining party has a good faith belief that use of the material in the manner complained of is not authorized by the copyright owner' as well as a statement that the information contained in the takedown notice was accurate.⁵³ The researchers decided not to pursue the experiment at that point. In the Dutch example, Nas carried out a similar experiment using webpages created on 10 Dutch ISPs. Of those, 7 removed or blocked the webpage containing public domain material by Dekker, sometimes before notifying the owner of the website. A further 2 ISPs ignored the requests, while 1 ISP refused to take the material down because it correctly determined it to be in the public domain.

These two experimental studies, while illustrative of the troubling fact that it 'takes only a Hotmail account to bring a website down',⁵⁴ are limited in certain important respects. Employees of the targeted ISPs, having various levels of subject-area expertise, could not be expected to consistently identify a literary text in the public domain, even a well-known one. Because they targeted a limited number of ISPs using a controlled scenario, these studies were unable to provide data on the actual levels of abuse of notice-and-takedown mechanisms, which would considerably extend the usefulness of this research. However, both studies

⁵¹ *ibid* 303.

⁵² See Christian Ahlert, Chris Marsden, and Chester Yung, 'How 'Liberty' Disappeared from Cyberspace: the Mystery Shopper Tests Internet Content Self-regulation' (2004) <<http://pcmlp.socleg.ox.ac.uk/wp-content/uploads/2014/12/liberty.pdf>>; Sjoera Nas 'The Multatuli Project: ISP Notice and Takedown' (2004) <<https://www-old.bof.nl/docs/researchpaperSANE.pdf>>.

⁵³ Ahlert, Marsden and Yung (n 52) 21.

⁵⁴ Nas (n 52) 6.

suggest that placing legal requirements on notice issuers (such as statements of ‘good faith’) may deter abusive behaviour.

4. Due Process and Transparency

Procedural justice, in particular relating to the interests of users, is an issue that surfaces in a number of contributions to this volume.⁵⁵ This also has been investigated empirically in relation to notice-and-takedown.

In a recent study, Fiala and Husovec studied the economic incentives that drive over-removal of content and under-use of the counter-notification mechanism among users.⁵⁶ The main problem identified by the authors is that, ‘according to theory and empirical evidence, [notice-and-takedown] leads to many false positives due to over-notification by concerned parties, over-compliance by providers, and under-assertion of rights by affected content creators.’⁵⁷ To investigate the causes of under-use of counter-notification, the authors designed a laboratory experiment to model the relationship between service providers (platforms) and content creators. In the experiment, players were given the task of evaluating whether a maze had a valid solution or not, in a 15-second time limit. This was intended to simulate the real-world decision by platform providers about whether to comply with a takedown request. Creators were given more time to study the maze, simulating their familiarity with their own work. A subsequent round allowed the opportunity to ‘punish’ service providers for incorrect removal of content. A stimulus condition simulated a hypothetical dispute resolution process in which creators were given more power and financial incentive to dispute incorrect decisions by providers. The authors ran the experiment with 80 subjects drawn from university students in the Netherlands, and used real payouts.

The researchers found that unlike the baseline condition in which providers tended to over-enforce, and creators tended not to dispute decisions, an alternative dispute resolution (ADR) treatment resulted in fewer mistakes by providers and a more profitable condition for creators overall.⁵⁸ After repeated iterations of the game, the researchers observed that the existence of a credible mechanism through which mistakes can be identified and corrected, represented by the ADR process, decreased the rate of incorrect takedowns from 35% to 19%.⁵⁹

Another problem for transparency in the notice-and-takedown process is the presence of automated or algorithmic methods of identification and removal. Algorithms are not subject to public oversight, often complex, walled off behind commercial secrecy, and unpredictable as they adapt to changing conditions over time. As Perel and Elkin-Koren write, ‘proper accountability mechanisms are vital for policymakers, legislators, courts and the general public to check algorithmic

⁵⁵ See in particular Chapter 34, 35.

⁵⁶ See Lenka Fiala and Martin Husovec, ‘Using Experimental Evidence to Design Optimal Notice and Takedown Process’ (2018) TILEC Discussion Paper no 2018-028 <<https://ssrn.com/abstract=3218286>>.

⁵⁷ *ibid* 1.

⁵⁸ *ibid* 15.

⁵⁹ *ibid* 17.

enforcement. Yet algorithmic enforcement largely remains a black box. It is unknown what decisions are made, how they are made, and what specific data and principles shape them.’⁶⁰

Perel and Elkin-Koren advocate ‘black box tinkering’ as a method to uncover the hidden functionality of algorithms and hold them accountable. In the context of the intermediary liability regime this means conducting experiments on live platforms under conditions controlled by the researcher to test how algorithms such as YouTube’s ContentID system react to various inputs. The authors did this by gathering data on the behaviour of online service providers over the lifecycle of a typical uploaded work of user-generated content: from filtering at the moment of upload, to receipt and handling of takedown notices, to the removal of content and notification of the uploader. To accomplish this, they uploaded and tracked various purpose-made clips with paired controls. For example, one clip contained non-infringing footage, but copyrighted music, while another contained only the footage. The researchers obtained ethical approval for the study from their university ethics committee, and notified the platforms at the conclusion of the study that they had conducted an experiment. They found that 25% of video sharing websites and 10% of the image sharing websites tested in Israel made use of some ex-ante filtering technology at the point of upload.⁶¹ 50% of the video sharing websites removed infringing content upon receipt of a notice, while only 12.5% of the image sharing websites did so. After removing content, all of the video sharing websites tested did notify the uploader about the removal, while only 11% of the image websites did so.⁶²

The wide variation in practices between online platforms suggest problems with procedural justice in the Israeli setting observed by the researchers. They also noted the presence of false positives (removal of non-infringing content when asked to do so) as evidence of the failure of human/algorithmic systems of handling notice-and-takedown procedures.⁶³

Methodologically, the authors noted that, ‘to study algorithmic enforcement by online intermediaries may often need to overcome different contractual barriers imposed by the examined platforms or software owners.’⁶⁴ The authors highlight terms of use which prohibit such tinkering. Internet companies have not readily shared information with researchers about how they process and handle takedown requests, likely because they are wary of increased scrutiny from rightsholders and regulators, or because the technical filtering mechanisms are a source of competitive advantage. In fact, none of the studies reviewed in this chapter obtained data with the cooperation of private companies, other than those made available via the Chilling Effects / Lumen database, or independently through experimentation such as by Perel and Elkin-Koren.

⁶⁰ *ibid* Mayaan Perel and Niva Elkin-Koren ‘Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement’ (2017) 69 Fla L Rev 181, 184.

⁶¹ *ibid* 208.

⁶² *ibid* 209.

⁶³ *ibid* 210.

⁶⁴ *ibid* 215.

5. Balancing of Responsibilities and Costs

An ongoing debate in copyright policy relates to the burden of responsibility for identifying and removing potentially infringing works. While the original DMCA takedown mechanism placed responsibility on the shoulders of rightsholders to monitor, identify and request takedown of infringing material, recent policy discussions have brought focus to re-evaluating the prior balance. Some rightsholder groups would like additional responsibilities placed on platforms (such as the obligation to ensure that content ‘stay down’ after removal).⁶⁵ Legislation adopted in Europe in 2019 would add a licensing obligation that may lead service providers to filter content at the point of upload.⁶⁶ Consequently, an important empirical question relates to understanding how costs of enforcement have been distributed so far, and what the effects of re-balancing those costs may be for internet companies, rightsholders and users.

In a study of the market for out-of-commerce musical works, Heald proposed that notice-and-takedown regimes, in tandem with automatic detection systems, may create a market for previously-unavailable works.⁶⁷ The labour of digitising, uploading and disseminating the work is borne by the uploader, while the rightsholder, once notified, may simply select to monetize the work and collect advertising revenue from it. Heald examined a dataset consisting of 90 songs which reached number one on the pop music charts in Brazil, France and the United States between 1930 and 1960, and an additional set of 385 songs dating from 1919 to 1926 (which should be out of copyright in the United States).⁶⁸ He found that 73% of the in-copyright works in his sample from the USA were monetized by a rightsholder, with a lower rate of monetization in France (62%) and Brazil (39%).⁶⁹ New uploads were less likely to have been monetized, while older uploads, particularly those with higher numbers of views, were more likely monetized.⁷⁰ Similarly to findings by Erickson and Kretschmer, Heald found that uploader creative practices were important in determining rightsholder response. Videos consisting of straight recordings were more likely to be monetized by rightsholders than amateur creative videos or cover performances.⁷¹ However, even these preferences varied by territory: French rightholders monetized a higher proportion of amateur videos and a lower proportion of straight recordings.

In general, Heald found that there were similarly high rates of availability of older in-copyright works (77% had an upload on YouTube) and public domain copyright songs from 1919-1926 (with 75% availability on YouTube).⁷² This rate

⁶⁵ See Martin Husovec, ‘The Promises of Algorithmic Copyright Enforcement: Takedown or Staydown? Which Is Superior? And Why’ (2018) 42 *Colum J L & Arts* 53.

⁶⁶ See Directive 2019/790/EU (n 7) Art 17.

⁶⁷ See Paul Heald, ‘How Notice-and-Takedown Regimes Create Markets for Music on YouTube: An Empirical Study’ (2014) 83 *UMKC L Rev* 313.

⁶⁸ *ibid* 314.

⁶⁹ *ibid* 316.

⁷⁰ *ibid* 319.

⁷¹ *ibid* 322.

⁷² *ibid* 324.

is high compared to other mediums such as books, for example, where only 27% of New York times bestsellers from 1926 – 1932 were found to have copies available to purchase.⁷³ The higher availability of in-copyright works on YouTube, despite the availability of takedown to rightsholders, leads Heald to conclude that the ContentID system creates an efficient form of licensing which reduces transaction costs and enables uploaders to communicate market demand to rightsholders.

Urban, Karaganis and Schofield found that algorithmic ‘DMCA plus’ techniques might be a source of competitive advantage for large incumbent platforms. Based on their qualitative interviews with large and small firms, the authors note that ‘In some striking cases, it appears that the vulnerability of smaller OSPs to the costs of implementing large-scale notice-and-takedown systems and adopting expensive DMCA Plus practices can police market entry, success, and competition’.⁷⁴ Respondents cited the high costs involved, for example, in replicating bespoke systems such as Google’s Content ID, or outsourcing to third-party fingerprinting services such as Audible Magic which was quoted as costing up to \$25,000 per month.⁷⁵ The ability of larger incumbent firms such as Google to monetize all kinds of user-generated content via AdSense and share that revenue with rightsholders via ContentID was also seen as a competitive advantage from the perspective of smaller OSPs. Rather than provide rightsholders the option of leaving such content up on their platforms, OSPs without this technology were limited to taking down videos in their entirety in response to takedown requests.

In addition to differences between large and small commercial enterprises, there are also concerns about the costs of complying with notice-and-takedown procedures for non-commercial institutions. For example, Schofield and Urban analysed the effects of DMCA and non-DMCA takedown requests on the practices of academic digital libraries. Since libraries have undertaken more digitisation as part of their open access public missions, and since public repositories increasingly allow contributions from user-uploaders, this potentially exposes them to DMCA takedown requests. As the authors point out, libraries have historically had ‘sophisticated, careful and public-minded approaches to copyright’ through their handling of physical material.⁷⁶ At the same time, libraries and academic repositories are not typically equipped with resources to handle large volumes of takedown requests such as those received by internet companies.

Schofield and Urban surveyed respondents about institutional practices (how libraries dealt with notices once received, whether they forwarded to other departments, etc.), the volume of requests received and the nature of those requests (copyright or non-copyright). In total, 11 libraries returned surveys and an additional 5 interviews were carried out.⁷⁷ Since 2013, libraries had noted an

⁷³ See Paul Heald, ‘How Copyright Keeps Works Disappeared’ (2014) 11 *J Empirical Legal Stud* 829.

⁷⁴ Urban, Karaganis and Schofield (n 16) 64.

⁷⁵ *ibid* 64.

⁷⁶ Brianna Schofield and Jennifer Urban, ‘Takedown and Today’s Academic Digital Library, 13 *I/S: A Journal of Law and Policy* 125 (2016) 130.

⁷⁷ *ibid* 131-132.

increase in notices received, and some had put in place procedures to deal with DMCA takedown requests. Respondents reported that handling DMCA notices put pressure on staff time. Some libraries expressed that a lack of legal confidence and a requirement to protect their reputation added uncertainty to their job roles. Some erroneous takedown requests were reported to have been received. In one case, the IT department removed a deposited article, which the library later re-instated after careful review (the article was in the public domain).⁷⁸ Overall, the authors found that librarians were more confident dealing with non-copyright removal requests. These included concerns about privacy, sensitivity and security. Librarians had developed institutional norms over time to deal with these matters, but had not yet accomplished this in the realm of digital copyright. This, combined with the high degree of scrutiny and attention paid to evaluating DMCA notices, made institutions potentially vulnerable to an increase in costs related to handling takedown requests.

6. Conclusion: Limitations, Gaps and Future Research

A number of studies reviewed here used data contained in the publicly accessible Chilling Effects / Lumen database.⁷⁹ These are rich data, and the Lumen project provides an interface for researchers to sort and query the voluminous archive. However, use of this database could skew empirical findings in the direction of a small group of intermediaries who share their data (such as Google) and focus attention on particular units of observation (individual notices and claims). A wider range of publicly accessible datasets on takedown would enrich the possibilities for more diverse empirical work. As noted in this chapter, some research has already considered the way that takedown practices are handled in settings such as universities and public libraries by seeking out data directly from those organisations.

The empirical studies reviewed here also demonstrate that the study of copyright notice-and-takedown is a moving target: patterns of behaviour measured by Urban and Quilter in 2006 had shifted in later observations using the same data source. There was an explosion in quantity and diversity of takedown requests and adoption of new practices by both intermediaries and issuers. Even if the current legal regime remains stable, it is likely that practices will continue to shift: new business models may emerge, and rightsholders that were once keen users of notice-and-takedown procedures may drop off as new users appear. For example, the adoption of subscription-based revenue models by firms like Microsoft and Adobe may result in waning investment in enforcement focused on piracy websites. Other rightsholders may find it advantageous to enforce in this manner. The concentration of takedown notices directed at Google Search and YouTube might also change if new platforms become dominant, or if new practices of sharing potentially infringing material emerge.

⁷⁸ *ibid* 138.

⁷⁹ See Lumen (n 12).

Empirical analysis of such trends is held back by a lack of access to data. As we have seen, what is happening inside of ‘black box’ systems often has to be reverse engineered, or revealed by experimental approaches. However, standardized and transparent automated data collection methods are entirely feasible. They will be increasingly demanded by regulators that are tasked with overseeing new obligations and duties on platforms imposed by legislators. It would have been more rational to enable better understanding first, before changes to the liability regime are enacted.

The current state of evidence suggests that, despite its flaws, the notice-and-takedown regime is working. A significant (and after 2013, vast) number of takedown notices are being sent by rightsholders of various types, and processed expeditiously by service providers large and small. The concept of providing safe harbour to innovators while enabling a mechanism for rightsholders to protect their copyrights, appears to be achieving its purpose. Links to infringing materials are being pushed out of the top search results, infringing videos are being removed from sharing websites, and institutions are removing infringing materials hosted on their networks.

The problems, as outlined in this review, remain significant. They relate to redressing contextual imbalances between differently-situated intermediaries, holding rightsholders and platforms to account for accuracy of takedown issuance and compliance, and providing meaningful due process for users whose content is removed. These shortcomings may be addressed through tweaking, rather than overhauling, the safe harbour regime.

There is a deep tension between bringing platforms into the regulatory sphere, and delegating regulatory functions (such as monitoring and filtering) to platforms themselves. The first approach may establish liability rules that platforms cannot escape; the second approach may lead to due process being bypassed. Who, for example, oversees Facebook’s machine learning, artificial intelligence and computer vision technology and their 30,000 human content moderators?⁸⁰ The online world appears to have entered a phase of radical experimentation, exploring new liability rules and new powers for regulatory agencies at the same time without fully understanding the efficacy or shortfalls of the safe harbour paradigm established two decades ago.

Our review indicates that designing effective reporting requirements is critical to enabling empirical assessment of changes to the liability regime. A range of regulatory agencies are now crowding the field, ranging from content, data, and competition to electoral regulators. Is notice-and-takedown still a valid mechanism for addressing these issues, or has it run its course?

⁸⁰ See Casey Newton, ‘The Trauma Floor: The secret lives of Facebook moderators in America’ (*The Verge*, 25 February 2019) <<https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona>>.