

scripted

Volume 15, Issue 2, October 2018

An Empirical Study of the Use of Automated Anti-Piracy Systems and Their Consequences for Cultural Diversity

Sabine Jacques, Krzysztof Garstka,** Morten Hviid,*** John Street*****



© 2018 Sabine Jacques, Krzysztof Garstka, Morten Hviid, John Street
Licensed under a Creative Commons Attribution-NonCommercial-
NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license

DOI: 10.2966/scrip.150218.277

Abstract

EU policy makers are currently debating copyright reforms which aim to provide an ecosystem for the effective combatting of online infringements through the use of algorithms while fostering, amongst other things, cultural diversity. As the goals set by the EU Commission for the Digital Single Market Strategy will only be met if cultural diversity is adequately preserved and promoted, it is important to analyse how algorithms operate as copyright enforcement mechanisms. This article provides an empirical analytical framework on how cultural diversity can be measured in the context of copyright reform suited for the digital environment.

Keywords

Intermediaries, liability, safe harbour, copyright, cultural diversity, algorithms, EU, United Kingdom, freedom of expression, parody

* Lecturer, School of Law and Centre for Competition Policy, University of East Anglia, Norwich, United Kingdom, sabine.jacques@uea.ac.uk

** Research Associate, School of Law and Centre for Intellectual Property and Information Law, University of Cambridge, Cambridge, United Kingdom, kkg22@cam.ac.uk

*** Director, Centre for Competition Policy and Professor, School of Law, University of East Anglia, Norwich, United Kingdom, m.hviid@uea.ac.uk

**** Professor, School of Politics, Philosophy, Language and Communication Studies and Centre for Competition Policy, University of East Anglia, Norwich, United Kingdom, j.street@uea.ac.uk

Jacques, Street, and Hviid are all members of the RCUK Centre for Copyright and New Business Models in the Creative Economy (CREATe) whose financial support is gratefully acknowledged. The usual disclaimer applies.

1 Introduction

The EU Commission aims to contribute to the functioning of the internal single market by increasing the dissemination of cultural expressions originating in other Member States for the benefit of users across the EU territory. Through its *Digital Single Market Strategy* ('DSMS'), the Commission intends to facilitate access and choice of online content as well as to create a fairer and more sustainable market place for creators and creative industries by means of a new Copyright Directive and intellectual property enforcement strategies.¹ As a result, it is presumed that cultural diversity within the Union will be preserved and promoted.² In the proposed Directive, an assumption is made that online video-sharing platforms introducing "effective measures" (which may include automated anti-piracy systems, or "AAPSs") contribute to the preservation of diversity of cultural works.³ To verify this assumption, we have examined the impact of AAPSs on the cultural diversity currently present on video-sharing platforms. The project focused upon YouTube⁴ in particular, and analysed the impact of Content ID on the diversity of expressions over a limited period (2012-2016). Given the significant lack of transparency regarding how Content ID

¹ European Commission, 6/5/2015, *A Digital Single Market for Europe: Commission sets out 16 initiatives to make it happen*, available at http://europa.eu/rapid/press-release_IP-15-4919_en.htm (accessed 20 March 2018).

² Sabine Jacques et al., "Automated anti-piracy systems as copyright enforcement mechanism: the need to consider cultural diversity" (2018) 40(4) *European Intellectual Property Review* 218-228, p. 218.

³ Speech by Vice-President Ansip at CEIPI/European Audiovisual Observatory event "Copyright Enforcement in the Online World" (22 November 2016), available at https://ec.europa.eu/commission/2014-2019/ansip/announcements/speech-vice-president-ansip-ceipieuropean-audiovisual-observatory-event-copyright-enforcement-online_en (accessed 20 March 2018).

⁴ A primarily consumer-to-consumer (C2C) ad-funded hosting intermediary but it is also a Business-to-consumers (B2C) intermediary by its partnership with the creative industries.

operates⁵ and the difficulty in accessing meaningful data,⁶ this study designs an analytical framework for measuring cultural diversity which revolves around variety, balance, and disparity, as well as distinguishing between supplied and consumed diversity in the online music sector.

Cultural diversity is often suggested as a policy goal. However, this concept remains largely undefined⁷ and existing studies have generally focused on a single dimension, namely variety.⁸ In the music field, diversity is generally measured through language, genre, and country of origin. The main problem with existing studies looking at diversity from a single dimension is that we may have a skewed picture of the diversity present in a given sample. While this may give a representation of quantitative diversity, these say very little in terms of diversity in the *content* from a qualitative perspective that focuses, for example, less on numerical appraisal and more on the viewpoints expressed.

In the UNESCO Convention on Cultural Diversity, “cultural diversity” is broadly defined as referring “to the manifold ways in which the cultures of groups and societies find expression”,⁹ which explains the importance of the human rights framework in ensuring its preservation.¹⁰ This broad definition

⁵ Content ID and the structure of other AAPSS’ is currently kept secret as these are used as competitive advantage by intermediaries.

⁶ While “manual” takedown notices are published in the Google transparency report and Lumen project, there is limited information available in relation to the operation of algorithms.

⁷ The UNESCO adopted a Convention on Cultural diversity on 20 October 2005.

⁸ There is no legal definition and no consensus within the literature. For example, Peterson and Berger measures it by referring to the number of different songs in the US top 10 hit parade over a limited period of time, Richard Peterson and David Berger, “Cycles in Symbol Production: The Case of Popular Music” (1975) 40 *American Sociological Review* 158-173, p. 159; or Peter Alexander, “Entropy and Popular Culture: Product Diversity in the Popular Music Recording Industry” (1996) 61 *American Sociological Review* 171-174, p. 171.

⁹ Article 4(1) UNESCO Convention on Cultural diversity on 20 October 2005.

¹⁰ *Ibid.*, Article 2(1).

takes a multi-dimensional approach to variety, balance, and disparity.¹¹ Here, we propose an analytical framework adequate for the digital environment which measures diversity from a multi-dimensional perspective by looking at quantitative and qualitative diversity in the production and consumption of cultural goods.¹²

Focusing on musical parodies, we assess the impact of algorithms on the diversity of cultural expressions under the current legal regime. Section 1 revisits general principles of freedom of expression and the copyright regime. Section 2 reviews the enforcement of these rights in the digital environment. Section 3 attempts to illustrate how these algorithms jeopardise the copyright paradigm and freedom of expression. Section 4 justifies our focus on parodies in this context. Sections 5 and 6 analyse the findings of our empirical study providing an example of the magnitude of the effect of these algorithms. Finally, Section 7 sets out the recommendations for Content ID specifically, and proposes a robust regulatory regime for delineating how online content should be filtered, blocked, or removed in compliance with freedom of expression and cultural diversity. The final Section speculates on what the future may hold.

2 Balancing freedom of expression with proprietary rights

In the absence of a robust regulatory framework, the preservation and promotion of cultural diversity is achieved through the human rights framework and most importantly through the promotion of freedom of expression. Increasingly,

¹¹ Andrew Stirling, "On the economics and analysis of diversity" (1998) *SPRU Electronic Working Papers Series 28*, available at <https://www.sussex.ac.uk/webteam/gateway/file.php?name=sewp28&site=25> (accessed 20 March 2018).

¹² Similar definition adopted by Benhamou and Peltier. See Françoise Benhamou and Stéphanie Peltier, "How should cultural diversity be measured? An application using the French publishing industry" (2007) 31(2) *Journal of Cultural Economics* 85-107, p. 86.

however, copyright is accused of curbing creativity, stifling freedom of expression, and jeopardising cultural diversity in favour of protecting exclusive proprietary rights.¹³ Creators and the wider public criticise the copyright system for not providing enough breathing space for new cultural expressions to thrive. Simultaneously, copyright owners argue that cultural diversity is under threat due to the volume of copyright infringements taking place online. Beyond these positions, it is asserted that, however copyright operates, there is an inherent tension between cultural diversity and freedom of expression, and that maximising one may not maximise the other.¹⁴

This is recognised in law. If freedom of expression consists of a broad¹⁵ right including the right to impart, seek and receive information, its exercise may be subject to restrictions which are partly defined in international instruments. Such restrictions include the protection of the rights of others, which encompasses the right to property and, consequently, copyright.

In European countries, any restriction, including copyright, on the right to freedom of expression must be construed strictly¹⁶ and is limited by a three-

¹³ Carys Craig, "Putting the Community in Communication: Dissolving the Conflict between Freedom of Expression and Copyright" (2006) 56 *University of Toronto Law Journal* 75-114, p. 76; David Fewer, "Constitutionalizing Copyright: Freedom of Expression and the Limits of Copyright in Canada" (1997) 55 *University of Toronto Faculty of Law Review* 175-240, p. 184.

¹⁴ On the ambiguous relation between copyright and freedom of expression, see Alexandra Couto, "Copyright and Freedom of Expression: A Philosophical Map" in Axel Gosseries, Alain Marciano, and Alain Strowel (eds.), *Intellectual Property and Theories of Justice* (Palgrave Macmillan, 2008), p. 160; Christina Angelopoulos, "Freedom of expression and copyright: the double balancing act" (2008) 3 *Intellectual Property Quarterly* 328-353, p. 328.

¹⁵ It includes expressions which might offend, shock or disturb the States or groups of individuals, because such expressions are part of pluralism - essential feature of a democratic society, requiring tolerance and broadmindedness. Being media-neutral, freedom of expression is applicable to all types of communication. *Handyside v UK* (1976) 1 EHRR 737, para. 49 and *Sunday Times (No 1) v UK* (1979) 2 EHRR 245 repeated in amongst others *Ashby Donald and others v France* (2013) Application No. 36769/08, para. 31; *Oberschlick v Austrian* (1991) 19 EHHR 389, para. 57.

¹⁶ Amongst others *Peta Deutschland v Germany* (2012), No. 43481/09, para. 46.

pronged test based on the facts of a particular case.¹⁷ The restriction must be provided by law (principle of predictability), pursue a particular aim (principle of legitimacy), and be necessary (principles of necessity and proportionality). The latter is interpreted as requiring the restriction to be “necessary in a democratic society”.¹⁸

Likewise, copyright legislation does not confer absolute rights to right-holders, because the right to property itself is not absolute. Copyright grants transferable economic and waivable non-economic exclusive rights to an author for a fixed period of time to authorise, amongst others, reproduction and communication to the public of copyright-protected works. Protection is further limited by the requirements for copyright to subsist. For example, copyright does not protect ideas but rather the expression of these ideas, provided that this expression satisfies the originality¹⁹ threshold in relation to authorial works.²⁰ Finally, national legislators can limit the scope of copyright protection by introducing exceptions and limitations, in accordance with the public interest or other policy to implement social, economic and cultural policies.

Given the absence of any hierarchy, no one freedom prevails over the other. Equal weight should be allocated to each right in a manner which ensures the realisation of all freedoms.²¹ This implies a balancing of the rights in play to maximise their realisation. Nevertheless, there is a series of factors that can be

This test was initially introduced in art. 19(3) ICCPR and later repeated in other legal instruments, including art. 10(2) ECHR. This test is applicable to online expressions as well. General Comment No. 34, CCPR/C/GC/34, adopted on 12/9/2011, para. 43.

¹⁸ See art. 10(2) ECHR.

¹⁹ In the EU, this threshold is the “author’s own intellectual creation”. See *Infopaq International A/S v Danske Dagblades Forening* (C-5/08) ECLI:EU:C:2009:465 [2009], para. 37.

²⁰ Literary, dramatic, musical, and artistic works.

²¹ Paul Torremans, “Copyright (and Other Intellectual Property Rights) as a Human Right” in Paul Torremans (eds.), *Intellectual Property and Human Rights* (3rd ed., Kluwer Law International, 2015), p. 252.

used to determine whether a restriction on the exercise of freedom of expression based on copyright is justified.

As the right to freedom of expression is a broad-ranging right, states' and courts' discretion is narrower in the application of restrictions based on copyright. Where a copyright exception aims at preserving freedom of expression, it must not be interpreted *restrictively* to meet the purposes and objectives of said exception. Exceptions such as fair dealing must, therefore, have their purposes interpreted broadly, and the requirements attached interpreted *strictly*.²² Most importantly, both online and offline, the impact of restrictions on the right to freedom of expression based on copyright or other grounds must be carefully scrutinised. Therefore, to respect due process, restrictions on freedom of expression must undergo the proportionality test in each and every case in order to be compatible with human rights. This exercise is consequently of paramount importance to the protection and promotion of cultural diversity online.

3 The rise of automated anti-piracy systems to fight online piracy

Both the EU and the USA have established copyright enforcement regimes which, if not inviting, are at least are compatible with automation.²³ This may be

²² *Johan Deckmyn and Vrijheidsfonds VZW v Helena Vandersteen and Others* (C-201/13) ECLI:EU:C:2014:2132 [2014], para. 22; Attorney General in *Johan Deckmyn and Vrijheidsfonds VZW v Helena Vandersteen and Others* (C-201/13) ECLI:EU:C:2014:458 [2014], para. 43; *NLA* (C-360/13), para. 23; Order in *Infopaq II* (C-302/10), para. 27; *Eva-Maria Painer v Standard VerlagsGmbH and Others* (C-145/10) ECLI:EU:C:2011:798 [2011], para. 109; *Infopaq International A/S v Danske Dagblades Forening* (C-5/08) ECLI:EU:C:2009:465 [2009], paras. 56-57.

²³ For the EU, see proposed new Copyright Directive (art. 13 and Recitals 38 and 39); in the US, the Office of US Intellectual Property Enforcement Coordinator (IPEC) also condoned these private arrangements. See for example Office of the US Intellectual Property Enforcement, *Joint Strategic Plan* (2013), p. 35; *Role of Voluntary Agreements in the US Intellectual Property*

justified by the international intellectual property framework, since art. 8 of the WIPO Copyright Treaty (together with its Agreed Statement) states: “mere provision of physical facilities for enabling or making a communication does not in itself amount to communication within the meaning of this Treaty or the Berne Convention”.²⁴ Hence, signatory parties must “permit effective action against any act of infringement of rights covered by this Treaty, including expeditious remedies to prevent infringements and remedies which constitute a deterrent to further infringements”.²⁵

While the E-Commerce Directive and Digital Millennium Copyright Act do not mention AAPSs explicitly, the trend towards the latter could be a path to practical compliance with the WIPO treaties. The burden currently falls on the right-holders to prove that an intermediary does not meet the safe harbour requirements which, in most cases, results in high costs for right-holders.²⁶ Preparing a valid takedown notification every time an infringing copyright work is discovered is also burdensome for the right-holders, making automated notification to intermediaries an attractive prospect. Faced with a potentially huge number of notices which must be dealt with expeditiously, intermediaries might wish to automate their *processing* as well as the procedure of *notifying* users that some of their content is being blocked. Hence, there are incentives for both right-holders and intermediaries to develop a fully automated systematisation of

System: Hearing Before the Subcommittee on Courts, Intellectual Property, and the Internet of the Committee On the Judiciary, House of Representatives (113th Cong. (2013)).

²⁴ WCT Agreed Statement regarding article 8.

²⁵ Arts. 14 WCT and 23 WPPT.

²⁶ Some right-holders argue that these intermediaries should act as gatekeepers as the advent of technologies puts them in a better position to tackle online copyright infringement. Benjamin Boroughf, “The next great YouTube: Improving Content ID to foster creativity, cooperation and fair compensation” (2015) 25(1) *Albany Law Journal of Science & Technology* 95, p. 103; Rebecca Alderfer-Rock, “Fair Use Analysis in DMCA Takedown Notices: Necessary or Noxious?” (2014) 86(3) *Temple Law Review* 691-720, p. 694.

online copyright enforcement. Moreover, there is an incentive for right-holders (typically represented by collecting rights societies) and intermediaries to collaborate on the design of such tools. Yet, one might wonder whether such collaboration in a digital environment (which facilitates the connection of cultures) creates a danger that unique cultural differences will be diluted by homogenisation, or misappropriated owing to ineptitude on the part of the legal framework.

Hence, in the absence of any over-arching legal obligation, algorithms have been introduced because of voluntary private initiatives, built upon agreements between right-holders and intermediaries, which increase the right-holder's control over works shared online by third parties.²⁷

3.1 The functioning of Content ID

YouTube operates the most well-known example of such an initiative, known as Content ID. First introduced mid-2007, this rights management system was the result of agreements between YouTube and a number of the major record labels and musical publishers, extended later to independent labels too. A proportion of revenue generated through advertising is shared between YouTube and the right-holder according to the agreed terms.

²⁷ Qualified as "DMCA Plus" intermediaries in the seminal empirical study from 2016, by Urban, Karaganis and Schofield. Later, Bridy distinguished two types of DMCA Plus intermediaries. DMCA Plus type 1 relates to intermediaries covered by the DMCA safe harbours and which voluntarily venture into private monitoring initiatives and type 2, comprising the intermediaries who venture into similar voluntary private initiative but which do not meet the safe harbour provisions" requirements. Jennifer Urban, Joe Karaganis and Brianna Schofield, "Notice and Takedown in Everyday Practice" (29 March 2016) UC Berkeley Public Law Research Paper No. 2755628; 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), pp. 185–208, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2628827 (accessed 21 March 2018).

Using digital fingerprinting technology,²⁸ the Content ID algorithm cross-checks all newly-uploaded content against an established database of copyright-protected works based upon the partners' repertoires. This database comprises a reference file for the copyright-protected work as well as "asset" metadata, i.e. a wide range of the content's inherent properties. The algorithm is able to detect any part of audio or video content, even if modified, which is stored in the database repository. A match results in an automated notification being sent to the relevant partners, who can customise their response. According to Youtube's least to most restrictive options, a partner can (1) do nothing, (2) add advertising and collect the revenue, (3) monitor its viewing statistics, (4) block its content (the content will not be audible or viewable on YouTube), or (5) issue a manual take-down request. Partners may choose a course of action in advance so it will be pursued automatically in the event that matching content is uploaded to the platform. The action can even be customised by the partner with respect to the extent of the match, its format (sound, video, or both), and the territory in which it is applied.

3.2 Limitations

The system is not without limitations. The algorithm is unable to detect whether a work is in the public domain, falling outside the scope of copyright, or covered by an exception or limitation.²⁹ To mitigate this shortcoming, YouTube introduced an internal dispute resolution system as required under s. 512(g) of

²⁸ Digital fingerprinting is to be distinguished from watermarking. See Dominic Milano, "Content control: digital watermarking and fingerprinting" (Rhozet White Paper), available at https://www.digimarc.com/docs/default-source/technology-resources/white-papers/rhozet_wp_fingerprinting_watermarking.pdf (accessed 21 March 2018).

²⁹ Jacques, *supra* n. 2, pp. 225-227.

the DMCA.³⁰ However, as explained in an earlier paper, the absence of independent oversight puts too much power in the hands of partners.³¹

4 Choosing musical parodies as a case study

Empirical research into the impact of algorithms in decision-making has mainly focused on the efficiency of these in content detection, and their implications for the music industry.³² With this research project, we design a theoretical framework to evaluate cultural diversity online from a multi-dimensional perspective, taking into consideration both quantitative and qualitative content aspects. To do so, we decided to focus on the case of parodies, because this particular form of expression enables individuals to critique, through comedy, mockery, or satire, an idea, value or person. Such critiques both represent a lynchpin of democracy, and offer qualitative aspects worthy of study. Furthermore, parodies rely on other works to exist, which is easier for the algorithm to detect. Thus, parodies constitute an effective means of measuring how content is reused and what kind of new cultural expressions flourish through this genre.

Before turning to the role of algorithms in preserving the diversity of parodic expressions, it is important to note some of the general challenges for parodies to thrive in the digital era. To participate in a democratic society, citizens must be able to accept or reject, via ridicule or parody, the messages, cultural values, attitudes, or other forms of behaviour that constitute their society.

³⁰ *Ibid.*

³¹ *Ibid.*

³² Mainly to process the manual notice and takedown. Paul Heald, "How notice-and-takedown regimes create markets for music on YouTube: an empirical study" in Andreas Rahmatian (ed.) *Concepts of Music and Copyright* (Edward Elgar, 2015), p. 195; Ben Depoorter and Robert Walker, "Copyright False Positives" (2013) 89(1) *Notre Dame Law Review* 319-360, p. 326.

Simultaneously, EU copyright legislation requires a high level of copyright protection,³³ providing strong incentives for the promotion of cultural endeavours.³⁴ The fragmentation of legislation and divergence in judicial interpretation means however that legal uncertainty has arisen.

Relying on the Berne Convention,³⁵ several legislators have introduced a parody exception in their copyright laws to strike a balance between the competing interests at stake. Whether the argument is based on further realising fundamental human rights,³⁶ fostering creativity to contribute to social welfare by favouring discursivity,³⁷ adjusting the market,³⁸ or a combination of these approaches, the parody defence provides an internal limit to copyright law in favour of users and future creators, and at the expense of the original creator.

³³ Preamble of the Infosoc Directive, Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market, OJ L 178, 17.7.2000, pp. 1–16.

³⁴ Impact assessment copyright package 2016, p. 63, available at <https://ec.europa.eu/digital-single-market/en/news/impact-assessment-modernisation-eu-copyright-rules> (accessed 21 March 2018).

³⁵ Berne Convention, art. 9(2).

³⁶ Mary Wong, "'Transformative' User-Generated Content in Copyright Law: Infringing Derivative Works or Fair Use?" (2009) 11 *Vanderbilt Journal of Entertainment and Technology Law* 1075-1139, p. 1081; Michael Spence, "Intellectual property and the problem of parody" (1998) 114 *Law Quarterly Review* 594-620, p. 601; Ellen Gredley and Spyros Maniatis, "Parody: A Fatal Attraction? Part 1: The Nature of Parody and its Treatment in Copyright" (1997) *European Intellectual Property Review* 330-344, pp. 343-344.

³⁷ In the UK, see HM Government, *Consultation on copyright* (2011), para. 7.103; *Digital Opportunity: A Review of Intellectual Property and Growth: An Independent Report by Professor Ian Hargreaves* (IPO, 2011); *Taking forward the Gowers Review of Intellectual Property: Second Stage Consultation on Copyright Exceptions* (IPO, 2009); *Taking forward the Gowers Review of Intellectual Property: Proposed changes to copyright exceptions* (IPO, 2008); *Gowers Review of Intellectual Property* (HMSO, 2009); More generally, see Séverine Dusollier, *Droit d'auteur et protection des œuvres dans l'univers numérique* (Larcier, 2007), p. 466.

³⁸ Anna Spies, "Revering Irreverence: A Fair Dealing Exception for Both Weapon and Target Parodies" (2011) 34(3) *UNSW Law Journal* 1122-1144, p. 1141; Wendy Gordon, "Excuse and justification in the law of fair use: transaction costs have always been part of the story" (2003) 50 *Journal of the Copyright Society of the U.S.A* 149-198, pp. 158 and 166; Spence, *supra* n. 36, p. 603.

While the boundaries of a specific parody exception in copyright law remain to be defined,³⁹ restrictions nevertheless operate on it. Firstly, various legal instruments do not define the term parody and related concepts, resulting in different definitions per territory. While the absence of a definition enables more flexibility to adapt the law to users' habits, it might have a deterrent effect until a body of case law provides more guidance. Secondly, there are differences in the interpretation and application of the parody exception amongst the various jurisdictions. These differences render the duties of intermediaries even more difficult, and hamper the development of a digital environment for copyright protection and creativity. Recently, the CJEU had its first opportunity to interpret the parody exception under EU copyright law in the *Deckmyn* case.⁴⁰ The Court noted that parody had to be interpreted as an autonomous concept requiring two core elements: an expression of humour or mockery which, while evoking an existing work, is noticeably different from that work.⁴¹ Therefore, if a parody does not satisfy these conditions, the parodist should get a licence from the right-holder to avoid copyright infringement. If an online video shared on YouTube meets these conditions, it should be deemed as lawful but, as discussed in Section 3, it is likely to be picked up by Content ID as infringing.

5 Methodology

The empirical analysis builds on the dataset used in two previous studies by Kris Erickson et al. in 2013 and 2014.⁴² These studies analysed 1845 user-generated

³⁹ Sabine Jacques, "Are the new 'fair dealing' provisions an improvement on the previous UK law, and why?" (2015) 10 (9) *Journal of Intellectual Property Law and Practice* 699-706, p. 703.

⁴⁰ *Deckmyn and Vrijheidsfonds VZW v Vandersteen*, C-201/13, EU:C:2014:2132.

⁴¹ *Ibid.*, para. 20.

⁴² Kris Erickson et al., "The reasons for copyright takedown on YouTube, and what they tell us about copyright exceptions", paper presented to the EURO CPR Conference, 24-25/3/2014, Brussels available at <http://eprints.gla.ac.uk/94718/1/94718.pdf> (accessed 21 March 2018);

parodies, stemming from 343 commercial pop music videos shared on YouTube. The original study was carried out in January 2012 and was based on content produced in the previous twelve months. The data collection relied on:

British Charts Company data to obtain a list of 343 music singles that charted in the UK in 2011 and had a corresponding licensed music video hosted on YouTube (either the VEVO or Warner licensed music channel). The authors then searched for parody videos referencing those commercial works by searching for “song name + parody” in YouTube’s internal search engine. The researchers located 8299 user-generated music video parodies referencing the original 343 commercial music videos. A randomly-selected sample of 1843 parodies from within that larger sample... The authors also recorded the location (URL) of each of those 1845 parody videos to enable future analysis.⁴³

The present research extends the data collection up to December 2016. By using the content’s location (URL), we assessed the *way* content is blocked or taken down by identifying and isolating which videos were taken down in 2016 due to a copyright claim. We also distinguished videos which were manually taken down from those blocked by the Content ID match policies.⁴⁴ This paper focuses on the evaluation of the supplied and consumed product diversity of a cohort

Kris Erickson et al., *Copyright and the Economic Effects of Parody: An Empirical Study of Music Videos on the YouTube Platform and an Assessment of the Regulatory Options Parody and Pastiche. Study III* (2013) an independent report commissioned by the Intellectual Property Office (IPO) available at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/309903/ipresearch-parody-report3-150313.pdf (accessed 21 March 2018).

⁴³ *Ibid.*, pp. 10-11. As the methodology for this dataset has been explained in both earlier studies, the present authors refer the reader to these for further details on the data collection.

⁴⁴ This was possible by combining checking each message displayed on unavailable content (“blocked”, referring to a Content ID match policy, and “removed” signifying a manual takedown) with the information about manual takedowns gathered from the Google Transparency Report and Lumen Project.

(Erickson et al.'s cleaned sample – 1839 videos) at three different moments t_1 (2012), t_2 (2013), and t_3 (2016).

To reflect fundamental rights as well as copyright concerns, we opted to evaluate both supplied and consumed diversity. Supplied diversity refers to the content available on the platform that *could* be accessed by the public. This measure focuses on the individual's right freely to express themselves. Consumed diversity focuses on what is *actually being watched* by the public and focuses on how voices are being heard.⁴⁵ While diversity legitimately may focus on just supplied expressions, in order to be relevant freedom of expression must combine both the supply and the demand sides.

If the focus is solely on supplied diversity, then counting (say videos or parodies) is sufficient and a useful index is simply the inverse of the number of videos or parodies. Once we are concerned with consumed diversity, we need a measure of diversity allowing for differences in popularity. One measure which does so is the *Herfindahl Index* (also termed the *Herfindahl-Hirschman Index*, hereinafter 'HHI'). While HHI is typically associated with industrial economics, where it is used to measure the concentration of firms in a market, it is also used to measure diversity where the subject of the research has been classified into types.⁴⁶ Initially developed by Edward H. Simpson ("Simpson Index of Diversity"), the index presents itself as follows:

⁴⁵ EU Parliament resolution of 12/09/2013 on promoting the European cultural and creative sectors as sources of economic growth and jobs (2012/2302(INI)), at M: "EU citizens need to be provided with a cultural and artistic education from an early age so as to develop their own understanding of arts and culture, make their voice *heard* and develop awareness of the great diversity of cultures present in Europe, and thereby promote their own creativity and expression, as well as cultural diversity".

⁴⁶ Radu Guiasu and Silviu Guiasu, "The weighted Gini-Simpson Index: Revitalizing an old index of biodiversity" (2012) *International Journal of Ecology*, available at <http://dx.doi.org/10.1155/2012/478728> (accessed 21 March 2018).

$$HHI = \sum_{i=1}^N p_i^2$$

N means the number of types of interest and p_i the proportion of views of the i^{th} type of video. With N types, if all types attract the same number of views, $p_i = \frac{1}{N}$ and $HHI = \frac{1}{N}$.⁴⁷ Thus if each type is equally popular, it is as if we are just counting varieties. If one type, say j , attract all the views, $p_j = 1$ and $p_i = 0$ for all other types' i s, and so $HHI = 1$. The value of this index will therefore vary from $\frac{1}{N}$ to 1. To put it simply, the closer the value is to 1, the lesser the diversity. An alternative way to use the HHI to illustrate diversity (when popularity is a legitimate aspect to consider) is by calculating the inverse of the index. This measures how many categories with identical numbers of views would yield the same HHI and is sometimes referred to in competition analysis as the "Numbers-Equivalent" or in political science as the "Effective Number of Parties" ("Eff N ").⁴⁸ In our context, this index tells us that it is as if there were only Eff N videos available to the public rather than the 1839 videos constituting our sample.⁴⁹

These indices enable a multi-dimensional approach that considers variety, balance, and disparity. Diversity is not just about the number of different varieties of video available, but also about how likely they are to be watched by people, since diversity may actually be located in the tail. If one variety attracts a majority of views, we may have diversity in theory but not in reality because network effects (the desire to watch what others are watching) may render the

⁴⁷ $HHI = \sum_{i=1}^N \left(\frac{1}{N}\right)^2 = N \cdot \frac{1}{N^2} = \frac{1}{N}$.

⁴⁸ Markku Laakso and Rein Taagepera, "Effective Number of Parties: A Measure with Application to West Europe" (1979) 12(1) *Comparative Political Studies* 3-4, pp. 3.

⁴⁹ Whether this is the relevant measure depends on how one assumes that a video is discovered – do all videos available on YouTube have an equal chance of being detected or does this depend on popularity. If the case is the latter, then the effective measure used above is relevant.

other varieties largely unseen. Given our purpose, the Eff N is a reasonable way to measure the number of types present on YouTube and their relative abundance. While most studies generally focus on variety (the number of different categories),⁵⁰ our system follows Stirling's⁵¹ properties and therefore also includes balance (the pattern of quantity across the categories)⁵² and disparity (how dissimilar the categories are within a set).⁵³ This gives a more in-depth analysis of the diversity present within the various categories.

Before turning to the preliminary results of this model, we need to make a few remarks about the dataset (beyond the difficulty of trying to quantify something inherently qualitative, namely cultural diversity). Firstly, this dataset relates exclusively to the UK music market. Secondly, given the sheer volume of content present, the size of the dataset may be distorted, given that the total number of views in 2012 (consumption of cultural content) is slightly above 253.4 million and that these are accounted for by a relatively small number of videos. The share of views of the top five videos in 2012 amounts to 36.5% of views. The corresponding numbers for the top ten was 50.7% in 2012, and of these only 6.6% remained viewable in 2013 with a further reduction of those of 84.3% in 2016. The top twenty-five accounted for 71.9% of all views of which only 1.6% remained in 2013 and even less in 2016. Thirdly, the original dataset comprises parodies based on 343 commercial pop music videos shared on YouTube and derived from British Charts Company data. These 343 commercial hits are therefore not fully representative of the whole market, but this problem is mitigated by the addition

⁵⁰ The more categories, the more diverse is the set.

⁵¹ Stirling (1998), *supra* n. 11, pp. 39-41.

⁵² The more equal are the shares of views in a category measured against the totality of the set, the more diversity is present.

⁵³ This is subjectively assessed and context-dependent explaining why this column could be overlooked. The more disparate the options are within the set for identical variety and balance, the more diverse the set is.

of user-generated parodies. This produces 8299 videos, from which the 1843 sample is derived. Finally, there are likely to be some problems of transparency and robustness because of issues of data quality. The data was collected for different initial purposes, and hence this has constrained us in the variables used and in the building of indexes. Given the limitation of the data, care must be exercised not to over-interpret the results. The analysis in Section 6 illustrates both what can be done with, and potential pitfalls from relying too mechanically on, the statistical analysis.

6 Findings

The assumption is that the most harmful impact on cultural diversity is that caused by the unavailability of content, rather than the tracking statistics or monetisation match policies which still enable the public to access the online content. The preliminary results demonstrate patterns in the application of YouTube's Content ID on the diversity of cultural expressions supplied and consumed on the platform.

6.1 Content blocking significantly impacts both supplied and consumed diversity

Before taking into consideration the role of algorithms, the following general observations can be made based on how the initial Erickson sample was constituted.

The first observation relates to whether consumed diversity of the original videos matches diversity in supply of content to viewers. Calculating the Eff N based on the share of views of each of the 343 original videos, we find that Eff N = 75.1. In other words, the consumed diversity is considerably lower than the supplied one. The second observation focuses on whether the incidence of

parodying is equally spread over the 343 original videos. Types in this case are each of the original videos, i.e. $N = 343$. The total number of parodies was 8299. Calculating the Eff N first for the number of parodies for each type, we find Eff N = 25.5. Contrasting this with $N = 343$, it is obvious that the supply of parodies is extremely skewed. If all the videos were equally popular as object for a parody, it is as if we only started with about twenty-five videos. If we turn to the popularity of these parodies, we can again compute the Eff N, in this case finding Eff N = 22.5. Thus, there seems to be a concentration (even if not by much) on a few original videos as object of parody. That does not imply that there is a close correlation between the number of parodies of a given video and their respective popularity. This illustrates a case where counting numbers (for example variety) is insufficient and requires considering balance and disparity.

Interestingly, the dataset seems to imply that it is not because an original hit attracts a lot of parodies that the sum of parodies is significant. Taken collectively, these parodies seem to account merely for a small share of overall views. For example, there may be many parodies of video X, but not many people are actually watching these. As an example, the video with the most parodies by number attracts almost 11% of all parodies both by number and by popularity, while the videos attracting the second and third highest number of parodies (9.9% and 8.9%, respectively) attract a much smaller share of views (6.3% and 2.7%, respectively). Equally, while the top ten parodies by views attracted more than half of all views, they accounted for less than a third of the actual number of parodies. Thus, it is not just important to decide whether the key consideration

is supplied or consumed diversity, but also what the object (of the hits or of the parodies) of diversity is.⁵⁴ That may depend on the policy question asked.⁵⁵

Turning to the availability of the parodies sampled in 2012, and later in 2013 and 2016, we observe the impact of the technology on the environment in which videos are shared.

⁵⁴ See Section 7.1 *infra*.

⁵⁵ For example, do we care about the number of parodies an original has attracted? Or should we focus on videos which generate a debate amongst these?

Descriptive statistics	2013		2016	
	N	Percent	N	Percent
Live	1471	80.0%	1088	59.2%
Manual copyright take-down			118	6.4%
	265	14.4%		
Algorithm-based blocking			591	32.1%
Not available for other reason*	103	5.6%	42	2.3%
Total N	1839	100	1839	100

	2013	2016
Effective number of videos viewed	27.2	20.2
Effective number of YouTubers	-	13.6
Top 5 YouTubers' share of total views	-	55.5%
Top 10 YouTubers' share of total views	-	72.2%

* This includes take down for other reasons than a copyright claim.

Table 1: General descriptive statistics

The first observation from Table 1 is that the effect of Content ID cannot be underestimated. If we combine all the content removal and blocking due to copyright concerns, we can see that in December 2016 Content ID accounted for

83.4% of the blocking, while manual take-down only accounted for 16.7%.⁵⁶ Such figures could demonstrate the efficiency of the technology for removing copyright infringing content, but this would be a fallacy because, as the original study from 2013 by Erickson et al. revealed, only a very small percentage of audio-visual works in the sample infringed copyright.⁵⁷

The second observation is that when we aggregate views up to the level of the numbers of YouTube users posting videos, only a few videos attract most of the views: the top five in the sample by number of views account for slightly over a half (55.5%) of all views, while the top ten account for almost a quarter (72.2%). Hence, when analysing diversity, it matters greatly whether we focus on “the number of voices” or “the number of voices *heard*”. In the former case, we can simply count the number of voices, i.e. YouTube users. In the latter case, the popular videos need to carry more weight in our measure and we will resort to the indices described in Section 5 *supra*. Using Eff N yields our third observation from Table 1, namely that out of our cohort of 1839 videos, only twenty-seven videos were effectively viewed in 2013 and only twenty remained in 2016. This tells us that viewing figures are very skewed in our sample. Even more astonishing is the number of effective YouTube users present in this sample. As opposed to the 940 YouTubers sharing content in 2012, it is if there are as few as thirteen effective YouTube users in 2016. While this can be explained by the fact that most of the views are concentrated in the top five (55.5%) and top ten (72.2%),

⁵⁶ Google advertised that overall, in 2016, that 98% of allegedly copyright infringing material is dealt through Content ID and 2% relates to manual notice and take-down. Google, *How Google fights piracy* (2016), p. 26 available at <https://drive.google.com/open?id=0BwxyRPFduTN2TmpGajI6TnRLaDA> (accessed 21 March 2018).

⁵⁷ Only 13% of all the UGC were found as mislabelled as “parody” and this does not mean that these 13% might be infringing as other exceptions might be applicable. 5% were also found are direct copy and therefore highly likely that these 5% constitute infringing materials Erickson, *supra* n. 42, pp. 27 and 31.

it is surprising that the sample is so skewed that out of so many audio-visual works shared, only a handful of content creators are able effectively to communicate their creative expression to the public. From the perspective of consumed diversity, this turns out to be surprisingly low. While this could be an artefact of the particular sample, it is also possible that this is a more robust finding if there are strong network effects at play when it comes to the popularity of items.

Furthermore, audio-visual works which copy the video are more likely to be picked up by Content ID than to be removed by a notice and take-down complaint.⁵⁸ This is easily explained by the functioning of digital fingerprinting technology as described in Section 3. As this is how the technology is meant to perform best (i.e. to find visual or audio matches) this is where the impact will be greater. Nevertheless, this raises concerns as to the likelihood of meeting the objectives sought by the introduction of a parody exception in copyright law. Indeed, while some countries seem to be permitting the copying of entire works (for example copying of sound recordings with altered lyrics),⁵⁹ it remains to be seen whether the application of the parody exception in other jurisdictions, including the UK, will result in the same or similar decisions.⁶⁰

6.2 Case study: Gender and diversity

While a number of features of a particular parody, such as its target, the degree of professionalism of the product, and the level of critique involved, may have

⁵⁸ By running a multinomial regression, our hypothesis leads us to believe that an audio-visual work which copies the sound recording of a copyright protected work has significant chances of being blocked by the algorithm, rather than remaining available.

⁵⁹ France: Cass. Civ. 1, 12 January 1988, *RTD Com.* 1988, 227; André Françon, "Questions de droit d'auteur relatives aux parodies et productions similaires" (June 1988) *Droit d'auteur* 303.

⁶⁰ Jacques, *supra* n. 39.

an impact on which of the rules are implemented in response to the algorithm's finding of a match, gender should not be one of these. We use this to investigate deeper into both the original data set and our extension of the data until 2016.

Looking at the 2012 data only, but before sampling of the parodies had taken place, i.e. the full set of 8299 parodies of 343 original videos, we find that among the original data, male performers dominate in terms of the original videos (accounting for 58.3% of the original videos in the sample), while female artists dominate when it comes to being parodied, whether by number of parodies or by the popularity of the parody (see Table 2).

	<i>Original videos</i>	<i>Parodies</i>	<i>Views of parodies</i>
<i>Female</i>	33.2%	55.0%	52.7%
<i>Male</i>	58.3%	42.5%	45.3%
<i>Mixed</i>	8.5%	2.5%	2.0%

Table 2: Share by gender of performers

In other words, in the sample, women performers are disproportionately likely to be parodied. Who then are engaging in the parody and to what extent is this affected by the algorithm? For both 2012 and 2016, we can divide the sample into five types (male solo, male group, mixed, female solo, female group). Calculating the Eff N for the 2012 sample and for those remaining in 2016, we get a value of 3.05 and 3.10 respectively, indicating that the parodies are concentrated on a subset of the five types and that there is not much difference between the two dates. Investigating bit further into the data (Table 3), the absence of female parodists is clear cut. Purely female parodists account for only 23%–24% of the parodies in the sample and more dramatically only for 3%–3.5% of parodies by views.

<i>Gender</i>	<i>By number 2012</i>	<i>By views 2012</i>	<i>By number 2016</i>	<i>By views 2016</i>
<i>Female Group</i>	11.9%	1.1%	11.2%	1.1%
<i>Female solo</i>	12.1%	2.5%	11.9%	2.0%
<i>Male Group</i>	20.8%	36.8%	21.5%	38.8%
<i>Male Solo</i>	37.0%	21.5%	35.7%	25.7%
<i>Mixed group</i>	18.2%	38.1%	19.6%	32.5%

Table 3: Share of parodies

For some time now, the platform has recognised that women were outnumbered as content creators. Our results show that for this sample, they are also outnumbered in terms of share of popularity. This can perhaps be explained by the environment of the platform. For example, considering the nature of the comments made on female performer videos, many of these are sexually suggestive, threatening, or hateful in nature. Even if one disables the comment function on the platform, this would not alter the central problem. While women remain a minority on YouTube, there will be relatively few safe spaces for women's expression to flourish.

6.3 Remarks on the data analysis

In a very specific dataset we have found that that both supplied and consumed diversity have only limited correlation with the presence of Content ID – in most cases the change in diversity has been very modest and the direction of change has not been consistent. That does not imply that there are no problems; it may simply be that the existing data is not sufficiently well suited to detect such effects. What we have found is that it matters greatly how we conceive of, and

hence quantify and qualify, diversity. A key question we should ask in future work is: how do people find a video? If we are dealing with a neutral search engine, it is possible that merely counting the number of videos in each category will be sufficient, but if videos are found by word of mouth recommendation, then some sort of epidemic model is needed which would possibly lead us (like a short-cut) to use the number of views as a basis to calculate shares. But it may also be that people use a search engine which uses a particular algorithm to decide the order in which it shows search results (based on advertising, for example).

Secondly, there is often a clear “most popular parody”. If people learn from each other, we would expect network effects to operate so that once a video gets to a level of popularity it dominates that category. One question we could not explore is what would happen if the AAPS removes the most popular parody: would the next most popular simply replace it, almost as if the AAPS functions as a forest fire which, in taking out the old trees, gives breathing space for new growth? If that is the case, then while the take-down by the AAPS 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.

Thirdly, the Eff N is a useful statistic to base an assessment of diversity on variety, instead of intensity. However, it can present a misleading picture where the identity of the category with the largest share changes, but not the relative shares. In cases such as some of those explored here, where number of types is small it is both feasible and essential to investigate further.

7 Proposals to preserve and promote supplied and consumed cultural diversity using AAPSS

Given the limited correlation between enforcement and cultural diversity, this

section aims to make a set of modest proposals for preserving and promoting the rich cultural diversity present in the online environment in the European Union. The proposals focus firstly on the EU regulatory framework and newly proposed amendments, before revisiting general principles related to due process and “best practices”. Finally, we discuss the idea that technologies could be used to facilitate licensing opportunities directly within online sharing platforms.

7.1 Fostering cultural diversity through a robust and coherent legal framework

The last fifteen years have taught us that a piecemeal regulatory framework fragments the EU digital single market. Instead of creating duplicates by repeating proposals made by others,⁶¹ we believe there is a real opportunity for the EU to implement a legal framework where a diversity of cultural expressions is promoted despite the current lack of robust international legal provisions. This can be achieved through the strengthening of copyright exceptions and further protection of the public domain by expanding the mandatory character of exceptions, and by providing adequate incentives for bigger industry players such as intermediaries, publishers, record labels and collecting rights societies, to respect both fundamental rights and the public interest.

The main concern is that the more trust that is placed in the hands of commercial private entities, the less likely it is that economic, cultural, or social diversity will be present in online sharing platforms. Advertising-funded online sharing platforms can increasingly bind together individuals living in different territories. The preservation and promotion of cultural diversity comprises two

⁶¹ Such as adding clarity in the legal framework by redrafting art. 13 of the new proposed directive, delineating the interaction between this new proposed directive and the other EU instruments such as the E-Commerce and Enforcement Directives and the EU Charter, and, in general, rationalising the adequacy of liability regimes for online intermediaries.

strands: (1) to enable individuals to have equal access to different cultures, and (2) to enable individuals located in one territory to be heard in another. Therefore, to avoid cultural homogenisation within the EU digital single market, legal provisions should ensure efficient enforcement of fundamental rights (freedom of expression and copyright, through the right to property), and support principles of democracy throughout the EU.

The problem with this proposal is getting agreement on how we define cultural diversity. If the concept spreads across policy debates but remains a vague concept, it will be hard to preserve and promote it. With this study, we have provided variables through which diversity in online sharing platforms can be measured, although statutory metrics would be welcomed and would facilitate its adjudication.⁶²

An alternative approach might be to promote democratic principles in the online environment by adopting a balanced copyright paradigm, compliant with freedom of expression.⁶³ In essence, such a proposal would ensure that the traditional instruments capable of promoting cultural diversity can be fostered in the digital era. By guaranteeing that legitimate expressions are not subject to market censorship, the public is able to determine which cultural expressions can be accessed and consumed. In the longer run, such a position enables different cultures to interact and foster creativity and innovation, possibly leading to economic growth and competition within the EU and the global digital market. In essence, there is a need to ensure cross-border flows of repertoire and user-generated content on platforms.⁶⁴

⁶² European Commission, *The AB Music Working Group Report December 2015-June 2016*, p. 15, available at https://ec.europa.eu/programmes/creative-europe/sites/creative-europe/files/ab-music-working-group_en.pdf (accessed 22 March 2018).

⁶³ *Ibid.*, p. 53.

⁶⁴ *Ibid.*, p. 39.

7.2 Ensuring due process online

Moving on to enforcement, blindly trusting private commercial interests to tackle online copyright infringement is a mistake. Not only do the industry players and intermediaries often have opposing commercial interests, but they rarely have the individual and wider public's best interest at heart. Therefore, we urge EU policy-makers to place obligations on intermediaries and right-holders to respect due process as a counter-weight to the reliance on algorithms to detect and identify infringing content.

Where business models develop in conjunction with the implementation of the legal framework, intermediaries try to protect themselves by applying match policies without due assessment of their illegality. Current business models seem to take little note of copyright exceptions, jeopardising the careful balance struck by legislators throughout the Union. The current copyright package falls short on this front as well. By requiring intermediaries to negotiate with right-holders and act as "judges" on their platforms, the risk is that major market players are going to abuse the system, curtailing freedom of expression and endangering cultural diversity.

Consequently, the new copyright paradigm could push for a more efficient and independent dispute resolution system. This is not to say that such system could not include automation. Indeed, the sheer volume of shared content makes it impractical to involve the judiciary every step of the way. Nevertheless, the possibility of using the judicial or administrative authorities appears essential for the safeguarding of cultural diversity.⁶⁵

⁶⁵ For example, see Bundesgerichtshof [Federal Court of Justice of Germany], *GEMA v Rapidshare*, I ZR 80/12 (15 August 2013).

For example, 90% or some other agreed-upon large portion of matching content could be left to the internal dispute resolution set up by intermediaries, using AAPSs as they currently operate. For smaller portions of matches of alleged infringement, intermediaries are not best placed to assess the legality of the use. Here, AAPSs are inefficient, and so the option of presenting the facts to an independent authority becomes necessary.⁶⁶

In any case, more transparency and harmonisation in the design of these internal dispute resolution systems and external recourses must be contemplated. While parts of these enforcement mechanisms can be left to self-regulation, some harmonisation would increase legal certainty and reduce the costs and efforts of individuals in trying to understand how to best dispute an automated notification of infringement across the myriad of online platforms.⁶⁷ Furthermore, as Garstka indicated, harmonisation in this field could give much needed attention to human rights such as freedom of expression, the right to privacy, and the freedom to conduct a business.⁶⁸ Finally, the intermediaries could be obliged to publish the outcomes of dispute resolution settlements on their website or on an external source, as in Google's *Transparency Report*⁶⁹ or the *Lumen Project*.⁷⁰ Not only would such an initiative enhance oversight of the intermediaries' actions and attitudes in the described copyright enforcement process, but it would also educate the wider public as to which uses are permitted.

⁶⁶ For example, in France where an administrative authority was established to target illegal activities.

⁶⁷ Krzysztof Garstka, "Looking above and beyond the blunt expectation: specified request as the recommended approach to intermediary liability in cyberspace" (2016) 7(3) *European Journal of Law and Technology* 1-23, p. 10.

⁶⁸ *Ibid.*, p. 12.

⁶⁹ <https://www.google.com/transparencyreport> (accessed 21 March 2018).

⁷⁰ <https://lumendatabase.org> (accessed 21 March 2018).

7.3 Streamlining intermediaries' "best practices" for cultural diversity to strive

The proposed art. 13(3) relies on hosting intermediaries and right-holders to negotiate best practices. There are dangers in such an approach. Firstly, the current online creative ecosystem already relies on private agreements reached between right-holders and intermediaries at the expense of the wider public. As an illustration, see the former PRS-YouTube deal:

If PRS for Music becomes aware that a members' work fits the definition of derogatory use, PRS for Music can notify YouTube on that member's behalf and work with YouTube to remove that content. Under the Joint Online Licence, a derogatory use is defined as a parodied work, or one that is insulting or detrimental to the composer of the commercially released sound recording.⁷¹

This illustrates how the defensive nature of copyright exceptions can be exploited by right-holders and intermediaries to secure the interests of right-holders. It can also further exemplify the danger of self-regulation in this area. Indeed, these agreements are generally reached in a non-disclosure form which removes the opportunity to verify compliance with copyright legislation or fundamental rights.

Secondly, we can identify best practice by looking at YouTube's Content ID system. In Section 2, we mentioned the choice of between using "tracking statistics" as a harsher policy than "monetisation". It is hard to justify these policies. Therefore, self-regulation is likely to create an open-door for intermediaries to choose the harshness factor linked to certain match policies.

⁷¹ See <http://www.musicbusinessworldwide.com/prs-deal-with-youtube-what-does-it-mean-for-songwriters> (accessed 21 March 2018).

This may benefit certain right-holders,⁷² but at the expense of smaller and newer content creators. It may also produce increased bureaucracy for individuals who wish to navigate the platform, and increase the incentive to outsmart the system.

The following guidelines might help. First, users could be given more information about the process of uploading content. While YouTube does offer some guidance, it is somewhat biased. Second, the match policy should reflect the careful balance struck by legislators in the legal instruments between exclusive rights and permissible uses. One of the key difficulties resulting from both proposals relates to the territoriality principle of copyright legislation. A way to mitigate this could be to adapt the interface to identify the location of the infringer, which is likely to coincide with the first territory in which the user enjoys exclusive rights. Third, the territoriality principle could be better respected by identifying the user's upload country as the default option, rather than it being "worldwide", as is currently the practice on YouTube. Fourth, intermediaries should be pushed to acquire more information about who the legitimate right-holder is. Right now, the first claimant of content is presumed to be the right-holder, and there is little opportunity to dispute this. An easy way could be to insert a field where the alleged right-holder could explain why he/she believes they are entitled to enforce his/her copyright prerogatives. Finally, diversity of cultural expressions could be promoted by imposing quotas in the display of search results on the platform.⁷³ Currently in the online world, there remains a fragmentation along national and linguistic lines which results in limited choices for consumers.

⁷² The ones who signed with a publisher or record label and are members of collecting rights societies.

⁷³ Such quotas (or targets) are already present in some member states such as France for broadcasting of content on the radio. This is also supported by European Commission, *The AB Music Working Group Report December 2015- June 2016*, *supra* n. 63, p. 39.

7.4 Facilitating the licensing system

A fair and sustainable digital environment will not be achieved unless progress is made with regards to licensing works in the digital single market. Some national initiatives (like the Copyright Hub⁷⁴ in the UK) or private initiatives (such as ICE)⁷⁵ are welcome, but should be scaled up to the EU level.

With the development of relevant technology, there is a real possibility for intermediaries to include, within their interface, a way for users to get a license for using content. Not only would this benefit the intermediaries by fostering the upload and promotion of UGC, but it would also benefit the creative industries and augment their revenue. An EU initiative similar to the UK Copyright Hub could enable customisation of licenses depending on the intended use. Such a user-friendly licensing system would require an efficient way to deal with license fees and ensure fair remuneration to collecting rights societies and content creators. An EU licensing hub would be of no value if there is no way to assure the transfer of money from the user to the right-holders. Finally, this would create a European database of works, as is currently underway due to the joint efforts of some EU collecting rights societies.

8 Conclusion

Scholars have already warned that the current legal framework, enforcement mechanisms, and market power of certain players create a “cultural blackhole” where the distribution of diverse cultural expression is threatened.⁷⁶ The global

⁷⁴ <http://www.copyrighthub.co.uk> (accessed 21 March 2018).

⁷⁵ <http://www.iceservices.com/about-ice> (accessed 21 March 2018).

⁷⁶ Johnlee Curtis, “Culture and the Digital Copyright Chimera: Assessing the International Regulatory System of the Music Industry in Relation to Cultural Diversity” (2006) 13 *International Journal of Cultural Property* 59-97, p. 64.

dominance of particular creative industries, together with their copyright enforcement strategies and increasing use of AAPs, are driving towards market censorship, thus creating concerns for cultural diversity. Given the lack of a robust legal framework promoting and preserving cultural expression, copyright and freedom of expression are the main instruments for achieving this greater objective. By granting a broad right for individuals to seek, impart, and receive information and copyright, freedom of expression rewards and incentivises creative endeavour. Against increasingly far-reaching copyright laws, legislators have introduced copyright exceptions, such as the parody exception, to ensure the preservation of this specific genre. This recalibration is not reflected in business practices and risks being reliant on the collaboration of private commercial entities, absent robust legal provisions that regulate their “best practices” and ensure due process.

If we recognise that audio-visual content has the power to shape culture by allowing the public to enjoy works in unexpected ways, but also to allow authors and individuals to disseminate their work in a globalised digital environment, then the copyright paradigm must tackle misappropriation and market censorship by implementing fair and efficient independent enforcement mechanisms. Without such a palliative, the agglomeration of culture is likely to be magnified by the ever-increasing reliance on “content recognition technologies” such as Content ID. This was demonstrated in our empirical research which demonstrated that Content ID has an important effect on some categories of content while leaving others unchanged. Here, the algorithm mainly blocked the most popular videos, giving the impression that diversity was therefore increased.

In sum, this article seeks to raise awareness about the need to secure full realisation of fundamental rights in the cultural digital ecosystem in order to uphold the promise of preserving and promoting cultural diversity. Our

approach to the problem of quantifying these hard-to-qualify values relied on two indices: HHI and Eff N. Whilst the HHI is meaningful for assessing variety, balance, and disparity of relevance, the Eff N enables the understanding of the weight of each category. This combination of indices is especially important in an epidemic model where there are important network effects. In this case, as the popularity of a video is self-reinforcing, counting the number of varieties is not sufficient because the public will not be able to access less popular categories. Consequently, instead of regulating online-sharing platforms by requiring them to introduce AAPSs, a better path would be to gain a deeper understanding as to how cultural expressions flow online and how the public access them.

As copyright law is a double-edged sword which can either foster cultural diversity or lead to abuses which endanger the diversity of cultural expression, the debates surrounding the new copyright legal framework should consider introducing vigorous and strict independent dispute resolution mechanisms to mitigate the inevitable reliance on algorithms to tackle large-scale piracy. Furthermore, instead of focusing on *control*, a better avenue – already advocated elsewhere – is to focus on providing adequate *incentives* for future authors to create and to reward the investment made by cultural industries. As a by-product, these balanced copyright and enforcement paradigms will boost competition in the digital single market by concentrating less market power in the hands of some actors (mainly collecting rights societies) to allow new creators to enter the market for cultural goods. Equally, such a balanced system would align better with the interests of consumers and citizens, as copyright will promote both corporate interests and cultural diversity.