

# Privacy and Legal Automation: The DMCA as a Case Study

Jonathon W. Penney\*

22 STAN. TECH. L. REV. 412 (2019)

## ABSTRACT

*Advances in artificial intelligence, machine learning, computing capacity, and big data analytics are creating exciting new possibilities for legal automation. At the same time, these changes pose serious risks for civil liberties and other societal interests. Yet, existing scholarship is narrow, leaving uncertainty on a range of issues, including a glaring lack of systematic empirical work as to how legal automation may impact people's privacy and freedom. This article addresses this gap with an original empirical analysis of the Digital Millennium Copyright Act (DMCA), which today sits at the forefront of algorithmic law due to its automated enforcement of copyright through DMCA notices at mass scale. With literally millions of such notices sent daily, this automation has been criticized for causing large scale chilling effects online, yet few empirical studies have examined this issue in depth. This article does so with a mixed-method empirical study synthesizing survey-based findings with an analysis of 500 Google Blogs and 500 Twitter accounts that have received DMCA notices. The findings offer a number of new insights, including evidence for DMCA notice*

---

\* The author would like to thank Victoria Nash, Urs Gasser, Joss Wright, Ian Brown, David Erdos, Jennifer Urban, Ari Waldman, Eric Goldman, Joe Kariganis, Annemarie Bridy, Niva Elkin-Koren, J. Nathan Matias, Ryan Budish, Kendra Albert, Andy Sellars, Molly Sauter, Derek Bambauer, Ben Zevenbergen, Neil Richards, Daphne Keller, and Jennifer Daskell for comments, advice, and/or suggestions or previous drafts or presentations based thereon. The author would also like to thank participants for his talks at the 2016 Internet Law Work in Progress Conference, Santa Clara Law School, Santa Clara University, March 2016; Berkman Klein Center for Internet & Society, Harvard University, May 2016; Luncheon Speaker Series, Center for Information Technology Policy, Princeton University, March 27, 2018; and Intellectual Property Law Discussion Group, Faculty of Law, University of Oxford, May 8, 2018.

*chilling effects across a range of activities; support for a privacy theory of automated law chilling effects; evidence of differential impacts including that women are disproportionately chilled and that legal information can mitigate chilling effects; and the effectiveness of automated DMCA notices as compared to non-automated ones. This article also explores the implications of these findings for future forms of automated law and lays the foundations for a new theory of governance for personal legal automation.*

#### TABLE OF CONTENTS

I. INTRODUCTION .....	414
II. THEORIZING LEGAL AUTOMATION AND THE DMCA.....	422
A. <i>The Automated Notice Scheme</i> .....	422
B. <i>Personal and Impersonal Automated Legal Enforcement</i> .....	426
C. <i>DMCA Automation and Chilling Effects</i> .....	430
III. METHOD AND DESIGN.....	436
A. <i>Case Study One: DMCA Online Survey</i> .....	436
B. <i>Case Study Two: Google Blogs and Twitter Analysis</i> .....	439
C. <i>Hypotheses / Predictions</i> .....	443
IV. RESULTS AND DISCUSSION.....	445
A. <i>DMCA Impact from Survey Analysis</i> .....	445
1. <i>DMCA Impact on Activities Online</i> .....	445
2. <i>Impact of a “Friend” Receiving DMCA Notice</i> .....	447
3. <i>Predicting Impacts</i> .....	448
4. <i>Legally Challenging DMCA Notices</i> .....	451
B. <i>Impact from Blog and Twitter Analysis</i> .....	453
1. <i>Information on Sample of DMCA Notices</i> .....	453
2. <i>Impact of the DMCA notices</i> .....	455
3. <i>Impact of Automated vs. Non-Automated Notices</i> .....	461
V. IMPLICATIONS .....	463
A. <i>Chilling Effects and Other Impacts</i> .....	463
1. <i>Evidence of Chilling Effects</i> .....	464
2. <i>“Networked”/Indirect Chilling Effects</i> .....	467
B. <i>A Privacy Theory of Automated Law Chilling Effects</i> .....	468
C. <i>Differential Impact</i> .....	470
1. <i>Women Disproportionately Impacted</i> .....	470
2. <i>Knowledge Is Power: Mitigating Chill with Legal Information</i> ..	471
D. <i>The Efficiency of Automated vs. Non-Automated Notices</i> .....	471
E. <i>Micro-Directives and the Future of Legal Automation</i> .....	472
F. <i>Toward a Theory of Personal Legal Automation</i> .....	475
VI. CONCLUSION.....	480
APPENDIX A. <i>BLOGGER SAMPLE: CODING ONLINE/OFFLINE CONTENT STATUS</i> .....	483
APPENDIX B. <i>TWITTER SAMPLE: CODING ONLINE/OFFLINE CONTENT STATUS</i> .....	484

## I. INTRODUCTION

Automated legal enforcement is here,<sup>1</sup> but its future is uncertain and its actual benefits, limits, and impact is unclear.<sup>2</sup> A combination of rapidly developing computing technology along with advances in artificial intelligence (“AI”), machine learning, and availability of big data are creating exciting new possibilities for the automation of law and legal enforcement, with the potential of significant benefits like greater efficiency, scalability, and costs savings.<sup>3</sup> Yet these same possibilities and new legal automation applications—like predictive policing, AI-powered surveillance, or personalized algorithmic legal enforcement at mass scale—raise a whole host of complex law and policy issues, including human rights, transparency, equality, and due process.<sup>4</sup> Despite these

---

1. Woodrow Hartzog et al., *Inefficiently Automated Law Enforcement*, 2015 MICH. ST. L. REV. 1763, 1764 (2015) (“While it may sound like science fiction, the automation of law enforcement is already here.”); Frank Pasquale & Glyn Cashwell, *Four Futures of Legal Automation*, 63 UCLA L. REV. DISCOURSE 26, 36, 39 (2015) (“The automation of law enforcement is already well documented in many fields.”); Lisa A. Shay et al., *Confronting Automated Law Enforcement*, in ROBOT LAW 235, 235 (Ryan Calo et al. eds., 2016); Thomas A. Smith, *From Law to Automation*, 1 CRITERION J. ON INNOVATION 535, 536 (2016); Manuel A. Utset, *Digital Surveillance and Preventive Policing*, 49 CONN. L. REV. 1453 (2017); Benjamin Alarie et al., *Regulation by Machine*, PROC. 29TH CONF. ON NEURAL INFO. PROCESSING SYS. (NIPS), Dec. 5-10, 2016, <https://perma.cc/Z6AF-ZVLX>. On the broader issues raised by automation of legal processes, see also Thomas H. Davenport & Jeanne G. Harris, *Automated Decision Making Comes of Age*, 46 MIT SLOAN MGMT. REV., Jul. 15, 2005, at 84, <https://perma.cc/9ZUF-689A>.

2. Pasquale & Cashwell, *supra* note 1, at 28 (“The future of law and computation is more open ended than most commentators suggest.”); Woodrow Hartzog, *On Questioning Automation*, 48 CUMB. L. REV. 1, 1–2 (2017) (describing the uncertainty of costs and benefits of technologies leveraging AI and algorithms); Anthony Niblett, *Regulatory Reform in Ontario: Machine Learning and Regulation* C.D. HOWE INST. COMMENT. No. 507, Mar. 2018, at 3–4; Frank Pasquale, *A Rule of Persons, Not Machines: The Limits of Legal Automation*, 87 GEO. WASH. L. REV. 1 (2019) (on the limits of legal automation); Frank Pasquale & Glyn Cashwell, *Prediction, Persuasion, and the Jurisprudence of Behaviourism*, 68 U. TORONTO L.J. 63, 63 (2018) (questioning predictions about the future of legal automation); See also FRANK PASQUALE, *THE BLACK BOX SOCIETY* (2015).

3. Hartzog et al., *supra* note 1, at 1765 (“The benefits that robotic technology will bring to law enforcement—particularly in the areas of efficiency and cost savings—are theoretically impressive.”); Hartzog, *supra* note 2, at 1 (“Given the rapid pace of innovation and adoption, it can be hard to make sense of automated technologies”).

4. See, e.g., Hartzog et al., *supra* note 1, at 1765 (“[E]mployment of these technologies without careful consideration poses a distinct danger to our civil liberties and can have detrimental effects on society.”); Pasquale & Cashwell, *supra* note 1, at

challenges and real uncertainties about the costs and benefits of legal automation, there remains a serious lack of guidance in the literature for lawmakers and policymakers.<sup>5</sup>

Part of the challenge is that while legal automation is already here, a lot of concrete debates about its existing applications and benefits, limits, or impact are not. Rather, they exist more in the abstract—arguments about *future* possibilities based on more advanced forms of AI, data, and technology not yet realized. For instance, when Anthony Casey and Anthony Niblett declare the “death” of “rules and standards” in favor of “micro-directives”—a form of highly specific machine enforced legal directions—they are speaking of a law and technological capability “of the future.”<sup>6</sup> This is also the case for Benjamin Alarie, who writes of the “legal singularity” wherein legal uncertainty is rendered “obsolete,” contingent upon the arrival of “massively more data and dramatically improved methods of inference.”<sup>7</sup> Similarly, the legal personalization movement, driven by the availability of large volumes of personal data and predictive algorithms, is the “wave of the future,” observes Cass Sunstein, as new forms of information computational capacity allow the law to be increasingly shaped to people’s specific circumstances.<sup>8</sup> Meanwhile,

---

47–48 (“[C]onflicting rights, unique fact patterns, and open-ended laws will likely remain excessively difficult to automate for an extended period of time. Deregulation may, however, effectively strip many persons of their rights and render once-hard cases simple.”); Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1252–53 (2007); Tal Z. Zarsky, *Automated Prediction: Perception, Law, and Policy*, 55 COMM. ACM, Sept. 2012, at 33–35; Tal Zarsky, *The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making*, 41 SCI., TECH., & HUM. VALUES 118 (2016); Orna Rabinovich-Einy & Ethan Katsh, *A New Relationship between Public and Private Dispute Resolution: Lessons from Online Dispute Resolution*, 32 OHIO ST. J. ON DISP. RESOL. 695, 717 (2017).

5. See Hartzog et al., *supra* note 1, at 1767 (“There is no guiding principle for policy makers and enforcement officers to ensure that automated law enforcement systems fulfill their objective in a way that respects privacy and civil liberties” and sets out to “remedy the dearth of guidance” in the literature).

6. Anthony J. Casey & Anthony Niblett, *The Death of Rules and Standards*, 92 IND. L.J. 1401 (2016).

7. Benjamin Alarie, *The Path of the Law: Towards Legal Singularity*, 66 U. TORONTO L.J. 443 (2016); Benjamin Alarie et al., *Law in the Future*, 66 U. TORONTO L.J. 423 (2016).

8. Cass R. Sunstein, *Deciding by Default*, 162 U. PA. L. REV. 1, 57 (2013) (“[P]ersonalized default rules are the wave of the future. We should expect to see a significant increase in personalization as greater information becomes available about the informed choices of diverse people.”). See also Ariel Porat & Lior Jacob Strahilevitz, *Personalizing Default Rules and Disclosure with Big Data*, 112 MICH. L. REV. 1417 (2014)

skeptics like Frank Pasquale, Elizabeth Joh, or Woodrow Hartzog, among others, either critique these predictive analyses,<sup>9</sup> implicitly accept them and advance solutions or regulatory responses,<sup>10</sup> or counter that the future is more “open ended” than most commentators allow.<sup>11</sup>

The result is that existing scholarship is growing but narrowly focused,<sup>12</sup> leaving a great deal of uncertainty on a range of issues relating to legal automation.<sup>13</sup> In particular, there is a need for more empirical research to understand and explore both the impact and effectiveness of legal automation and its implementation,<sup>14</sup> especially in light of concerns from a range of scholars about how more personalized forms of algorithmic law and automated legal enforcement pose serious risks for civil liberties and human rights, both on individuals and society more generally.<sup>15</sup> Niva Elkin-Koren and Michael Gal, for example, recently raised

---

(arguing for the benefits of personalization, including using automation in certain contexts); Omri Ben-Shahar & Ariel Porat, *Personalizing Negligence Law*, 91 N.Y.U. L. REV. 627 (2016).

9. See, e.g., Pasquale & Cashwell, *supra* note 1 (critiquing predictions as to future legal automation); Pasquale, *A Rule of Persons, Not Machines*, *supra* note 3 (discussing the limits of legal automation).

10. Alarie et al., *supra* note 1, at 1 (noting the work of Hartzog, Elizabeth Joh, Cathy O’Neil, among others, each providing different regulatory solutions. See, e.g., Hartzog et al., *supra* note 1.).

11. Pasquale & Cashwell, *supra* note 1, at 28 (“The future of law and computation is more open ended than most commentators suggest.”).

12. Alarie et al., *supra* note 1, at 1 (“Legal scholars investigating artificial intelligence are preoccupied with regulation. The literature has largely focused on the need for humans to regulate the behavior of automated systems.”).

13. Hartzog, *supra* note 2, at 1 (“People making decisions related to technology law, policy, and ethics have not faced such uncertainty since the advent of the Internet.”).

14. Frank Pasquale, *Professional Judgment in an Era of Artificial Intelligence and Machine Learning*, 46 BOUNDARY 2, 73 (2019) (noting the appropriateness of calling for more empirical research concerning the impact of AI on the legal profession); Peter J. Denning, *Remaining Trouble Spots with Computational Thinking*, 60 COMM. ACM 33, 37–38 (2017) (“[F]or example, physicians, surgeons, psychologists, architects, artists, lawyers, ethicists, realtors, and more. . . . It would be useful to see some studies of how essential computational thinking is in those professions.”); Dru Stevenson & Nicholas J. Wagoner, *Bargaining in the Shadow of Big Data*, 67 FLA. L. REV. 1337, 1352 (2015) (noting that empirical legal studies of law using big data is relatively recent and remains largely narrow).

15. Hartzog et al., *supra* note 1, at 1765 (“[E]mployment of these technologies without careful consideration poses a distinct danger to our civil liberties and can have detrimental effects on society.”); Niva Elkin-Koren & Michal S. Gal, *The Chilling of Governance-by-Data on Data Markets Symposium: Personalized Law*, 86 U. CHI. L. REV.

concerns about how personalized law and similar algorithmic, data-driven, and individual-focused legal and regulatory approaches can have a broader “chilling effects” as well as impacts on civil liberties.<sup>16</sup> Ryan Calo, Lisa Shay, Hartzog, and others have similarly raised privacy and chilling effect concerns about automated or robotic legal enforcement.<sup>17</sup> Yuval Feldman and Yotam Kaplan, after analyzing different forms of personalized legal enforcement, acknowledge similar concerns and the need to “minimize chilling effects” for these “enforcement mechanisms” to be used “successfully.”<sup>18</sup> And Casey and Niblett themselves admit that the personalized machine-enforced law they herald raises serious privacy and autonomy concerns.<sup>19</sup> In short, researchers must shift focus and

---

403 (2019); Primavera De Filippi & Samer Hassan, *Blockchain Technology as a Regulatory Technology: From Code is Law to Law is Code*, 21 FIRST MONDAY, Dec. 5, 2016 (cautioning about “automated legal governance” as it may “reduce the freedoms and autonomy of individuals”); Pasquale & Cashwell, *supra* note 2, at 63 (“The rise of a ‘black box society’ portends profound threats to individual autonomy”); Pasquale, *A Rule of Persons, Not Machines*, *supra* note 2, at 59 (“Software can radically simplify compliance efforts, but when it does so by downplaying, trivializing, or ignoring important aspects of the language of law, it is a betrayal of the rule of law.”); Utset, *supra* note 1, at 1453 (discussing privacy and surveillance concerns of automated legal processes and surveillance); Solon Barocas & Andrew D. Selbst, *Big Data’s Disparate Impact*, 104 CALIF. L. REV. 671 (2016); Elizabeth E. Joh, *The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing Symposium: Policing in America on the 50th Anniversary of Miranda v. Arizona*, 10 HARV. L. & POL’Y REV. 15 (2016). These technologies may also, in the long run, have international law and policy impacts. See generally Jonathon Penney, *The Cycles of Global Telecommunication Censorship and Surveillance*, 36 U. PA. J. INT’L L. 693 (2015) (examining how international law and policy impact, and are impacted by, global technologies).

16. Elkin-Koren & Gal, *supra* note 15, at 1, 4 (writing generally about chilling effects and also noting that “personalized law may undermine important values, raising concerns regarding privacy, equality under the law, and civil liberties”). De Filippi & Hassan raise similar concerns. See De Filippi & Hassan, *supra* note 15 (cautioning about “automated legal governance” as it may “reduce the freedoms and autonomy of individuals”).

17. PATRICK LIN ET AL., *Robots and Privacy*, in ROBOT ETHICS 187 (2011); Shay et al., *supra* note 1, at 30 (noting chilling effects likely caused by automated legal systems); Hartzog et al., *supra* note 1, at 1789–90 (arguing that preserving indeterminacy and inefficiency is a means to address chilling effects).

18. Yuval Feldman & Yotam Kaplan, *Differentiated Regulation Across People and Situations: A Behavioral Ethics Perspective to Personalized Law* (Feb. 15, 2018) (unpublished manuscript).

19. Casey & Niblett, *supra* note 6, at 1405, 1441–45.

interrogate more directly automated law and its practices to better understand the present and future.<sup>20</sup>

This article addresses this gap in the literature with an empirical study on the impact, and potential chilling effects, of perhaps the most high profile automated law today: the Digital Millennium Copyright Act (DMCA).<sup>21</sup> Under the DMCA's notice and takedown scheme, online service providers (OSPs) receive "safe harbor" protection from copyright liability in return for removing allegedly infringing content once they receive copyright removal or notice ("DMCA notice") sent on by copyright holders.<sup>22</sup> To enjoy this safe harbor protection, most large OSPs have thus incorporated measures in their platforms to respond to these DMCA notices.<sup>23</sup> In the last decade, however, the number of DMCA notices sent to OSPs has increased exponentially, largely due to "bots" and automated processes powered by machine learning and algorithms that constantly scan the internet and for infringing content and send on removal requests on detection.<sup>24</sup> Google, for example, deals with approximately 2 million DMCA takedown requests per day and in 2016, removed 900 million links

---

20. Pasquale & Cashwell, *supra* note 1, at 28–29.

21. Digital Millennium Copyright Act of 1998, Pub. L. No. 105-304, § 103, 112 Stat. 2860, 2863–76 (1998) (codified at 17 U.S.C. §§ 1201–1202 (2012)). The "notice and takedown" system can be found in section 512, which was actually enacted by the Online Copyright Infringement Liability Limitation Act (OCILLA) and is now codified as Title II of the DMCA at 17 U.S.C. § 512 (2012). *See also* David Nimmer, *Riff on Fair Use in the Digital Millennium Copyright Act*, 148 U. PA. L. REV. 673 (1999); JESSICA LITMAN, DIGITAL COPYRIGHT 14 (2006); Emily M. Asp, *Section 512 of the Digital Millennium Copyright Act: User Experience and User Frustration Notes*, 103 IOWA L. REV. 751 (2017).

22. Maayan Perel & Niva Elkin-Koren, *Accountability in Algorithmic Copyright Enforcement*, 19 STAN. TECH. L. REV. 473, 481 (2016); David Nimmer, *Puzzles of the Digital Millennium Copyright Act Part I*, 46 J. COPYRIGHT SOC'Y U.S.A. 401, 434–35 (1998); Jane C. Ginsburg, *Copyright 1992-2012: The Most Significant Development*, 23 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 465, 494 (2012); *See also* Niva Elkin-Koren, *Fair Use by Design Melville B. Nimmer Memorial Lecture*, 64 UCLA L. REV. 1082, 1084 (2017) ("Nowadays, the vast majority of copyrighted materials are distributed digitally, and much of copyright enforcement is performed using algorithms."); Jennifer M. Urban et al., *Notice and Takedown in Everyday Practice 7* (UC Berkeley Pub. Law Research Paper No. 2755628, 2017), <https://perma.cc/8BZ7-AVW8>.

23. Perel & Elkin-Koren, *supra* note 22, at 477; Elkin-Koren, *supra* note 22, at 1085; Urban et al., *supra* note 22, at 8.

24. Perel & Elkin-Koren, *supra* note 22, at 477; Elkin-Koren, *supra* note 22, at 1085–86; Urban et al., *supra* note 22, at 8; Michael W. Carroll, *Pinterest and Copyright's Safe Harbors for Internet Providers*, 68 U. MIAMI L. REV. 421, 424 (2014).

to allegedly infringing content.<sup>25</sup> In response, major content platforms have deployed their own automated processes to deal with the large volume of DMCA notices they receive, using algorithms to process and respond to notices and removal requests at scale.<sup>26</sup> As a result, the DMCA sits at the vanguard of what Elkin-Koren calls “algorithmic copyright enforcement,” where robots, automated processes, and machine algorithms make decisions about copyright and content removal.<sup>27</sup>

But critics argue that legal automation under the DMCA has had impact beyond copyright law and its development. Just like micro-directives, personalized law, and other forms of personally focused automated legal enforcement noted earlier, DMCA notices have also long been criticized for potentially having serious chilling effects, impacting people’s autonomy, freedom of expression, and privacy online.<sup>28</sup> And despite automation

---

25. Gina Hall, *How Many Copyright Takedown Notices Does Google Handle Each Day? About 2 Million*, SILICON VALLEY BUS. J. (Mar. 7, 2016), <https://perma.cc/7BML-9RBN>; *Google Asked to Remove 558 Million “Pirate” Links in 2015*, TORRENTFREAK (Dec. 30, 2015), <https://perma.cc/NJ6V-EETC>. These numbers have increased almost exponentially year over year for Google. In 2014 and 2015, those numbers were 558 million and 345 million removal requests. *Id.*; Joe Mullin, *Google Handled 345 Million Copyright Takedowns in 2014*, ARS TECHNICA (Jan. 1, 2015), <https://perma.cc/D4JM-RMSD>. In 2012, Google received 441,370 DMCA notices containing over 54 million content takedown requests, while in 2013, it dealt with 230 million takedown requests. *Google Discarded 21,000,000 Takedown Requests in 2013*, TORRENTFREAK (Dec. 27, 2013), <https://perma.cc/G8YX-MHPL>. See also Perel & Elkin-Koren, *supra* note 22, at 477; Daniel Seng, *The State of the Discordant Union: An Empirical Analysis of DMCA Takedown Notices*, 18 VA. J.L. & TECH. 369, 444, 460–61 (2013). Meanwhile, Facebook reported removing 1.8 million posts or files in response to DMCA notice requests in the first six months of 2016. *Facebook Rejects 31% of All Piracy Takedown Requests*, TORRENTFREAK (Dec. 19, 2017), <https://perma.cc/SZB6-YQH6>. And Reddit reports that its DMCA notice requests increased by 138% in 2017. *Reddit Copyright Complaints Jump 138% but Almost Half Get Rejected*, TORRENTFREAK (Apr. 11, 2018), <https://perma.cc/6UVT-GXHZ>.

26. Perel & Elkin-Koren, *supra* note 22, at 477. Elkin-Koren, *supra* note 22, at 1085–86; Urban et al., *supra* note 22, at 8; Carroll, *supra* note 24, at 424.

27. Elkin-Koren, *supra* note 22, at 1086–89 (speaking to the challenge for fair use in copyright law). For other issues in copyright and automation, see Toni Lester & Dessislava Pachamano, *The Dilemma of False Positives: Making Content ID Algorithms More Conducive to Fostering Innovative Fair Use in Media Creation*, 24 UCLA ENT. L. REV. 51 (2017); Zoe Carpou, Note, *Robots, Pirates, and the Rise of the Automated Takedown Regime: Using the DMCA to Fight Piracy and Protect End-Users*, 39 COLUM. J.L. & ARTS 551 (2016); Rabinovich-Einy & Katsh, *supra* note 4, at 717 (noting algorithms raise “serious questions about the full impact of automated processes on the fairness of such processes, an issue whose implications remain unknown”).

28. Carpou, *supra* note 27, at 585 (“Much literature has been devoted to ways in which automated takedown procedures necessarily result in the chilling of free speech

increasing the number of DMCA notices exponentially in recent years adding a troubling new dimension to these concerns,<sup>29</sup> there are surprisingly few empirical studies on point<sup>30</sup> and none specifically investigate chilling effects and similar impacts. This, too, is consistent with the earlier noted dearth of empirical research on the automation of law and legal enforcement. As a result, skepticism about such chilling effect

---

and should, therefore, be disallowed or at least discouraged by the DMCA"). See, e.g., Asp, *supra* note 21, at 753–54 ("This 'notice and takedown' scheme not only raises questions about the chilling of free speech, it also arguably fails to provide consumers with adequate due process."); RONALD J. DEIBERT, BLACK CODE: INSIDE THE BATTLE FOR CYBERSPACE 229–30 (2013); Wendy Seltzer, *Free Speech Unmoored in Copyright's Safe Harbor: Chilling Effects of the DMCA on the First Amendment*, 24 HARV. J. L. & TECH. 171 (2010); FRED VON LOHMANN, ELECTRONIC FRONTIER FOUNDATION, UNINTENDED CONSEQUENCES: TWELVE YEARS UNDER THE DMCA (2010); JAMES BOYLE, THE PUBLIC DOMAIN: ENCLOSING THE COMMONS OF THE MIND (2008); JONATHAN ZITTRAIN, THE FUTURE OF THE INTERNET—AND HOW TO STOP IT 216 (2008) (exploring the potential chilling effects of perfect enforcement of legal norms by technology measures as well as those caused by citizen surveillance due to the proliferation of devices like smartphones); Jennifer M. Urban & Laura Quilter, Symposium Review, *Efficient Process or Chilling Effects—Takedown Notices Under Section 512 of the Digital Millennium Copyright Act*, 22 SANTA CLARA COMPUTER & HIGH TECH. L.J. 621, 622 (2005); MARJORIE HEINS & TRICIA BECKLES, BRENNAN CTR. FOR JUST. N.Y.U. SCH. L., WILL FAIR USE SURVIVE? FREE EXPRESSION IN THE AGE OF COPYRIGHT CONTROL 24–27, 36 (2005) (discussing the DMCA's chilling effect on classroom experiences, especially in relation to internet use and the web); Yochai Benkler, *Through the Looking Glass: Alice and the Constitutional Foundations of the Public Domain*, 66 LAW & CONTEMP. PROBS. 173, 216–18 (2003) (arguing that the NET Act and Digital Millennium Copyright Act expand protections for certain legal rights online in such a way that will chill expression); Sonia K. Katyal, *The New Surveillance*, 54 CASE W. RES. L. REV. 297, 370 (2003).

29. Urban et al., *supra* note 22, at 8 ("Faced with large-scale infringement, large corporations now use automated 'bots' to search for copyright violations and generate millions of automated 'takedown' notices to OSPs. While this allows some copyright owners to police their copyrights on today's Internet, relying on machines to make decisions about sometimes-nuanced copyright law raises questions about the effect on expression").

30. *Id.* ("Despite the enormous changes since the law was passed, there have been few empirical studies of how notice and takedown actually works in practice"); Seng, *supra* note 26, at 375 (noting the "paucity" of empirical studies on the DMCA's notice and takedown system); Daniel Kiat Boon Seng, *"Who Watches the Watchmen?" An Empirical Analysis of Errors in DMCA Takedown Notices* 7 (SSRN, Working Paper No. 2563202, 2015), <https://perma.cc/XUL7-Q9CU>; Urban & Quilter, *supra* note 28; Heins & Beckles, *supra* note 28. However, there are a few newer and more narrow studies: Kristofer Erickson & Martin Kretschmer, *"This Video is Unavailable": Analyzing Copyright Takedown of User-Generated Content on YouTube*, 9 J. INTELL. PROP. INFO. TECH. & ELECTRONIC COM. L., 75 (2018) (empirical study on the factors that motivate DMCA takedown requests on YouTube); BRUCE BOYDEN, CTR. FOR PROTECTION INTELL. PROP., THE FAILURE OF THE DMCA NOTICE AND TAKEDOWN SYSTEM: A TWENTIETH CENTURY SOLUTION TO A TWENTY-FIRST CENTURY PROBLEM 8 (2013).

claims persists among judges, lawyers, and researchers.<sup>31</sup> Do DMCA notices—today sent by the tens of millions daily at mass scale via automated and algorithmic processes—have chilling effects on people’s online activities, an impact that goes beyond content targeted by notices? If so, are different groups—like women or minorities—disproportionately impacted? And what are the implications, including for how we understand and respond to legal automation today and in the future?

This article sets out to address these and other related questions with a new empirical analysis examining any chilling effects on speech, expression, and other activities online associated with the DMCA’s notice and takedown system. This analysis explores the potential chilling effect caused by the DMCA, but also, by extension, concerns expressed by Elkin-Koren, Calo, Shay, Hartzog, and others, about the impact of more personalized forms of automated legal enforcement today and of the future. The DMCA’s notice and takedown system, I argue, offers a reasonable approximation of such future automated law to understand their impact. The analysis also explores related issues, such as how receiving a DMCA notice may impact certain people (e.g., older, younger, more educated) or certain activities (e.g., expression, sharing, search) differently. Moreover, while the DMCA today is predominantly enforced through automated and algorithmic processes, there are still notices sent manually by people. Comparing the impact of notices sent by automated and non-automated means may also provide insights into legal automation.<sup>32</sup>

This empirical analysis uses two new and first-of-their-kind case studies to triangulate and investigate these questions. The first is an online

---

31. See, e.g., Leslie Kendrick, *Speech, Intent, and the Chilling Effect*, 54 WM. & MARY L. REV. 1633, 1656–57 (2013) (after surveying both scholarship and case law on point, described how the empirical basis for such chilling effect concerns and claims were “weak” and “flimsy” and concluding additional research was required for the “unsubstantiated empirical judgments” of chilling effects claims). See also Jonathon W. Penney, *Chilling Effects: Online Surveillance and Wikipedia Use*, 31 BERKELEY TECH. L.J. 117, 120–21 (2016) (describing long time skepticism from judges, lawyers, legal scholars, and researchers, about chilling effect claims in law).

32. Jake Goldenfein, *The Future of Automated Privacy Enforcement*, in TRANS-ATLANTIC DATA PRIVACY RELATIONS AS A CHALLENGE FOR DEMOCRACY 507, 518 (Dan Svantesson & Dariusz Kloza eds., 2017) (“This then raises further questions around whether purely automated action has the same impact on individuals as action mediated by human intelligence and agency—a question that requires very serious consideration at a general level”).

survey examining internet user responses to hypothetical scenarios involving DMCA notices and analyzes, explores, and compares responses. The survey investigates how the hypothetical DMCA scenario impacts the participant's online activities in a range of contexts, including online speech, sharing, search, and privacy concerns. It also explores user's willingness to challenge the DMCA notice received in the scenario. This second case study explores DMCA impacts by examining 500 Google Blogger accounts and 500 Twitter accounts that have received actual DMCA notices.

Part II of the Article provides theoretical context related to legal automation under the DMCA and its impact. It also sets out and explains the DMCA's notice and takedown scheme. Part III sets out the overall research design and methodology for this empirical analysis, and Part IV sets out and discusses results of the study. Implications are discussed in Part V. I conclude in Part VII, and discuss some important limitations in the studies herein.

## II. THEORIZING LEGAL AUTOMATION AND THE DMCA

### A. *The Automated Notice Scheme*

When the DMCA was enacted by the United States Congress in 1998, few of its framers likely predicted its central enforcement scheme—the “notice and takedown” system—would prove so influential globally, serving as a model for similar legislation around the world.<sup>33</sup> Even fewer, it can be surmised, would have foreseen its position, twenty years later, at the “forefront” of algorithmic law enforcement.<sup>34</sup> And yet in many ways,

---

33. Julie E. Cohen, *Pervasively Distributed Copyright Enforcement*, 95 GEO. L.J. 1, 16 (2006) (“The DMCA was enacted as part of U.S. accession to a 1996 treaty that requires effective legal protection for technological measures applied to copyrighted works, and has served as a model for implementing legislation in other countries”); Urban et al., *supra* note 22, at 7 (“In the eighteen years since it was passed, the ‘notice and takedown’ system established by section 512 . . . has become a primary tool for raising and resolving copyright disputes in the United States, and has served as a model for other countries.”); Matthew Rimmer, *Back to the Future: The Digital Millennium Copyright Act and the Trans-Pacific Partnership*, 6 LAWS 2 (2017) (describing American trade efforts to promote the DMCA abroad). *See also* LAWRENCE LESSIG, CODE, at 117 (version 2.0 2006) (noting that the framers of the DMCA likely did not envision some of its provisions or applications.).

34. Perel & Elkin-Koren, *supra* note 22, at 476–77 (“Copyright law was at the

the DMCA's unique notice and takedown system is precisely the kind of regulatory scheme contemplated by advocates of greater legal automation today as it largely uses decentralized algorithmic enforcement of legal claims replaces "inefficient" and "costly" traditional mechanisms of legal enforcement like retained lawyers or a centralized state bureaucracy with decentralized algorithmic enforcement of legal claims.<sup>35</sup> Moreover, the DMCA, in the form of its DMCA notices, also employs a form of what I call *personal* and *personalized* legal automation and does so at mass scale, another benefit often advanced.<sup>36</sup> This point requires elaboration, including further detailing of the DMCA's notice and takedown system itself.

When content is uploaded or posted by users to an OSP like Twitter, YouTube, or Google Blogger, these OSPs constitute a "service provider" under the DMCA.<sup>37</sup> Service providers provide internet services like social media platforms, hosting, and linking. In order for service providers to remain in the DMCA's "safe harbor" and avoid liability for the actions of their users is their ignorance of infringing acts: an OSP must "not have actual knowledge"<sup>38</sup> of copyright infringement on its network, servers, or

---

forefront of algorithmic law enforcement beginning in the early 1990s, conferring safe harbor protection to online intermediaries who removed allegedly infringing content upon notice under the Notice and Takedown procedure designed by the Digital Millennium Copyright Act."); Elkin-Koren, *supra* note 22, at 1084.

35. See, e.g., John O. McGinnis & Russell G. Pearce, *The Great Disruption: How Machine Intelligence Will Transform the Role of Lawyers in the Delivery of Legal Services Colloquium: The Legal Profession's Monopoly on the Practice of Law*, 82 FORDHAM L. REV. 3041, 3041–42 (2014); Bruce H. Kobayashi & Larry E. Ribstein, *Law's Information Revolution*, 53 ARIZ. L. REV. 1169, 1170 (2011) (advocating for the disruption of "legacy" legal service providers by technology corporations). The DMCA's scheme, of course, does not fully eliminate the need for lawyers, but the vast majority of copyright claims made under the DMCA's authority are today determined by automated processes.

36. Goldenfein, *supra* note 32, at 508 ("As these techniques and practices become more sophisticated and automated, greater reliance is placed on information processing such as data mining, predictive analytics and other artificial intelligence techniques, deployed at mass-scale, to detect patterns 'hidden in the data' for the purpose of flagging or identifying individuals as suspicious reasonableness of their actions or search for the content of specific law.").

37. 17 U.S.C. § 512(k)(1) (2012); Jeffrey Cobia, *The Digital Millennium Copyright Act Takedown Notice Procedure: Misuses, Abuses, and Shortcomings of the Process Note*, 10 MINN. J.L. SCI. & TECH. 387, 390 (2009); Charles S. Wright, *Actual Versus Legal Control: Reading Vicarious Liability for Copyright Infringement into the Digital Millennium Copyright Act of 1998*, 75 WASH. L. REV. 1005, 1024 (2000).

38. 17 U.S.C. § 512(c)(1)(A)(i) (2012).

system, and if it does obtain such knowledge, then it must act “expeditiously” to remove or disable access to the infringing material.<sup>39</sup> This is where DMCA “takedown notices” (DMCA notices) come into the scheme:<sup>40</sup>

- If a copyright holder discovers her work has been copied and uploaded to a blog or website without permission, she can send a takedown “notification” or notice to notify the website or OSP.<sup>41</sup>
- The DMCA notice must conform to certain statutory requirements or it is “void.”<sup>42</sup> Once an OSP has received the takedown notice it must act quickly or “expeditiously” to remove or disable access to the infringing content.<sup>43</sup>
- Importantly, the OSP must also “promptly” notify the user whose content has been removed or disabled about receiving the DMCA notice, which often involves providing a copy of the DMCA notice or information therein to the targeted user.<sup>44</sup> For example, Google Blogger sends a message to a blogger on its service indicating it has received a DMCA notice and has disabled access to the targeted content, as well as information to view the DMCA notice at the Lumen Database.<sup>45</sup> This notification by the OSP is

---

39. 17 U.S.C. § 512(c)(1)(A)(iii)(2012); Cobia, *supra* note 37, at 393–94.

40. For a corresponding exposition of the scheme, see also Urban et al., *supra* note 22, at 15–17.

41. 17 U.S.C. § 512(c)-(d) (2012).

42. Generally, the notice must be a “written communication” and it should contain: (i) a signature (digital or otherwise) of the person acting on behalf of the copyright owner; (ii) identification of copyrighted work that is allegedly infringed (like a URL to the work’s location); (iii) identification of allegedly infringing content that should be removed or access to which should be disabled (often a URL if content is on the web); (iv) contact information for the complainant; (v) a statement as to the complainant’s good faith belief and accuracy of the notice’s contents and claims. See 17 U.S.C. § 512(c)(3)(A) (2012).

43. 17 U.S.C. § 512(c)-(d) (2012).

44. 17 U.S.C. § 512(g) (2012).

45. OSPs like Google and Twitter, as a matter of corporate policy, send DMCA notices they receive and act upon to the Lumen Database (formerly the “Chilling Effects Repository”) hosted by the Berkman Klein Center for Internet & Society at Harvard University. LUMEN DATABASE, <https://lumendatabase.org/>; see *infra* note 119 for policies. For additional background, see David F. Gallagher, *New Economy; A Copyright Dispute with the Church of Scientology is Forcing Google to Do Some Creative Linking*, N.Y. TIMES (Apr. 22, 2002), <https://perma.cc/LXY7-RHBK> (providing insight into the origins of Google’s policy to share DMCA notices with the Chilling Effects Database).

mandatory because it gives the user (here, the blogger) a chance to send a “counter-notice”—which also statutorily requires certain information—back to the OSP denying that content or links posted infringed copyright.<sup>46</sup>

- If no counter-notice is sent by the user, the content remains removed or disabled.<sup>47</sup> If a counter-notice is sent, the OSP must notify the complainer (who sent the original takedown notice) and replace or re-enable the removed content 10-14 days *after* the counter-notice is received. (Thus, material is inevitably offline for at least 10 days whether or not a counter-notice is filed).<sup>48</sup>

If the OSP does all of this, it is protected by Section 512’s safe harbor from lawsuits both by copyright holders *and* the users whose allegedly infringing content was removed.<sup>49</sup> It should be also noted that if a counter-notice is filed, the original complainer (who sent the takedown notice) may then file an action in court seeking to restrain the infringing activity before the content is replaced by the OSP,<sup>50</sup> though they need not wait for that point to bring such a court action—they can do so at any time. DMCA notices are sometimes sent directly to users but most often are sent to “service providers” (OSPs), counter-notices are very rare, and there have been only a few lawsuits filed after the initial notice-and-takedown

---

46. 17 U.S.C. § 512(g)(3) (2012).

47. It should be noted that on some online services like Google Blogger, the user may have an opportunity to edit their blog post, remove any infringing material, and re-post. Brett Wiltshire, *New Tools for Handling Copyright on Blogger*, GOOGLE BLOG (July 25, 2011), <https://perma.cc/9RPF-PD6C> (“At this point, Joe Blogger has the right to file a counter-notice and request the post be restored if he believes the takedown was improper. He may also edit the post to remove the allegedly infringing content himself.”). A recent example of a notification sent to a Blogger user, which indicates Google has placed the targeted blog post on “draft” so the user can re-post without the alleged infringing content. *Blogger Blog Take-Down Notification for Copyright Content*, GOOGLE BLOGGER HELP (Oct. 25, 2017), <https://perma.cc/N523-ZBYN> (here, a Blogger user posts the DMCA notification he has received and Google Blogger tech support responds and explains why he has received the notice).

48. 17 U.S.C. § 512(g)(2) (2012).

49. 17 U.S.C. § 512(c)-(d), (g) (2012); Cobia, *supra* note 37, at 393–94; Ginsburg, *supra* note 22, at 494; Nimmer, *supra* note 23, at 434; LAURA QUILTER & MARJORIE HEINS, BRENNAN CTR. FOR JUST. N.Y.U. SCH. L., *INTELLECTUAL PROPERTY AND FREE SPEECH IN THE ONLINE WORLD* 49 (2007).

50. 17 U.S.C. § 512(g)(2)(C) (2012).

procedure.<sup>51</sup> In other words, most copyright claims raised through the notice and takedown system are settled outside the judicial process.

This is, essentially, the DMCA's notice and takedown system, as designed by its framers. Today, however, it has become increasingly an automated law. The volume of DMCA notices sent to OSPs has increased exponentially, largely due to algorithmic and automated processes scanning the internet for infringing content and sending removal requests upon detection.<sup>52</sup> Google deals with two million DMCA takedown requests per day and removed 900 million links to allegedly infringing content in 2016.<sup>53</sup> To deal with this volume, major OSPs now automated their responses to DMCA notices, using algorithms to process notices and remove or disable targeted content at mass scale.<sup>54</sup> The DMCA's notice and takedown scheme is thus predominantly "algorithmic copyright enforcement".<sup>55</sup> This has led to substantial scholarship critiquing the automation of notices under the DMCA.<sup>56</sup>

### *B. Personal and Impersonal Automated Legal Enforcement*

However, there remains little systematic theoretical or empirical work done to understand the DMCA's automation and its broader impact.<sup>57</sup> Here, I am interested in DMCA notices—which are automated at mass-scale—that target individual users, and may affect their activities beyond copyright law. Automated legal enforcement, defined as "any computer-

---

51. Urban et al., *supra* note 22, at 95 (noting their study "and other research on this issue consistently shows that counter notices are rarely used"). See also Urban & Quilter, *supra* note 28, at 679; Cobia, *supra* note 37, at 391.

52. Perel & Elkin-Koren, *supra* note 22, at 477; Elkin-Koren, *supra* note 22, at 1085–86; Urban et al., *supra* note 22, at 8.

53. See works cited *supra* note 24.

54. Perel & Elkin-Koren, *supra* note 22, at 477; Elkin-Koren, *supra* note 22, at 1085–86; Urban et al., *supra* note 22, at 8.

55. See Elkin-Koren, *supra* note 22, at 1086–89 (speaking to the challenge for fair use in copyright law). For other issues in copyright and automation, see generally, Lester & Pachamano, *supra* note 27; Carpou, *supra* note 27; Rabinovich-Einy & Katsh, *supra* note 4, at 717 (noting algorithms raise "serious questions about the full impact of automated processes on the fairness of such processes, an issue whose implications remain unknown").

56. Carpou, *supra* note 27, at 585. See also works cited *supra* note 27.

57. Hartzog et al., *supra* note 1, at 1767. The work of Perel and Elkin-Koren is an exception. See Perel & Elkin-Koren, *supra* note 22.

based system that uses input from unattended sensors” to “algorithmically” determine, detect, surveil, or take responsive action for, legal issues or infringements of legal rights and interests, can take a diverse range of forms in civil and criminal law systems.<sup>58</sup> Hartzog et al., for instance, offer a taxonomy that includes “three” major actors (the subject of law enforcement; law enforcement agencies; and the judicial system) and three different forms of legal automation that includes (1) surveillance/detection; (2) analysis (of the legal issues); (3) action (in response to a legal determination, and with impact on subjects or persons; entities or persons conducting surveillance, detection, and enforcement; and determinations of legality or innocence and guilty.<sup>59</sup> This taxonomy revised an earlier one, with similar components, but also spoke to aggregation as separate component of automation.<sup>60</sup>

Yet, this taxonomy could be even more granular, delving more specifically into each of those components—for instance, precisely *how* an automated or algorithmic process is deployed in the detection or enforcement process.<sup>61</sup> Detection and enforcement could simply assist non-automated actors or other processes to contact targeted subjects or persons. This would be a form of what I call *impersonal* legal automation—a role for automation and algorithms in the legal process that helps facilitate, operate, propel, or implement aspects of the system, but never directly touches, impacts, or interacts with subjects or persons. On the other hand, automated processes could be deployed to directly interact with, and give legal directions to, subjects or persons. I call this *personal* legal automation, as it involves automated processes having personal contact and interactions with people in their legal functions.<sup>62</sup>

---

58. Hartzog et al., *supra* note 1, at 1768.

59. *Id.* at 1768–69.

60. Shay et al., *supra* note 1, at 9–22.

61. In fairness to Shay, Hartzog, and their co-authors, they do delve deeper—including examining time, location, tracking, velocity, and identification, among other aspects—just not specifically on these counts. *See id.*; Hartzog et al., *supra* note 1, at 1768–69.

62. This is a slightly different distinction to that drawn by advocates of legal personalization of like Ariel Porat and Lior Strahilevitz, who use “impersonal” law to refer to more general legal standards like an enacted statute and “personal” to refer to law that is not general but customized to the specific circumstances of a person. *See* Porat & Strahilevitz, *supra* note 8, at 1418–22.

This distinction is important for a few reasons. First, whether machines or automated processes directly interact with people can have important legal, ethical, and policy implications.<sup>63</sup> For example, Calo argues that as automation and machines become more integrated in provision of services like health, direct interactions between people and automated service providers be an important factor in legal liability in the event of health harms or medical malpractice.<sup>64</sup> There is also a wealth of research on how people perceive threats and other forms of interactions with automated or robotic systems differently than human or non-automated kinds.<sup>65</sup> Kate Darling has argued for laws providing legal protections for these machines based on similar findings on how humans tend to perceive anthropomorphic machines.<sup>66</sup> And Hartzog has argued for specialized consumer protection laws taking these realities into account.<sup>67</sup> Whether or not automated processes interact directly and personally with people has noteworthy legal, ethical, and policy implications, and may require a unique legal or regulatory response.

Another reason the distinction between personal and impersonal legal automation is important is that personal forms of legal automation are precisely the kind often heralded as key to more advanced automated law of the future. Advocates of legal personalization argue that using personal information and forms of data to tailor the law to a person's circumstances is superior to what they call impersonal law—like general rules and

---

63. Goldenfein, *supra* note 32, at 518 (“This then raises further questions around whether purely automated action has the same impact on individuals as action mediated by human intelligence and agency—a question that requires very serious consideration at a general level.”); Hartzog, *supra* note 2, at 1 (“Virtual assistants can interact with us as though they were human. But what type of relationship is appropriate to form with them?”); *see also* Woodrow Hartzog, *Unfair and Deceptive Robots Focus on Cyberlaw*, 74 MD. L. REV. 785, 787–89 (2015) (noting that robots “raise common consumer protection issues”); Christina Mulligan, *Revenge Against Robots*, 69 S.C. L. REV. 579 (2018); Kate Darling, *Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior Towards Robotic Objects*, WE ROBOT CONF. (Apr. 21–22, 2012), <https://perma.cc/3EFR-N3SU>.

64. Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. 513, 547 (2015).

65. *Id.* at 547–48.

66. Darling, *supra* note 63.

67. Hartzog, *supra* note 63, at 787–89 (proposing an approach, via the Federal Trade Commission, to address consumer protection issues raised by robots).

standards set by a general statute.<sup>68</sup> Advocates advance many reasons for this, but one is it can provide greater legal certainty and guidance to people when personalized legal information or notice is directly and personally communicated.<sup>69</sup> Ariel Porat and Lior Strahilevitz, for example, argue that advances in big data analytics and other technologies will eventually allow for personalized legal disclosures and determinations to be directly communicated to consumers via software or other interfaces.<sup>70</sup> Similarly, Caryn Devins and colleagues argue that the “personalized law model” can help “automate” the application of general law to specific circumstances, and suggests software apps delivering “simple directives” for consumers to comply with the law.<sup>71</sup> All of these personalized law forms contemplate personal legal automation.

In fact, there is a body of personalized law literature specifically devoted to exploring how the state can “personalize directives to citizens”.<sup>72</sup> One prominent recent example is Casey and Niblett’s innovative work on “micro-directives”<sup>73</sup> a prototypical example of personal legal automation. Micro-directives, they argue, will involve machines essentially enforcing the law through highly context specific legal directives determined on an ongoing basis by sophisticated

---

68. Sunstein, *supra* note 8, at 57; Porat & Strahilevitz, *supra* note 8; Ben-Shahar & Porat, *supra* note 8.

69. Andrew Verstein, *Privatizing Personalized Law Symposium: Personalized Law*, 86 U. CHI. L. REV. 551, 563–64 (2019) (“The great advantage of personalized law is that it may be better at guiding and informing individuals about their obligations. Where technology permits micro-directives, lawmakers can give certainty directly to individuals, instructing them on the law’s content given their exact circumstances.”); Caryn Devins et al., *The Law and Big Data*, 27 CORNELL J.L. & PUB. POL’Y 357, 367 (2017) (“Private technology such as software apps could also provide simple directives for legal consumers to comply with the law without having to weigh the reasonableness of their actions or search for the content of specific law.”).

70. See generally Porat & Strahilevitz, *supra* note 8; Ben-Shahar & Porat, *supra* note 8.

71. Devins et al., *supra* note 69, at 367.

72. Verstein, *supra* note 69, at 552 (“Some papers have focused on ways in which the state could personalize its directives to citizens.”); Devins et al., *supra* note 69, at 367 (“Private technology such as software apps could also provide simple directives for legal consumers to comply with the law without having to weigh the reasonableness of their actions or search for the content of specific law.”). See, e.g., Sunstein, *supra* note 8, at 7–10, 30; Porat & Strahilevitz, *supra* note 8, at 1442–50.

73. Casey & Niblett, *supra* note 6, at 1401.

algorithms and delivered directly and personally to people.<sup>74</sup> This new legal possibility is contingent upon development of two forms of technology, both predictive technology and communication technology.<sup>75</sup> Predictive technology will be driven by big data access and advances in big data computational capacity, to allow for fast and nuanced legal determinations specific to a given context.<sup>76</sup> Importantly, they argue, advances in *communication* technology will allow these micro-directives—highly specific legal determinations and directives—to be directly and personally communicated to persons.<sup>77</sup> The typical micro-directive example is a car dashboard informing a driver of the legal speeding limit. This form of personal legal automation is predicted to be the law’s future.<sup>78</sup> The future, therefore, is a move away from impersonal non-automated law, to personal *and* personalized automated law enforced at mass or population-wide scale.<sup>79</sup>

### C. DMCA Automation and Chilling Effects

The DMCA employs precisely this form of legal automation: *personal* and *personalized* legal automation at mass scale,<sup>80</sup> via a notice and

---

74. *Id.* at 1404 (“With micro-directives, however, the law looks quite different. The legislature merely states its goal. Machines then design the law as a vast catalog of context-specific rules to optimize that goal. From this catalog, a specific micro-directive is selected and communicated”). One important difference is that Casey and Niblett envision micro-directives to be *ex ante* behavioral directives, that is, they provide a legal directive for specific situations and it is then up to the recipient to decide how to respond. DMCA notices are *ex post facto*, to the extent that they are a response to an alleged act of copyright infringement. Nevertheless, the DMCA notice also doubles as a legal directive (or threat), for the recipient to no longer infringe the targeted copyright material.

75. *Id.* at 1410.

76. *Id.*

77. *Id.*

78. *Id.*

79. *See, e.g.,* Goldenfein, *supra* note 32, at 508 (“As these techniques and practices become more sophisticated and automated, greater reliance is placed on information processing such as data mining, predictive analytics and other artificial intelligence techniques, deployed at mass-scale, to detect patterns ‘hidden in the data’ for the purpose of flagging or identifying individuals as suspicious.”); Pasquale, *A Rule of Persons, Not Machines*, *supra* note 2, at 4.

80. *See, e.g.,* Goldenfein, *supra* note 32, at 508; Pasquale, *A Rule of Persons, Not Machines*, *supra* note 2, at 4.

takedown system. In contrast to impersonal statutory laws, each of the millions of automated DMCA notices sent out daily are form of *personalized law* as they constitute specific legal directives based on personalized circumstances of the sender, the recipient, and the specific alleged infringing actions in relation to specific copyrighted material. Notices are also *personal* legal automation as their legal directives are, if acted on by OSPs, personally received by the user, either directly or via the OSP. The DMCA notice is arguably a precursor to Casey and Niblett's notion of "micro-directives,"<sup>81</sup> although not as personalized as the micro-directives they describe.

Like micro-directives, each automated DMCA notice, once acted on by the OSP and received by the targeted user, acts as a context specific legal directive (to cease infringing a claimed legal right in relation to specific content and a specific use). That legal directive is pre-determined by automated processes and algorithms, deployed both by the legal claimant as well as OSP, based on the specifics of the targeted content and user.

Given the personal and personalized nature of the DMCA's legal automation, it is not surprising there is a growing body of legal scholarship criticizing the DMCA for causing serious chilling effects, impacting freedom of speech, expression, privacy, and autonomy online.<sup>82</sup> The idea that laws, state actions, or legal processes might have such a chilling effect was first comprehensively explored by Frederick Schauer, with his account of chilling effects law and theory often described as the "definitive treatment."<sup>83</sup> He conceived of chilling effects as mainly due to people fearing legal prosecution, sanction, or liability, coupled with uncertainties in the legal system; and, to be specific, his account is concerned with chilling of legally activities like First Amendment protected speech.<sup>84</sup> Thus,

---

81. Casey & Niblett, *supra* note 6, at 1401.

82. Carpou, *supra* note 27, at 585 ("Much literature has been devoted to ways in which automated takedown procedures necessarily result in the chilling of free speech and should, therefore, be disallowed or at least discouraged by the DMCA."). *See also* works cited *supra* note 27.

83. Julie E. Cohen, *Right to Read Anonymously: A Closer Look at Copyright Management in Cyberspace*, 28 CONN. L. REV. 981, 1011 n.117 (1996) (speaking of Schauer's working as leading and definitive). The work to which Cohen was referring was his 1978 article. Frederick Schauer, *Fear, Risk and the First Amendment: Unraveling the Chilling Effect*, 58 B.U. L. REV. 685, 730 (1978).

84. Schauer, *supra* note 83, at 687–89.

many of the examples Schauer discusses involve vague or uncertain laws impacting free expression, where legal outcomes are unclear.<sup>85</sup> Here, due to the costs and uncertainty in defending oneself in the legal system, a person may avoid saying or doing something entirely legal out of the fear that the act may lead to legal penalty or liability.

Daniel Solove built upon Schauer's work by theorizing how modern information practices and data gathering also create a kind of broader regulatory "environmental pollution" encouraging chilling effects and self-censorship.<sup>86</sup> Here, people are chilled not because they fear actual legal punishment, but to avoid other kinds of risks and harms, such as the social stigma of being labeled and tracked by the state as a non-conformist, threat, or criminal, or the risk that information gathered about them may be leaked, shared, disclosed, or misappropriated, leading to reputational, economic, or other types of harms.<sup>87</sup> Though some like scholars like Anne-Marie Bridy, Paul Ohm, and Sonya Katyal have highlighted the privacy implications of copyright surveillance, the DMCA's chilling effects have predominantly been theorized not in privacy terms, but through Schauer's framing: how the vagueness and legal uncertainties involved with copyright law and doctrine chills legal and legitimate forms of speech and expression.<sup>88</sup>

---

85. *Id.* at 705–25.

86. Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. PA. L. REV. 477, 488 (2006); Neil M. Richards, *The Dangers of Surveillance Symposium: Privacy and Technology*, 126 HARV. L. REV. 1934, 1949–50 (2012); NEIL RICHARDS, *INTELLECTUAL PRIVACY: RETHINKING CIVIL LIBERTIES IN THE DIGITAL AGE* 107 (2015).

87. Solove, *supra* note 86, at 496 (discussing the example of how information obtained by surveillance was used to discredit and blackmail Martin Luther King, Jr.).

88. For the typical framing, see Seltzer, *supra* note 28, at 178, 187 (citing Melville Nimmer on the "balance" and "conflict" at the heart of copyright law); Neil Weinstock Netanel, *Copyright and a Democratic Civil Society*, 106 YALE L.J. 283, 381, 385 (1996) (describing uncertainty in copyright law and how the long often involves balancing the interests of user rights, copyright holders, and the public interest); see also R. Polk Wagner, *The Perfect Storm: Intellectual Property and Public Values Symposium: Law and the Information Society: Panel I: Intellectual Property and Public Values*, 74 FORDHAM L. REV. 423, 434 (2005). See also works cited *supra* notes 25–26. For examples of works speaking to privacy impact of copyright enforcement, see, e.g., Annemarie Bridy, *Graduated Response and the Turn to Private Ordering in Online Copyright Enforcement*, 89 OR. L. REV. 81, 96 (2010); Paul Ohm, *The Rise and Fall of Invasive ISP Surveillance*, 2009 U. ILL. L. REV. 1417, 1432 (2009); Sonia K. Katyal, *Filtering, Piracy Surveillance and Disobedience*, 32 COLUM. J.L. & ARTS 401, 402 (2008).

Through the lens of this work, it is easy to see how DMCA notices may cause chilling effects. First, copyright law itself is rife with uncertainty, often balancing competing interests, and the DMCA's scheme requiring "bright line" decisions about legal and impermissible forms of expression exacerbates this problem in practice.<sup>89</sup> Second, the legislative scheme "describing the procedure is vague and filled with flaws,"<sup>90</sup> ranging from the provisions governing the content required for DMCA notices to be valid,<sup>91</sup> to the definition of "service provider"<sup>92</sup> requirement that OSPs "take reasonable steps to promptly notify" users whose content has been removed or disabled, to what an OSP but communicate to users upon receiving a DMCA notice.<sup>93</sup> Undefined and vague terms like this pervade Section 512 rendering its application from uncertain and unclear.

Perhaps most problematic, there is little guidance as to the role of the "fair use" copyright defense in the notice and takedown system. Originally a judicially created legal doctrine, but now recognized in Section 107 of the U.S. Copyright Act,<sup>94</sup> fair use is the "most important" defense to the copyright infringement and is an essential element of copyright law because it renders expansive copyrights both "possible and bearable."<sup>95</sup> A key public interest served by fair use is freedom of expression.<sup>96</sup> There is no "bright line rule" for fair use; these factors must be applied on a case-by-case basis.<sup>97</sup> Burk and Cohen, for example, call it an essential "safety

---

89. Seltzer, *supra* note 28, at 178.

90. Asp, *supra* note 21, at 769.

91. 17 U.S.C. § 512(g)(3)(A) (2012). *See also* Nimmer, *supra* note 21, at 711 (noting the statute only says notices must contain "substantial[ly]" the specified elements, creating great "leeway" as to what may qualify).

92. Jane Ginsburg described it as "exceedingly vague." Ginsburg, *supra* note 22, at 494.

93. Asp, *supra* note 21, at 769–70.

94. 17 U.S.C. § 107 (2012).

95. Barton Beebe, *An Empirical Study of U.S. Copyright Fair Use Opinions, 1978–2005*, 156 U. PA. L. REV. 549, 551 (2007).

96. *Id.*; Seltzer, *supra* note 28, at 187 (noting the importance of fair use to the balance at the heart of copyright law); BOYLE, *supra* note 28, at 50; Dan L. Burk & Julie E. Cohen, *Fair Use Infrastructure for Rights Management Systems*, 15 HARV. J. L. TECH 41, 43–44 (2001).

97. *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 577 (1994) ("The task [of fair use analysis] is not to be simplified with bright-line rules, for the statute, like the doctrine it recognizes, calls for case-by-case analysis..."); Seltzer, *supra* note 28, at 178.

valve” between expansive copyrights and freedom of expression.<sup>98</sup> Copyright protection is a restriction on freedom of expression as it prevents people from using copyrighted materials for their own speech and expression without permission. Fair use ameliorates these restrictions, at least in part, by allowing certain uses of copyrighted works, “such as” criticism, comment, news reporting, teaching, scholarship, research, that would otherwise require permission. This balances copyright restrictions with the “public interest in open dialogue, deliberation, and the advance of knowledge.”<sup>99</sup>

Yet the way in which fair use is considered in the DMCA notice and takedown scheme is vague and unclear.<sup>100</sup> Though some OSPs like Google consider potential fair use issues on reviewing DMCA takedown notices it has received,<sup>101</sup> there is no explicit requirement to do so. This is because parties sending DMCA notices are legally required to consider fair use before sending, but this does not always happen; both good and bad faith mistakes about fair use occur.<sup>102</sup> And there are few safeguards in the DMCA to detect, penalize, or punish abuses of the notice and takedown system, despite anecdotal and high profile cases suggesting abuse and overreach.<sup>103</sup>

Based on these problems and other concerns, commentators, legal scholars, and researchers have criticized the DMCA notice and takedown

---

98. Burk & Cohen, *supra* note 96, at 43.

99. *Id.*

100. Seltzer, *supra* note 28, at 172–73, 178, 203, 210 (identifying a DMCA case where fair use was completely neglected, and noting, “Additionally, many scenarios simply fall outside the core of copyright’s policy justifications, and others are too close to the edge between infringement and fair use to be decided accurately by the summary procedures of a service provider reviewing a § 512[c] notice.”).

101. See *Google Search Removals Due to Copyright Infringement FAQs*, GOOGLE TRANSPARENCY REP. HELP CTR., <https://perma.cc/BX6M-VAS9> (“It is our policy to respond to clear and specific notices of alleged copyright infringement. Upon review, we may discover that one or more URLs specified in a copyright removal request clearly did not infringe copyrights. In those cases we will decline to remove those URLs from Search. Reasons we may decline to remove URLs include not having in enough information about why the URL is allegedly infringing; not finding the allegedly infringing content referenced in the request; deducing that the copyright removal process is being used improperly . . . or fair use.”).

102. Seltzer, *supra* note 28, at 178; Ira Steven Nathenson, *Looking for Fair Use in the DMCA’s Safety Dance*, 3 AKRON INTELL. PROP. J. 121, 143, 147 (2009) (discussing bad faith in the *Lenz* decision).

103. Seltzer, *supra* note 28, at 178.

system for creating chilling effects on people's activities online.<sup>104</sup> Drawing on Schauer's work, a simple way of understanding these concerns is that DMCA notices constitute a clear legal threat (or legal directive to users who receive notices directly or are notified via the OSP: to remove or take down the copyright infringing content targeted by the notice. The threat could be the potential for a lawsuit that could lead to court-imposed civil or criminal penalties under the Copyright Act, but due to the DMCA's vague scheme, whether such litigation could be successfully opposed is quite uncertain. The threat could also be the subpoena powers given to copyright holders in the DMCA, which allows them to potentially unmask anonymous activities online.<sup>105</sup> So such communicated legal claims, threats, or directives in DMCA notices are analogous to libel and defamation claims that have, in more traditional media settings, been shown to "chill" or deter certain media practices, publishing, or content coverage.<sup>106</sup>

Chilling effects can be theorized by this account of legal uncertainties and harms, but a privacy-focused theory based on surveillance concerns explains other forms of chilling effects. Solove's privacy work explores how forms of tracking, monitoring, and surveillance encourages chilling effects and self-censorship is surely relevant to understanding the broader impacts of large scale automated legal enforcement under the DMCA.<sup>107</sup> When a person receives a personal and personalized DMCA notice, it suggests someone or something is watching, tracking, and monitoring her online actions, which could create noteworthy chilling effects due to concerns about present and possibly ongoing surveillance.<sup>108</sup> When you combine this range of legal and privacy concerns with the fact that DMCA

---

104. Carpou, *supra* note 27, at 585 ("Much literature has been devoted to ways in which automated takedown procedures necessarily result in the chilling of free speech and should, therefore, be disallowed or at least discouraged by the DMCA."). *See also* works cited *supra* note 26 (not exhaustive).

105. Katyal, *supra* note 28, at 337.

106. LIBEL AND THE MEDIA: THE CHILLING EFFECT (E.M. Barendt ed., 1997); *see generally* THE COST OF LIBEL: ECONOMIC AND POLICY IMPLICATIONS (Everette E. Dennis & Eli M. Noam eds., 1989).

107. Solove, *supra* note 86, at 488.

108. *See* Katyal, *supra* note 28, at 342–43 (discussing "piracy surveillance" that is greatly enhanced by the DMCA, as well as discussing how use of bots to detect/enforce copyright through notices creates a "panopticon" effect with privacy implications).

notices are being sent out daily by the million by automated processes it is easy to foresee substantial chilling effects. These concerns are tested and explored in this empirical analysis, with method and design set out in the next section.<sup>109</sup>

### III.METHOD AND DESIGN

#### A. Case Study One: DMCA Online Survey

Survey respondents were recruited using an online crowdsourcing platform,<sup>110</sup> from which past samples were found to be relatively representative of the U.S. internet using population.<sup>111</sup> The survey presented respondents with two hypothetical scenarios, one where the respondent received a DMCA-like notice and another where a “friend”

---

109. The ethical dimensions of this study were informed by the recent Menlo Report as well as Association of Internet Researchers ethical guidelines. The research also received ethical approval from the Oxford Internet Institute’s departmental Central University Research Ethics Committee (CUREC) at the University of Oxford. The design and method overall were informed by an empirical legal studies framework and methodological approach. *See e.g.*, Lee Epstein & Andrew D. Martin, *Quantitative Approaches to Empirical Legal Research*, in *THE OXFORD HANDBOOK OF EMPIRICAL LEGAL RESEARCH* 901-925 (Peter Cane & Herbert M. Kritzer, eds., 2010); Lisa Webley, *Qualitative Approaches to Empirical Legal Research*, in *THE OXFORD HANDBOOK OF EMPIRICAL LEGAL RESEARCH* 926-50 (Peter Cane & Herbert M. Kritzer, eds., 2010); Mark A. Hall & Ronald F. Wright, *Systematic Content Analysis of Judicial Opinions*, 96 CALIF. L. REV. 63, 63-122 (2008).

110. Amazon’s Mechanical Turk provides an “open” crowdsourcing platform for “task creation,” “recruitment,” “compensation,” and “data collection.” Michael Buhrmester et al., *Amazon’s Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?*, 6 PERSP. ON PSYCHOL. SCI. 3, 3 (2011). The platform has been used and “validated” as a tool for a broad range of empirical, experimental, and behavioral studies, including for conducting survey research. Indeed, Paolacci and Chandler recently concluded, after extensively canvassing existing research and evidence, that researchers can this platform for “virtually any study that is feasible to conduct online.” Gabriele Paolacci & Jesse Chandler, *Inside the Turk: Understanding Mechanical Turk as a Participant Pool*, 23 CURRENT DIRECTIONS PSYCHOL. SCI. 184, 186 (2014); *see also* Scott Clifford et al., *Are Samples Drawn from Mechanical Turk Valid for Research on Political Ideology?*, 2 RES. & POL., Oct.-Dec. 2015; Matthew J.C. Crump et al., *Evaluating Amazon’s Mechanical Turk as a Tool for Experimental Behavioral Research*, 8 PLOS ONE e57410 (2013).

111. Panos Ipeirotis, *Turker Demographics vs Internet Demographics*, COMPUTER SCIENTIST BUS. SCH. BLOG (Mar. 16, 2009), <https://perma.cc/22MD-8L9E>; Gabriele Paolacci et al., *Running Experiments on Amazon Mechanical Turk*, 5 JUDGMENT & DECISION MAKING 411, 411-12 (2010).

received a notice.<sup>112</sup> These scenarios are followed by questions aimed at eliciting likely behavioral responses to the hypothetical scenario rather than having participants self-report or indicate attitudes.<sup>113</sup> Surveys based on hypothetical scenarios have been used in similar empirical studies and for research in a range of fields, including studying online behavior.<sup>114</sup> Finally, the online activities tracked for impact—online speech and writing, online search, sharing of personally created content online, and contributions to social media networks—are all *presumptively* legal activities; Schauer and Solove are not concerned with chilling of illegal activities (all criminal laws should chill criminal activities).

---

112. The scenarios are meant to replicate or reproduce the impact for a user receiving a DMCA notice, but to avoid biasing responses neither scenario expressly mentions “DMCA.” For reference, this is DMCA scenario presented to respondents:

You post some content on a website that you really want to share with other people. However, a week later you receive by email a legal notice from the website's administrators. The legal notice states that the website received a complaint that the content you posted on the website violated a law. The legal notice states that the content you posted has been removed from the website to avoid a lawsuit. The legal notice also states that further legal steps will be taken against you, if you re-post the content online.

The “friend” DMCA scenario, for reference, in full:

A friend of yours posts on Facebook that he received a legal notice from his internet service company. The legal notice warned him that some of his online activities may have been illegal, including downloading unauthorized copies of computer programs. The legal notice warned him to avoid such activities in the future.

113. For example, research has found respondents' stated attitudes or self-reports do not always match their actual behavior or choices in practice online. Annika Bergström, *Online Privacy Concerns: A Broad Approach to Understanding the Concerns of Different Groups for Different Uses*, 53 COMPUTERS HUM. BEHAV. 419, 419 (2015).

114. Stephen M. Renas, Charles J. Hartmann, & James L. Walker, *An Empirical Analysis of the Chilling Effect*, in THE COST OF LIBEL: ECONOMIC AND POLICY IMPLICATIONS (E. E. Dennis & E. M. Noam, eds., 1989) Judith Townend, *Online Chilling Effects in England and Wales*, 3 INTERNET POL'Y REV., Apr. 3, 2014; *see also, e.g.*, Stanislav Mamonov & Marios Koufaris, *The Impact of Perceived Privacy Breach on Smartphone User Attitudes and Intention to Terminate the Relationship with the Mobile Carrier*, 34 COMM. ASS'N FOR INFO. SYS. 1158 (2014); Yi-Tai Seih et al., *Do People Want to Be Flattered or Understood? The Cross-Cultural Universality of Self-Verification*, 49 J. EXPERIMENTAL SOC. PSYCHOL. 169, 170 (2013). However, hypothetical questions also have limitations. Improperly designed or overly abstract hypothetical questions can lead to biased and unreliable responses. This case study thus employs questions with “response categories beyond simple yes and no responses,” which avoids forcing respondents into unreliable commitments with dichotomous. *See generally* B. Douglas Bernheim et al., *Do Hypothetical Choices and Non-Choice Ratings Reveal Preferences?*, (Nat'l Bureau of Econ. Research, Working Paper No. 19269, 2013). Moreover, the scenarios themselves also avoid complexity and abstraction to minimize bias.

The DMCA scenario set out in the survey is based on the already noted prior literature asserting chilling effect claims, criticisms, and concerns about the DMCA's notice and takedown system. Here, respondents were presented with two scenarios: first, that they received a DMCA notice about content the respondent had personally posted online, and second, where the respondent learns through a social media post on Facebook that a friend has received a personal legal notice about content posted online ("the third-party DMCA scenario"). This scenario aims to understand the impact of DMCA notices within networks.<sup>115</sup> The survey also elicits basic demographic information about respondents, and tests chilling effects by asking respondents how each scenario impacted how they engaged in online activities. Each scenario is described and followed by the same series of questions, which are generally based on a five-point scale, aiming to measure how the threat in the scenario would impact their behavior. It is from these responses that any potential or expected regulatory chilling effects are observed.

The survey was hosted and field tested using online survey software,<sup>116</sup> with 1,296 total survey responses collected in March 2015.<sup>117</sup> Responses to questions are first analyzed (percentages, summary, and descriptive statistics) to understand any apparent "chilling effect," including comparisons between results for each scenario, for comparative insights. To identify potential factors that may impact or influence any apparent "chilling effects" (e.g. being less likely to speak or share online, due to receiving a DMCA notice) results were statistically analyzed.<sup>118</sup>

---

115. Some social network studies have found internet users self-censor based on factors in their online networks, including audience, and that certain social effects can spread through networks. *See generally* Sauvik Das & Adam Kramer, *Self-Censorship on Facebook*, PROC. 7TH INT'L AAAI CONF. ON WEBLOGS & SOC. MEDIA 120–27 (July 8–11, 2013); Lorenzo Coviello et al., *Detecting Emotional Contagion in Massive Social Networks*, 9 PLOS ONE e90315 (2014).

116. SurveyMonkey was used, an online survey design and delivery service that has been "used for surveys in a number of areas including health research." Eugene Waclawski, *How I Use It: Survey Monkey*, 62 OCCUPATIONAL MED. 477 (2012).

117. The 1,296 total survey responses were collected in March 2015, with 64 survey responses excluded for being substantially incomplete (defined by 10 or more questions left unanswered; many of these were likely false-starts by respondents); another 18 excluded for being completed too quickly, and 2 more screened because the respondents had completed a version of the survey previously (in a field test).

118. In the survey case study, ordinary least squares (OLS) regression was used to statistically analyze findings, as it allowed for all relevant variables to be controlled in

### B. Case Study Two: Google Blogs and Twitter Analysis

This case study explores how receiving a DMCA notice impacts Google Blogger and Twitter users. To do so, it examined 500 Google Blogger accounts and 500 Twitter accounts that had received DMCA notices for content posted online (along with the corresponding notices themselves). The Google Blogger and Twitter accounts were located by drawing a random sample of 500 DMCA notices in the Lumen Database<sup>119</sup> for each platform.<sup>120</sup> Only notices that led Google and Twitter to remove or disable

---

order to isolate relationships. STEVEN G. HEERINGA ET AL., *APPLIED SURVEY DATA ANALYSIS* 235 (2010). Cohen's  $f^2$  was also used to test the effect size of findings, using the conventional interpretation of small (0.02), medium (0.15), and large (0.35) values. JACOB COHEN, *STATISTICAL POWER ANALYSIS FOR THE BEHAVIORAL SCIENCES* 413–14 (2d ed. 1988); TIMOTHY Z. KEITH, *MULTIPLE REGRESSION AND BEYOND: AN INTRODUCTION TO MULTIPLE REGRESSION AND STRUCTURAL EQUATION MODELING* 62–63 (2d ed. 2014). In case study two, involving blog and Twitter content analysis, the Goodman and Kruskal's gamma ( $\gamma$ ) test statistic was used to analyze statistically significant associations, including the effect size and direction of any such associations. It is a proportional reduction in error statistic measuring of how many fewer errors may be made in predicting the value of one variable by taking to account another. This was supplemented, for robustness, with Pearson's chi-square ( $\chi^2$ ) and Fisher's exact tests. Where any expected frequency was less than 1 in any cell or where the expected values were less than 5 in more than 20% of all cells, Fisher's exact test was used to test significance in these cases or variables recoded. Gamma is an appropriate measure of effect size when using ordinal categorical data (as here) and has an identical values range to  $r$  (Less than + or - 0.10: very weak association; + or - 0.10 to 0.19: weak; + or - 0.20 to 0.29: moderate; + or - 0.30 or above: strong). Gamma is expressed on a spectrum of -1 to 1, with -1 suggesting a perfect negative association and 1 a perfect positive association; a return of 0 suggests no association between variables at all. The gamma test of significance was calculated as significant at the  $p < 0.05$  level where  $\gamma / \text{ASE} = \pm 1.96$  (95% confidence) and at the  $p < 0.01$  level where  $\gamma / \text{ASE} = \pm 2.575$  (99% confidence). See LOUIS M. REA, *DESIGNING AND CONDUCTING SURVEY RESEARCH A COMPREHENSIVE GUIDE* 229–30 (4th ed. 2014); HUGH COOLICAN, *RESEARCH METHODS AND STATISTICS IN PSYCHOLOGY* (2017); Christopher Ferguson, *An Effect Size Primer: A Guide for Clinicians and Researchers*, 40 *PROF. PSYCHOL.: RES. & PRAC.* 532 (2009); JOSEPH F. HEALEY, *THE ESSENTIALS OF STATISTICS: A TOOL FOR SOCIAL RESEARCH* (2007); D.H. STAMATIS, *SIX SIGMA AND BEYOND* 78–79 (2002).

119. OSPs like Google and Twitter as a matter of policy send DMCA notices they receive and act upon to the Lumen Database (formerly the “Chilling Effects Repository”). See *infra* note 120.

120. The random sample was obtained from the Lumen Database repository of notices sent to Google Blogger accounts in 2012 and 2013. The recent study conducted by Urban et al. similarly sampled DMCA notices from 2013. Urban et al., *supra* note 22, at 31. Part of the benefit of an earlier sampling frame is to be able to carry out one of the stated aims of this analysis—to investigate the impact of automated DMCA notices with non-automated ones. In recent years, automated DMCA notices have constituted an overwhelming proportion of all DMCA notices sent to platforms like Google and Twitter. For example, from 2014 onward, the top 10% of entities sending DMCA notice

access to content were used. The bloggers and Twitter users in question would have received notice that the DMCA notices were received and acted upon.<sup>121</sup> This case study aims to glean evidence as to the impact of DMCA notices on targeted users, including any regulatory chilling effects.

The analysis proceeded iteratively. Theoretical and empirical observations and assumptions based on existing literature provided the

---

(in terms of total notices sent) are responsible for 90% of all DMCA notices sent to Google. Moreover, because each notice can include thousands of URLs to be removed, the top 0.1% of notice senders actually accounted for over 76 million removal requests per month as compared to the 50,000 URLs requested per month by the bottom 99.9% of DMCA notice senders. This is due to automated processes creating and sending DMCA notices to Google, Twitter, and other platforms. As automation of notices largely came to prominence in 2011, and increased yearly, thereafter, samples drawn from 2012 and 2013 increase the statistical odds of including non-automated notices. See Stuart Efstathis, *Who Polices the Internet?*, DIGITAL SOC. CONT. (Oct. 3, 2016), <https://perma.cc/7FJK-7NVW>. See also Daniel Seng's discussion of the automated notices emerging in late 2011. See Seng, *supra* note 25, at 408–09. For context, there were 48,915 total DMCA takedown notices sent to Google Blogger in 2012 and 32,365 in 2013 (up to July), while there was 3,334 total DMCA takedown notices sent to Twitter accounts in 2012 and 4,631 in 2013 (up to July). All of these numbers were confirmed in writing with Wendy Seltzer, founder of the Lumen Database Project, and Adam Holland, Director of the project. E-mails from Adam Holland, Director of the Lumen Database Project, and Wendy Seltzer, founder of the Lumen Database Project, to the author (May 17, 2014) (on file with author).

121. The Lumen database repository categorizes DMCA notices in its repository between “unspecified” (unknown action), “yes” (removal action taken in response to DMCA notice), “no” (no action taken in response), and partial (partial action taken in response to notice), see LUMEN DATABASE, *supra* note 45. In fact, Google’s stated policy (both at the time this study was conducted, but also presently) is to share with the repository only those notices it has acted on (removed or disabled content pursuant to) with the Chilling Effects/Lumen repository. See Google’s Transparency Report FAQ, stating Google will “share a copy of qualifying copyright removal requests with the public site Chilling Effects.” Google Transparency Rep. Help Ctr, *supra* note 101. This is further supported by the DMCA policy linked to Google’s Legal Help site, which states, “We may also document notices of alleged infringement on which we act. We may forward the content in your notice to the nonprofit organization Lumen, which publishes these notices after removing certain personal information.” *The Digital Millennium Copyright Act*, GOOGLE: LEGAL HELP, <https://perma.cc/M6HY-QHFG> (archived Sept. 24, 2019). Twitter similarly states in their Copyright Policy that “[i]f we decide to remove or disable access to the material, we will notify the affected user(s) and provide them with a full copy of the reporter’s complaint (including the provided contact information) along with instructions on how to file a counter-notice. We will also forward a redacted copy of the complaint to [Lumen](#), with your personal information removed.” *Twitter Rules and Policies: Copyright Policy*, TWITTER HELP CTR., <https://perma.cc/Z4W9-A488> (archived Sept. 24, 2019). Google and Twitter, as required by the DMCA, each notifies their users of DMCA takedown notices so that their users or subscribers may file counter-notifications. See Seltzer, *supra* note 28; GOOGLE, TRANSPARENCY REPORT FAQ (2015) and TWITTER, COPYRIGHT POLICY (2019).

foundation for coding and method. Other empirical legal studies methodological approaches<sup>122</sup> provided guidance for coding and data collection, particularly systematizing data collection to increase reliability. Thus, code also evolved inductively in the process of data collection, analysis, and interpretation, as coding was used to categorize blogs and Twitter accounts. Each DMCA notice and corresponding Blogger or Twitter account was viewed and coded for a range of relevant variables. First, the accounts were coded as to whether the blog post, tweet, or content targeted by the DMCA notice was *offline or inaccessible*.<sup>123</sup> In the case of Blogger, this would mean either the targeted blog post is offline or inaccessible (but the blog remains online); the entire blog has been suspended by Google; or the blog has been removed or deleted by the user.<sup>124</sup> In the case of Twitter, this would mean the targeted tweet or content is offline or access has been disabled (but the Twitter account remains online and public); the entire account has been suspended by Twitter; or the user has protected their tweets from public access. Other variables coded include: the nature of the content targeted by the DMCA

---

122. See works cited *supra* note 110. For example, inter-coder reliability was tested to ensure reliability and validity in coding.

123. Each Blogger or Twitter account (and corresponding DMCA takedown notice each account) were coded for this variable on a range of 1 to 4. If the targeted blog post, tweet, or content was online and accessible, it was coded as 1; if the targeted blog post, tweet, or content was offline or inaccessible but the blog or Twitter account in question was still online, it was coded as 2; if the blog or Twitter account targeted by the DMCA takedown notice was deleted or removed by the user, it was coded as 3; if the targeted blog or Twitter account was suspended (e.g., by Google or Twitter, as the case may be) it was coded as 4. This coded variable was approached as ordinal, as a “suspended” Google Blog or Twitter account meant that the user could no longer tweet or blog indefinitely (this is different from user deleted blogs or Twitter accounts, wherein users could re-open their blog or Twitter account at a later point). Examples of this coding can be found at Appendix A and B.

124. When Google’s Blogger platform receives a valid DMCA notice, it usually changes the publicly published blog post containing the allegedly infringing to “draft status,” which moves it offline and inaccessible. However, it may also suspend the account (possibly for repeated terms of service violations). *New Tools for Handling Copyright on Blogger*, *supra* note 47 (“At this point, Joe Blogger has the right to file a counter-notice and request the post be restored if he believes the takedown was improper. He may also edit the post to remove the allegedly infringing content himself.”); *Blogger Blog Take-Down Notification for Copyright Content*, *supra* note 47 (discussing a recent example of a notification Blogger has sent to a Blogger user); *BLOGGER CONTENT POL’Y*, <https://perma.cc/92VK-858T>.

notice; the type of the blog targeted (Blogger only);<sup>125</sup> the number of total followers (Twitter) or comments (Blogger) for the recipient of the DMCA notice;<sup>126</sup> whether the user blogged or tweeted about receiving a DMCA notice or copyright more generally;<sup>127</sup> and the apparent location of the recipient Blogger or Twitter user.<sup>128</sup> The purpose of coding for this wide array of variables was to explore potential factors that may impact and influence any apparent findings.

Seltzer theorizes the DMCA as a regulatory scheme that imposes costs and barriers that encourages chilling effects. Building on this approach, this case study uses the success that DMCA notices may have in getting content removed or disabled online—that is, their offline or online status— as the main proxy for understanding how those DMCA notices may chill Internet users’ speech, expression, or content. Analysis of blog posts or tweets that are made in response to receiving DMCA notices provide additional means of investigating any potential chilling effects. Finally, the statistical software Stata was used to perform cross tabulations and statistical measures of association to analyze, test predictions, and explore relationships in the data.

---

125. Often, there was not enough information to determine the “type” of Twitter account; this was different from Google Blogs, as clear blog type categories emerged from the data during the coding process.

126. Each Blogger or Twitter account (and corresponding DMCA takedown notice each account received) was coded for this variable on a range of 1 to 5. If the targeted blog or Twitter account had no comments or followers, it was coded as 1; if the targeted blog or Twitter account had 1-49 comments or followers, it was coded as 2; if the targeted blog or Twitter account had 50-99 comments or followers, it was coded as 3; if the targeted blog or Twitter account had 100-999 comments or followers, it was coded as 4; if the targeted blog or Twitter account had 1000 or more comments or followers it was coded as 5. While followers are clearly stated on Twitter accounts, a simple script was used to scrape/compile the number of comments on a Google Blog.

127. To determine this, a Blogger’s blog was searched, using Google search, for any mention of keywords “DMCA” or “copyright” and any responses or statements coded and logged. A similar process was completed for the Twitter accounts. Using Twitter’s special tweet search function, each Twitter user’s tweet stream was searched for the keywords “DMCA” or “copyright” and, if found, logged and coded.

128. Here, if the targeted blog or Twitter account indicated a location inside the United States, it was coded as 2; if the targeted blog or Twitter account indicated a location outside the United States, it was coded as 3; if location was unclear, the targeted blog or Twitter account was coded as 1.

### C. Hypotheses / Predictions

In the first case study, the online survey was designed to test whether a user who received a DMCA notice would be less likely to engage in certain legal activities online or more careful about how they do so, which would be evidence of the DMCA notice's chilling effects. Based on a theory of chilling effects associated with DMCA notices, it is predicted that *internet users will be less likely to engage in a range of legal online activities or will be more careful and cautious about how they engage in such online activities in response to receiving a DMCA notice (H1)*. I also ask whether a Schauer theory focused on legal uncertainties and avoidance of legal harms and punishments or a Solove theory focused on privacy has greater explanatory power for understanding chilling effects in this context. This aspect of the study also explores other possible predictors for any chilling effects observed, including age, gender, income, education, level of internet use and social network engagement. Since this work is exploratory, there are no hypotheses as to predictors.

In the second case study, when users are notified that a DMCA notice was received for content they posted on his or her Blog or Twitter account, the notice will, in theory, chill or deter the user from re-enabling or re-posting the content or filing a counter-notice. Fear of criminal or civil copyright penalties, awareness that third parties are monitoring users' activities, or the additional "costs" associated with having to re-assert the legality of their online expression<sup>129</sup> could all chill speech online. In some cases, the users may also be so fearful of legal claims that they will delete their Blog or Twitter account, which effectively eliminates any future expression by the user: a significant chilling effect. Thus, based on a theory of chilling effects, it is predicted that *receiving a DMCA notice will lead to the removal of targeted content on a majority of Blog and Twitter accounts (H2)*.<sup>130</sup>

---

129. Seltzer, *supra* note 28, at 176–77.

130. It may be hypothesized that a user, on receiving a DMCA notice, may simply delete their current account and create a new one, with a different Blogger/Twitter profile, or even on a different platform, and re-post or re-share the targeted content. However, there is good reason to believe that this would be uncommon, even rare among social media users. Twitter and Blogger are what Ellison and Boyd call social network sites ("SNS"), and a "key driver" of SNS is a "desire to communicate and share content" and "much of what is novel" on these sites is how users incorporate their

This analysis also compares the impact of automated legal enforcement to that of non-automated enforcement. DMCA notices were thus also coded as either automatically or manually sent to explore this differential impact in the second case study.<sup>131</sup> Based on the work of Perel and Elkin-Koren and Seltzer, I predict that *since algorithmic copyright enforcement may not take into account more flexible standards required by copyright law determinations—like issues of fair use, which cannot be determined by bright lines enforced by algorithms—that content targeted by automated DMCA notices will be more likely to be offline or inaccessible*

---

“connections”—their “friends,” followers, users, etc. into their SNS “online practices.” See Nicole B. Ellison & Danah M. Boyd, *Sociality Through Social Network Sites*, in THE OXFORD HANDBOOK OF INTERNET STUDIES 9 (2013). In other words, SNS users build up a network or following—an audience or community in which they are embedded—and they become tied to that community or audience; the primary driver of their SNS use and online practices is connecting, communicating, and sharing with these articulated audience or network of connections. For a Twitter or Blogger user to delete or abandon their accounts—and thus their network of friends, followers, connections—in order to simply re-post or re-share a single targeted piece of content, would run contrary to this fundamental aspect of SNS use. Indeed, Justine Gangneux recently reviewed a growing body of research on how such SNS user account deletion/termination constitutes a “great social loss” and that instead users engage in a range of other “soft strategies” short of account deletion/termination—like “taking a break” from SNS, “not liking or friending”—that allow users to remain on the platforms. Justine Gangneux, *Logged in or Locked in? Young Adults’ Negotiations of Social Media Platforms and Their Features*, J. YOUTH STUD. 4–5 (2019). All of this suggests such account deletion to re-share/re-post elsewhere, disconnected from existing audiences and followers, would be uncommon.

In fact, perhaps the most compelling evidence that these practices are uncommon are empirical findings of the study discussed in this article. The survey responses show that 71% of internet users indicated they would be unlikely to re-post or re-share, either publicly or privately, the content targeted by the received DMCA notice. This is consistent both with the research noted above as well as chilling effects theory more generally.

131. Notices with less than 10 URLs for removal were treated as individual sent DMCA notices while those with over ten URLs were treated as automated notices. Based on Seng’s findings, coding this way provides a proxy to distinguish automated from non-automated DMCA notices. Seng found in his empirical study of all DMCA notices in the Lumen Database between January 2001 through December 2012, that automated and algorithmic DMCA notices—which became more prominent starting in 2011 and increased substantially in 2012—tended to send “mega” DMCA notices—notice that include greater numbers of URL removal requests—while individuals typically sent “micro” notices, that is, notices with a single URL or takedown request. Despite the emergence of “mega” notices in 2012 (with over 25,000 removal requests per DMCA notice) thanks to automated programs and processes, Seng infers that notices with two and four requests per notice in 2011 and 2012, respectively, represent notices sent by individuals and not automated processes. Seng, *supra* note 24 at 408-409 (Seng notes that despite mega notices in 2012, the median number of takedown requests per notice was two in 2011 and four in 2012).

(H3).<sup>132</sup> That is, the user, being chilled, is more thus likely to comply with the DMCA notice—the user would not file a counter-notice to have the targeted content put back online or repost/re-enable the targeted content. The user may also delete, lock, or protect their account, also rendering the targeted content offline/inaccessible.

#### IV. RESULTS AND DISCUSSION

##### A. DMCA Impact from Survey Analysis

There were 1,296 total respondents to the survey. The population had a few biases typical of participant pools, from previous studies, recruited via online recruitment platforms (i.e. participants were slightly younger and with incomes slightly lower than the general U.S. internet population),<sup>133</sup> although unlike previous such participant pools, this one was more balanced in gender with 49.7% male and 50.3% female participants. This participant pool, also like samples in previous studies, consisted of heavy internet users, with nearly half (49.5%) reporting “continual” online connectivity, and another 46.1% reporting connecting to the internet several times a day. A majority (51.9%) shared content or posted online “several times a week” or more and almost 17% shared content daily.<sup>134</sup>

##### 1. DMCA Impact on Activities Online

Respondents were presented with a hypothetical scenario comparable to receiving a DMCA notice. They were presented with a personalized legal

---

132. See Perel & Elkin-Koren, *supra* note 22, at 486–88; Urban et al., *supra* note 22, at 95; Seltzer, *supra* note 28, at 178.

133. There were 1,296 total responses and the final set analyzed was 1,212 (84 responses excluded per reasons at *supra* note 72), which reflected other “relatively representative” pools recruited through Amazon’s platform. See Panos Ipeirotis, *supra* note 111; Paolacci et al., *supra* note 111, at 411–12.

134. Similarly, responses to Questions 32 and 34 indicated very nearly half (49.8%) contributed to online networks and related communities at least several times a week, with almost 16% contributing several times a day. Sharing of content respondents “personally created” was less pronounced, with 41.2% of respondents indicating “[r]arely or never” share such content, while 27% sharing such “personally created” content once a week, 21.7% answered “[s]everal times a week”, 4.8% sharing “[o]nce a day” and 5.4% sharing “[s]everal times a day or more.”

notice about content they had posted online.<sup>135</sup> Respondents were then questioned as to how they would react to this scenario in relation to various online activities. The results suggest a notice that is personalized and personally received has a substantial chilling effect, leading to compliance, that is, removing targeted content and not re-posting or re-sharing thereafter. When respondents were asked if they would “try to re-post or re-share the content” targeted by the legal notice publicly or even privately, a total of 71% were unlikely to do so, with 55% of respondents indicating they would be “very unlikely” and 16% indicating they would be “somewhat unlikely.”

Results also offered evidence of chilling effects on a range of other activities in response to this DMCA scenario, including online speech, search, sharing, and creating personal content. To investigate effects on online speech, respondents were asked whether they would be “more likely or less likely to speak or write about certain topics online” in response to the DMCA scenario. Here, a full 75% of respondents reported being “much less likely” (40%) or “somewhat less likely” (35%) to “speak or write about certain topics online” as a result. Respondents were also asked to indicate their level of agreement with the following statement, “I would be more careful about what I say or discuss in certain contexts online” in response to the DMCA notice scenario. Responses again suggested significant chilling effects with a total of 81% of respondents either “strongly agreed” (50%) or “somewhat agreed” (31%) with the statement.

Responses also suggested a chilling effect on willingness to share content that respondents had personally created. Sharing such content online would be demonstrably legal (as they own the copyright) and has socially beneficial dimensions as sharing personally created content can enrich through new forms of expression, can lead to additional creativity and innovation.<sup>136</sup> Here, respondents were asked in response to the DMCA scenario whether they would be “more likely or less likely to share content

---

135. For the full description of the scenarios as used in the survey, see *supra* note 112.

136. For a discussion of different theories of the value of user creation including in innovation, see generally Anja Bechmann & Stine Lomborg, *Mapping Actor Roles in Social Media: Different Perspectives on Value Creation in Theories of User Participation*, 15 NEW MEDIA & SOC'Y 765 (2013).

on the internet that [the respondent] personally created, authored, or made (e.g., a digital photo, song, blog post, Facebook status update, etc.)” Again, responses suggested a noteworthy chilling effect with 72% of respondents either much less likely (38%) or somewhat less likely (34%) to share such content online after receiving a personalized legal notice. This kind of chilling effect also likely impacts expression, speech, and creativity itself, as sharing such content is likely a motivation to create in the beginning. If that sharing is chilled, then the creativity itself may also be so chilled.

Social media engagement was also chilled. Respondents were asked whether they would be “more likely or less likely to contribute to online social networks, communities, and discussion forums” in response to the DMCA scenario. Again, responses suggested a noteworthy chilling effect. A total of 75% of respondents were either “much less likely” (37%) or “somewhat less likely” (38%) to contribute to online social networks and communities on receiving a personal legal notice about content posted online. And in terms of online search, the DMCA scenario similarly proved impactful. Respondents were asked to indicate their level of agreement with the statement, “I would be more careful about what I search for online.” 30% of respondents “strongly agreed” and an additional 29% “somewhat agreed.”

This scenario also triggered a host of other chilling effects, creating caution and privacy concerns for respondents. For instance, a full 81% of respondents agreed with the statement that the scenario would make them “more concerned about my privacy, taking extra steps to protect it,” with 55% strongly agreeing and another 26% somewhat agreeing. 86% would be more careful about what they “post or share” online as well, with 57% strongly and 30% somewhat agreeing.

## *2. Impact of a “Friend” Receiving DMCA Notice*

The results also suggested a form of secondary or indirect chilling effects from the “friend” DMCA scenario—here, a friend of the respondent posts online about receiving a legal notice about online activities. The most noteworthy finding was evidence of significant privacy chilling effects, with 90% of respondents either strongly agreeing (64%) or somewhat agreeing (26%) with the statement that the scenario would make them more “concerned” about their privacy, leading them to take steps to

protect it. Clearly, as with a personally received legal notice, awareness that a friend or other person in one's social network has received a legal notice, also raises privacy concerns. The scenario would also chill more robust forms of speech and expression, with 76% of respondents "strongly" (41%) or "somewhat" agreeing (35%) that they "would more careful about what I say or discuss in certain contexts online" if a friend of theirs were to receive such a personal legal notice. Online search would also be impacted, with 77% of respondents "strongly" (39%) or "somewhat" (38%) agreeing that they would be "more careful" about online searches in response to this scenario. 83% of respondents agreed that they would also be more careful about what they post or share online as a result.

All of these findings offer strong support for the existence of serious chilling effects on a range of activities due to a personalized and personally received legal notice that targets a person and their online activities. Moreover, receiving such a notice creates privacy-related chilling effects, likely arising from the legal notice indicating that a third party has been watching, tracking, and monitoring their activities.

### 3. *Predicting Impacts*

Results were also statistically analyzed to identify factors that may influence or correlate with potential chilling effects suggested by responses, including factors like age, gender, education, income, internet use, how often they share personally created content online, how often they contribute to social network sites and other online forums, and their level of knowledge about internet laws. The question here is whether certain demographic or similar factors can predict certain chilling effect outcomes. The results are set out in Table 1 and Table 2:

**Table 1: OLS Regression Results Predicting Respondent Privacy Concerns and Willingness to Discuss, Search, Share Content, or Engage with Social Networks (SN) After Personally Receiving a Legal Notice About Content Posted Online (Regression Coefficient with Standard Errors in Parenthesis)**

Predictor	Less Likely to Speak / Write Online	More Caution in Online Speech	More Caution in Online Search	Less Likely to Share Online	Less Likely to Engage with SN
Age†	-0.00(0.04)	-0.00(0.04)	<b>-0.22(0.04)***</b>	-0.02(0.04)	0.02(0.04)
Gender	<b>0.22(0.06)***</b>	<b>0.16(0.06)***</b>	<b>0.18(0.07)***</b>	0.08(0.05)	<b>0.11(0.05)**</b>
Education level†	0.09(0.05)*	0.07(0.04)	-0.05(0.05)	0.04(0.04)	0.01(0.04)
Income level	0.03(0.03)	0.01(0.03)	<b>0.10(0.03)***</b>	0.00(0.02)	0.01(0.02)
Internet usage level	0.05(0.04)	-0.01(0.04)	<b>-0.10(0.05)**</b>	0.07(0.04)*	0.07(0.04)*
Online sharing	0.0(0.03)	0.02(0.03)	0.01(0.03)	0.03(0.03)	-0.01(0.03)
Legal familiarity†	<b>-0.08(0.04)**</b>	-0.05(0.03)	0.06(0.04)	<b>-0.08(0.03)**</b>	<b>-0.09(0.03)***</b>
NSA awareness	-0.03(0.04)	-0.04(0.03)	-0.05(0.04)	0.02(0.03)	0.02(0.03)
SN engagement	<b>-0.05(0.02)**</b>	-0.03(0.02)	-0.01(0.03)	<b>-0.06(0.02)***</b>	-0.03(0.02)
Privacy concerns	<b>0.27(0.03)***</b>	<b>0.55(0.03)***</b>	<b>0.69(0.03)***</b>	<b>0.29(0.02)***</b>	<b>0.30(0.02)***</b>

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ , † = Recoded

All models significant (Prob > F = 0.00) with medium / near medium effect sizes (Cohen's  $f^2$ ).

**Table 2 – Friend Scenario: OLS Regression Results Predicting Respondents Being More Careful and Cautious in their Online Speech and Searches After Made Aware A Friend Has Received a Threatening Legal Notice For Unauthorized/Illegal Activities Online (Regression Coefficient with Standard Errors in Parenthesis)**

Predictor	More Cautious/Careful in Online Speech	More Cautious/Careful in Online Search
Age†	-0.01(0.06)	-0.11(0.06)*
Gender	<b>0.17(0.09)**</b>	0.08(0.09)
Education level†	<b>-0.19(0.7)**</b>	<b>-0.24(0.09)***</b>
Income level	<b>0.11(0.04)**</b>	<b>0.09(0.04)**</b>
Internet usage level	-0.04(0.06)	-0.04(0.07)
Online sharing	0.03(0.04)	<b>0.10(0.05)**</b>
Legal familiarity†	-0.02(0.05)	0.09(0.06)
NSA news awareness†	-0.02(0.05)	-0.03(0.06)
SN engagement	<b>-0.11(0.04)***</b>	<b>-0.11(0.04)***</b>
Privacy concerns	<b>0.52(0.05)***</b>	<b>0.35(0.05)***</b>

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ , † = Recoded

All models significant (Prob > F = 0.00) with medium / near medium effect sizes (Cohen's  $f^2$ ).

There are a number of noteworthy results here. First, the results suggest the single strongest predictor of chilling effects was greater privacy concerns. That is, participants indicating stronger privacy concerns about receiving a DMCA notice were statistically more likely to be chilled. In every single impact measured, across a range of contexts and activities in both Table 1 and Table 2, greater privacy concerns had a statistically significant association with greater impact. Concerns about legal harms were also a factor, though much less than privacy. Participants

reporting more legal knowledge were less chilled in three of seven impacts measured. That is, respondents reporting more familiarity with internet laws were less likely to report being chilled in relation to speech, online sharing, or social network engagement. This might be the product of the respondent simply having a better understanding of what a targeted legal threat might mean, and therefore the respondent is more affected and impacted in their activities. Nevertheless, while legal knowledge was a factor, greater privacy concerns was a stronger predictor of chilling effects across all contexts.

Second, findings for both DMCA scenarios set out in Tables 1 and 2 show a statistically significant gender effect. Based on how gender was coded in the data, the positive association showing for three of four impacts measured in Table 1 suggests female respondents were “chilled” in this DMCA scenario more than men. That is, female participants are less likely to a statistically significant degree to speak or write online in certain contexts, less likely to engage with social networks, and would be more careful in their online speech as well as online search activities. There was also a gender effect in the “secondary” DMCA scenario, where a friend is targeted with a personal legal notice about content posted online (Table 2), with female respondents being more cautious or careful about their online speech.

Third, a greater level of engagement with social networks is largely associated with less chilling effects for multiple activities tracked in Tables 1 and 2. That is, respondents that reported contributing to social networks and online communities more often were less likely to be chilled in terms of their speech, sharing, and social network engagement online in response to the DMCA notice scenario. This may be the product of these more engaged internet users making a decision that staying engaged with their online social networks is more important than the risk associated with receiving a DMCA notice.<sup>137</sup> There were several other noteworthy findings, like an age effect when it came to caution in online search (younger participants were more chilled than older ones). Education was also a factor in a few contexts—there was a statistically significant negative association between greater levels of education and greater levels of chill

---

137. Schauer, *supra* note 83, at 694 (theorizing that chilling effects was the product of similar cost/benefit analyses by rational and deliberative individuals).

on the two activities tracked (online speech and search). This suggests that the more educated respondents were, the less likely they reported being more affected or chilled in the speech and search as a result of the friend's DMCA notice.

#### 4. *Legally Challenging DMCA Notices*

As noted in Part II.1, DMCA counter-notices are very rare with few copyright lawsuits filed arising from the initial notice-and-takedown occurrence; few internet users appear willing to challenge DMCA notices they receive.<sup>138</sup> To explore why, questions were also posed in relation to the DMCA notice scenario how likely the respondents would take steps to legally challenge the notice if they believed it was wrong or mistaken. Here, 34% indicated they would challenge; 36% indicated they would not, and 30% responded that they “[d]on’t know.” For those indicating they would not legally challenge notice, or did not know, legal costs for doing so was the most common factor cited (81%), while time costs (66%), wishing to avoid additional legal trouble (53%), inability to understand the issues at stake (28%), or not knowing how to legally challenge the notice (34%), were also common reasons for not legally challenging. And in terms of the chilling effect on re-posting and re-sharing content targeted by such a notice—it was substantial, with 70% “very unlikely” (54%) or “somewhat unlikely” (16%) to re-post or re-share. On the other hand, 15% of respondents said receiving the notice would make them “very likely” (4%) or “somewhat likely” (11%) to re-post or re-share. We might label these respondents as resilient to chilling effects. These findings were also statistically analyzed to identify any factors can predict the likelihood of challenging (see Table 4):

---

138. Urban et al., *supra* note 22, at 95 (noting their study “and other research on this issue consistently shows that counter notices are rarely used”); *see also* Urban & Quilter, *supra* note 28, at 679; Cobia, *supra* note 37, at 391.

**Table 3: Notice Compliance: OLS Regression Results Predicting Respondent's Willingness to Challenge Notice Received Or Re-Post/Re-Share Targeted Content Elsewhere Online (Regression Coefficient with Standard Errors in Parenthesis)**

Predictor	More Likely to Legally Challenge the Notice Received	Less Likely to Re-Post or Re-Share Targeted Content Online
Age†	-0.03(0.02)	-0.04(0.05)
Gender	-0.03(0.03)	<b>0.27(.07)***</b>
Education level†	0.02(0.02)	-0.10(0.05)*
Income level	0.00(0.01)	<b>0.09(0.03)***</b>
Internet usage level	0.05(0.02)	<b>0.13(0.05)**</b>
Online Sharing	0.00(0.02)	<b>-0.08(0.04)**</b>
Legal knowledge†	<b>0.09(0.02)***</b>	<b>-0.11(0.04)**</b>
NSA News Awareness†	0.01(0.01)	<b>-0.12(0.04)***</b>
SN Engagement	0.02(0.02)	-0.00(0.03)
Privacy concerns	<b>0.04(0.01)***</b>	<b>0.16(0.03)***</b>

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ , † = Recoded

All models significant (Prob > F = 0.00) with medium / near medium effect sizes (Cohen's  $f^2$ ).

Here, again, respondents' concerns about privacy were a noteworthy predictor. Greater privacy concerns were statistically associated with participants being more likely to challenge a DMCA notice but less likely to re-share or re-post content targeted by notices. Similarly, gender effects were apparent here too, with women statistically less likely to re-post or re-share targeted content.

Income and overall internet usage were associated with less willingness to re-post and share while, on the other hand, greater online sharing, legal knowledge, awareness of news about NSA surveillance, and social network site engagement, all were associated with more willingness to re-post. However, it should be kept in mind that the vast majority of respondents (70%) were either "very unlikely" (54%) or "somewhat unlikely" (16%) to re-post or re-share targeted content. In terms of challenges, when all variables were controlled for, only legal knowledge was a factor in whether people were willing to legally challenge a DMCA notice they believed to be wrong. If you are more familiar with internet laws, it makes sense that you may be more willing to challenge a legal threat if you believe it is wrong or inaccurate. This also suggests that providing users with information as to their legal rights may be a means to mitigate chilling effects.

Interestingly, legal costs were the most commonly cited reason for *not* taking steps to legally challenge a notice (with 81% of responses citing legal costs), but findings suggest no statistically significant association between income levels and willingness to challenge. There was also no

education effect here. This might mean legal costs were perceived as so prohibitive that it remained a concern no matter what income or education level respondents reported. The results might also mean that other factors are more decisive, such as the social media engagement or legal familiarity, which both show positive associations with willingness to legally challenge the notice. If internet users are making rational decisions about their benefits and costs of responding to a notice, these personal factors may ultimately weigh more heavily in the determination.

### *B. Impact from Blog and Twitter Analysis*

#### *1. Information on Sample of DMCA Notices*

This part of the analysis examines the impact a random sample of actual DMCA notices had on Google Blogger and Twitter users. The Google Blogs sample totaled 500 (n=500) though one blog returned a malware warning (and was thus not visited) so in some coded variables observations will total 499 (n=499). The total number of accounts for the Twitter sample was also 500 (n=500). The types of the blogs targeted by DMCA notices in the Blogger sample were quite wide ranging (Figure 1):

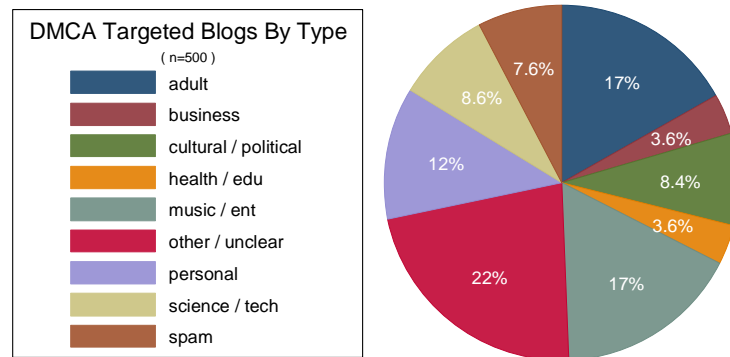


FIGURE 1: GOOGLE BLOGGER BLOG TYPES

While the largest percentage of blogs were not clearly definable (the “other / unclear” constituted 22%)—there was no obvious or apparent “category” for these blogs—blogs that concerned music, entertainment,

and sports (“music / ent”) and adult content (“adult”) were the second largest groups (17% each). Notably, blogs that were clearly spam accounts constituted only 7.6% of the sample. Personal (12%), social, cultural, and political (8.4%), and science and technology blogs (8.6%) all constituted a notable share of the Blogger sample. The content targeted by the DMCA notices in each sample was also wide ranging. For the content targeted in the Blogger sample (Figure 2), the largest percentage of content targeted by DMCA notices was mixed media (25%), that is, content involved more than one media type (e.g., images and video or text and images):

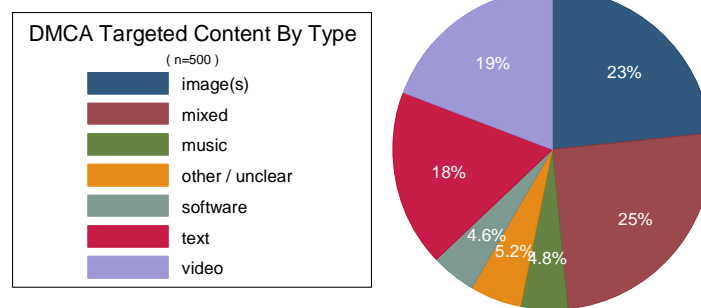


FIGURE 2: BLOGGER SAMPLE

Images (23%) and video (19%) also constituted large percentages of content targeted, with text-based content also constituting a notable share of content targeted (18%). Software and “other” categories of content all took smaller shares. In the Twitter sample (Figure 3), the largest percentage of content targeted by DMCA notices was multimedia (54%) (which included video, software, and anime).

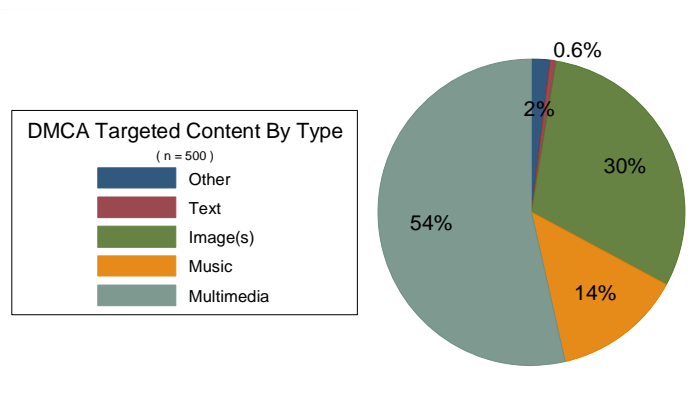


FIGURE 3: TWITTER SAMPLE

Images were the next largest type of content targeted (30%) followed by music, which constituted 14% of the content targeted. Much smaller percentages were text-based content (0.6%) and miscellaneous content that did not fall into other main categories (“other” constituted 2%).

## 2. *Impact of the DMCA notices*

This section explores and analyzes any evidence of chilling effects in the data collected, assessing whether there is any foundation or substance to concerns that the DMCA notice and takedown system has a chilling effect on legal activities and expression online. On this count, perhaps some of the most convincing (and surprising) evidence of the DMCA’s chilling effects in this case study is how incredibly effective DMCA notices are in having the blog posts, tweets, or other content they target removed or rendered offline and inaccessible (usually permanently)—suggesting an important effect not only on content, but also on the internet users behind the content. In the Blogger sample, 499 of 500 total blogs (and DMCA notices) were analyzed, while all 500 Twitter accounts were visited and analyzed in its sample. See Figure 4 for the results on this count:

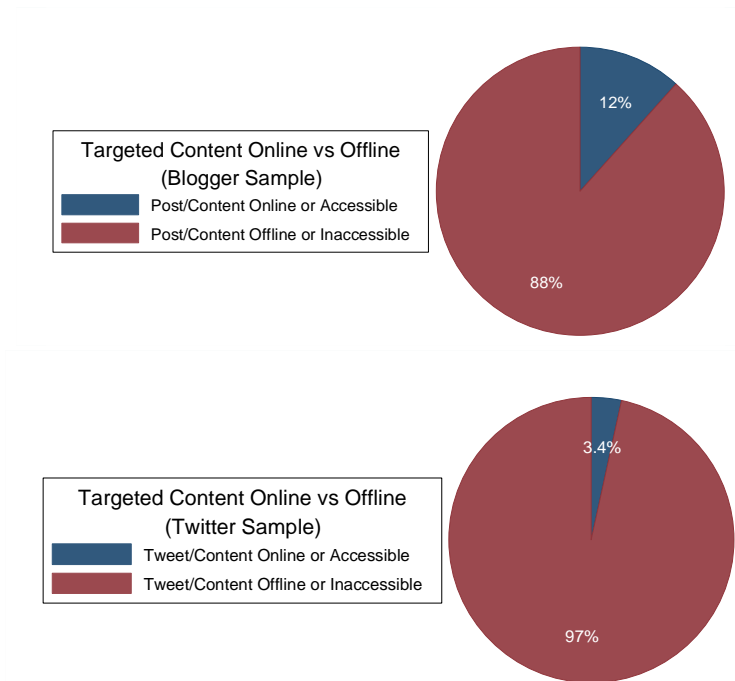


FIGURE 4: DMCA NOTICE IMPACT

In the Blogger sample, 88% (n=441) of blog posts or content targeted by the DMCA notices were found to be offline, removed, disabled, or inaccessible. Only 12% (n=58) were online and accessible. In the Twitter sample, the percentage of tweets or tweeted content withheld, removed, disabled, or inaccessible, was even higher: 97% (n=483). Also, only 3.4% of content targeted was currently online and accessible. These numbers suggest (roughly) 90% of the Bloggers and 97% of the Twitter users failed to take steps, legal or otherwise, to get their content back online, at least directly. But this data can be broken down further, to provide a clearer picture as to the impact of DMCA notices and *why* the specific content is online or offline (see Figure 5):

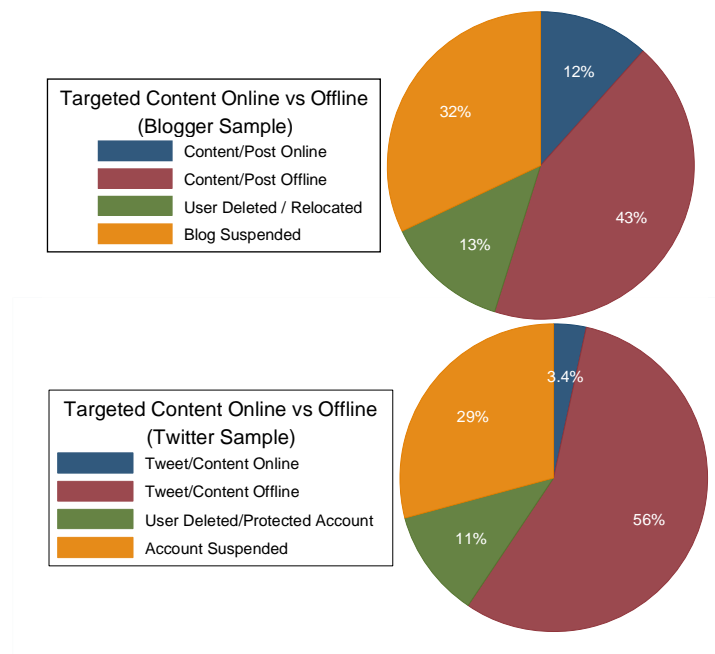


FIGURE 5: DMCA NOTICE IMPACT—FURTHER BREAKDOWN

These figures provide some additional context. For the Blogger sample, the 88% of offline and inaccessible content can be broken down three ways: 43% involved cases where the targeted content or blog post was offline and inaccessible but the blog itself was still online; 32% involved cases where the blog was offline because it was suspended by Google; and 13% involved cases where the blog was offline because it was deleted by the user (or potentially relocated). In the Twitter sample, 56% of cases where content was offline or inaccessible involved cases where tweets or tweet content were removed, offline or inaccessible but the Twitter account itself was still online; in 29% of cases, the Twitter account was suspended and in 11% of cases, it was closed or deleted by users.

A few observations are warranted here. First, there are some noteworthy similarities between the results despite Google Blogger and Twitter being relatively distinct social web platforms.<sup>139</sup> Interestingly, the

139. For example, applying a chi-squared test as recommended by Hanneman et al.,

single category where the greatest disparity is seen is “percentage of content online,” where 12% of content was left online on Blogger but only 3.4% was left online on Twitter. The most straightforward means by which content can be restored after being removed in response to a DMCA notice, as earlier noted, is to file a counter-notice. However, counter-notices have been found to be very rare under the DMCA and, similarly, this study found no evidence any counter-notices were filed in relation to these samples.<sup>140</sup> These different percentages may, instead, reflect an important difference in how each service responds to DMCA notices. Twitter deletes or “withholds” a targeted tweet, leaving users with the explicit option of filing a counter-notice to have the withheld tweet put back online option. A Twitter user could also re-tweet the targeted content again, a legally risky undertaking for a user that has just received a legal threat for the content in the form of a DMCA notice. Google, on the other hand, gives Bloggers the option, beyond mere counter-notices, to simply edit and re-post targeted blog posts, providing a faster and more nuanced means of responding to a DMCA notice.<sup>141</sup> This platform difference may be reflected in those percentages.

These numbers are important when understood in the context of the DMCA’s scheme and, on their own, suggest noteworthy DMCA notice chilling effects. The key to understanding this point is to approach the status of content targeted by a DMCA notice as a measure of the notice’s chilling effect on the user who posted the content and received the notice. First, if content targeted by a DMCA notice remains offline or inaccessible,

---

there is no statistically significant difference between the percentages for either account suspensions (32% for Blogger and 29% for Twitter) or the percentages of accounts deleted or protected (13% for Blogger and 11% for Twitter). There is, however, a statistically significant difference between percentages for content being online (12% vs 3.4%) and being offline or inaccessible (88% and 97%). See ROBERT A. HANNEMAN ET AL., BASIC STATISTICS FOR SOCIAL RESEARCH 332 (2012).

140. Each DMCA notice was searched in the database/dataset for a corresponding counter-notice. None were found. The Lumen Database, in fact, contains very few counter-notices compared to the DMCA notices received. Urban et al., *supra* note 22, at 95. It may be that counter-notices were filed but not deposited in the Lumen Database. Or, in the alternative, content was restored through in response to more informal requests for restoration (the user requesting via communications), or Google or Twitter restoring the content at a later stage (e.g. reversing an earlier removal action in response to a DMCA for some unspecified reason, perhaps internal quality control).

141. With respect to users possibly deleting their accounts and simply creating new accounts elsewhere to re-share/re-post content, see the discussion *supra* note 130.

it suggests users have been chilled or deterred from filing counter-notices to have the content replaced or re-enabled. The counter-notice procedure is the clearest and most simple way to get posts, content, tweets, and other expression put back online within a matter of days (ten to fourteen) after being taken down pursuant to a DMCA notice. As noted earlier, all notices in each sample were acted on by Google or Twitter and therefore Bloggers and Twitter users in question would have been notified of these DMCA notices and their right to file counter-notices; they also have the option to re-post (or re-tweet) content.<sup>142</sup> Failing to take any of these steps suggests the Blogger and Twitter users were chilled from doing so. This point stands whether only the specific targeted blog post or tweet only remains offline or, in the more serious case (which is actually quite common, based on Figure 5), the Blog or Twitter account itself has been deleted, removed, or suspended.

In all of these cases, if the content remains offline, it is because a counter-notice has not been filed. This is a kind of self-censorship as the original content or material was online *expression* and *activity*, that user had chosen to engage in and are now forgoing very like due to receiving a DMCA notice. As Wendy Seltzer has argued, the structure or “architecture” of the DMCA notice and takedown system, combined with the personally received and individualized legal threats posed by the DMCA notices imposes implicit regulatory barriers on speech, expression, and activities online that can promote chilling effects.<sup>143</sup>

Second, if the Blogger or Twitter accounts are deleted or removed in response to a DMCA notice, then the notice has had a *significant* regulatory chilling effect as it means the Blogger or Twitter user has been deterred from all future expression and activities, at least through their existing Twitter or Blogger profile. This chill is even more onerous if the account has been deleted or removed by the user themselves as this self-imposed censorship is substantial in scope and possibly permanent. With a deleted account, it is not just one tweet or blog post that is chilled—or tweets or posts relating to that targeted content—but all communications in the past and future via the platforms are censored, possibly forever. As Figure 5 showed, 13% of Google Blogs and 11% of Twitter accounts targeted by

---

142. Re-posting, as earlier noted, is arguably much easier with Google than Twitter.

143. See Seltzer, *supra* note 28.

DMCA notices were deleted by the user. There may be other reasons why these Blogs and Twitter accounts were closed, but those are noteworthy percentages (13%, n=65; 11%, n=57), suggesting at least some of these accounts were closed by users in response to receiving a DMCA notice.

It is also possible that some of these blogs may have been relocated elsewhere on Blogger or on some other unknown site on the internet. However, relocation can also suggest a noteworthy chilling effect, as removing a blog and re-locating elsewhere severs the blog (and blogger) from all existing online and social media networks, links, audiences and communities in which the user and his or her blog is presently situated. Deleting or relocating a blog is similar to “blog scrubbing” practices observed by Child et al. where bloggers deleted/removed content due to privacy concerns and the Twitter account locking/tweet protecting practices observed by Liu et al. and Almuhimedi et al. also due to privacy concerns.<sup>144</sup> The main difference is that here the DMCA notice is the “trigger” leading the user to engage in “deletion” practices. User blog deletion, relocation, or in the case of Twitter “protection” (making tweets private) are all different forms of chilling effects and self-censorship.

However, others might argue that any chilling effects observed are indicators that the DMCA is operating as expected by targeting and thus chilling illegal content. Indeed, as Schauer emphasized,<sup>145</sup> chilling effects theory primarily concerns the deterrence or chilling of *legal* activities. Unfortunately, fair use and other legalities were not considered and coded here due to methodological limitations—fair use is too nuanced and contextual to determine with the limited information afforded by notices and accounts targeted. However, other empirical studies *have* identified legal issues with DMCA notices. The recent large-scale study of DMCA notices by Urban, Kariganis, and Schoefield found issues with the legal validity of the notice with 28% of the notices, such as failure to comply with statutory notice requirements (e.g., not properly identifying the

---

144. Jeffrey T. Child et al., *Blog Scrubbing: Exploring Triggers that Change Privacy Rules*, 27 *COMPUTERS IN HUMAN BEHAVIOR* 2017 (2011); Yabing Liu et al., *The Tweets They Are A-Changin’: Evolution of Twitter Users and Behavior*, *PROCEEDINGS OF THE EIGHTH INTERNATIONAL AAAI CONFERENCE ON WEBLOGS AND SOCIAL MEDIA* 305, 5 (University of Michigan); Hazim Almuhimedi et al., *Tweets Are Forever: A Large-Scale Quantitative Analysis of Deleted Tweets*, *PROCEEDINGS OF THE 2013 CONFERENCE ON COMPUTER SUPPORTED COOPERATIVE WORK* 897, 897–908 (ACM Press San Antonio, Texas, USA 2013).

145. Schauer, *supra* note 83, at 694.

allegedly infringed work in the notice) or targeting content with potential fair use defenses.<sup>146</sup> Another study of DMCA notices concerning images removal requests found 36.3% of DMCA takedown requests were legally questionable, including 11.5% raising fair use issues<sup>147</sup> and a study conducted in Israel on 10,000 DMCA notices found even more striking problems—only 34% of the notices contained copyright allegations, and the rest concerned claims about defamation or the accuracy of online information not relevant.<sup>148</sup>

These results are corroborated by earlier findings by Urban and Quilter, and Heins and Beckles in their analyses of DMCA notices. Urban and Quilter examined all 876 DMCA notices in the Chilling Effects Project's repository as of August, 2005 and found that 30% "presented an obvious question for a court," most often, a "clear" fair use argument or a complaint over "uncopyrightable" materials.<sup>149</sup> Similarly, Heins and Beckles found 20% of notices targeted material for removal that had either a "strong" or "reasonable" fair use or other copyright defense, while another 27% had a "possible" fair use or other copyright defense.<sup>150</sup> Neither of these studies specifically examines chilling effects issues. However, their work *does* suggest that DMCA notices here may be targeting and thus chilling legal activity.

### 3. *Impact of Automated vs. Non-Automated Notices*

The emergence of algorithmic enforcement of copyright, where bots and automated programs increasingly make decisions about copyright and content removal, has significant implications for copyright.<sup>151</sup> The likely future enforcement and application of the law through automated processes raises issues of fairness, transparency, and accountability,

---

146. Urban et al., *supra* note 22, at 11–12.

147. *Id.*

148. Elkin-Koren, *supra* note 22, at 1088.

149. Urban & Quilter, *supra* note 28, at 666–67.

150. Heins & Beckles, *supra* note 28, at 54; Seng, *supra* note 30, at 416.

151. Elkin-Koren, *supra* note 22, at 1095 (speaking to the challenges for automating fair use, particularly the "high degree of complexity" involved in fair use determinations that is likely beyond present algorithmic capacity).

among others.<sup>152</sup> This automation raises additional chilling effect concerns and claims about the impact of DMCA notices explored in this study. Notices with less than ten URLs for removal were treated as individual sent DMCA notices while those with over ten URLs were treated as automated notices, and are here tabulated by content targeted (Figure 6):

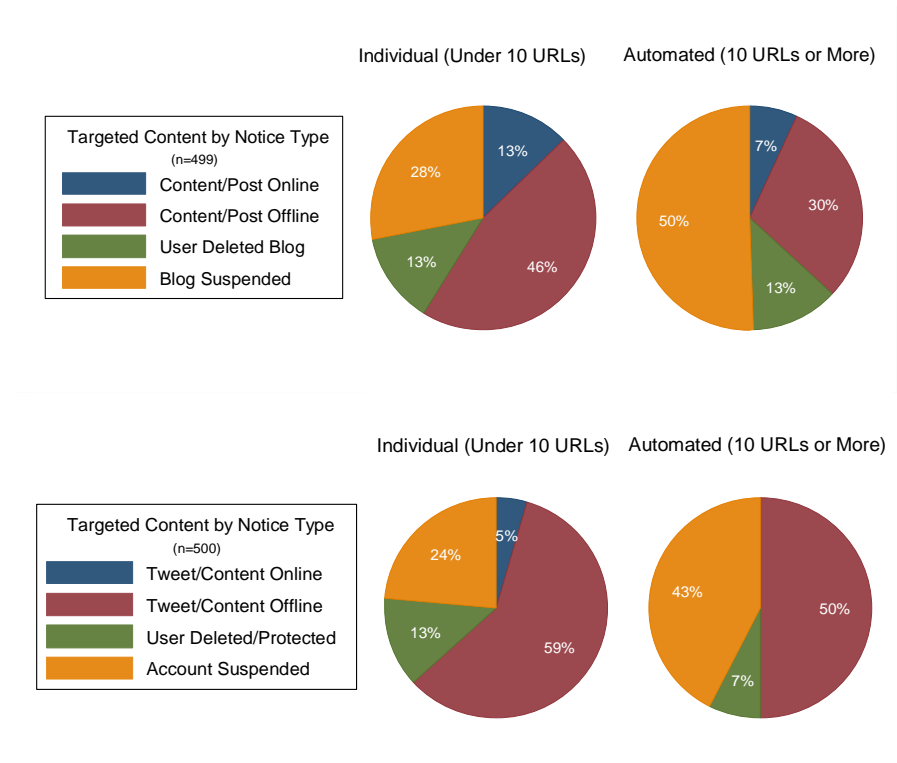


FIGURE 6: AUTOMATED VS. NON-AUTOMATED NOTICES IMPACT  
(TOP: BLOGGER, BOTTOM: TWITTER)

Blogs and Twitter content targeted by automated DMCA notices is more likely to be offline or inaccessible. In the Blogger sample, 13% of the content targeted by non-automated individual notices remains online while only 7% for automated targeted content. The Twitter sample is similar, with 5% of non-automated targeted content online and accessible

152. On the broader issues raised by automation of legal processes, see works cited *supra* note 14.

while *none* of the automated targeted content remained online and accessible. Second, not only is automated content more likely to be offline and inaccessible, the results suggest Blogs and Twitter accounts targeted by automated notices are much more likely to be suspended. Among Blogs targeted by automated DMCA notices, 50% were suspended, compared to only 28% for non-automated individually sent notices. The Twitter sample is comparable. Here, 43% of accounts targeted through automation were suspended, while only 24% of non-automated targeted accounts were suspended.

The impact of automated copyright enforcement on account activity are statistically significant. In both the Blogger sample and the Twitter sample there was a noteworthy positive association between the number of URLs in notices and greater levels of impact (targeted accounts were more likely be suspended or deleted/relocated accounts.) (Blogger:  $\gamma=0.18$ ;  $\chi^2=21.82$ ;  $d=6$ ;  $p<0.01$ ; Twitter:  $\gamma=0.23$ ;  $\chi^2=32.4$ ;  $d=6$ ;  $p<0.01$ ). However, this association became even stronger when the variable was re-coded to distinguish between presumptively individual (less than 10 URLs) and automated (10 URLs or more) in both samples (Blogger:  $\gamma=0.37$ ;  $\chi^2=17.95$ ;  $d=3$ ;  $p<0.01$ ; Twitter:  $\gamma=0.33$ ;  $\chi^2=23.8$ ;  $d=3$ ;  $p<0.01$ ). These were all highly statistically significant findings, with important implications for automated copyright enforcement.

## V. IMPLICATIONS

### A. Chilling Effects and Other Impacts

Despite substantial literature critiquing the DMCA for chilling effects, there are few empirical studies on the DMCA and its notice and takedown system.<sup>153</sup> Studies by Urban and Quilter, Heins and Beckles, Seng and Urban, Kariganis, and Schofield remain the major empirical studies on the DMCA's notice and takedown system.<sup>154</sup> And none of these empirical

---

153. Urban et al., *supra* note 22, at 8 (“Despite the enormous changes since the law was passed, there have been few empirical studies of how notice and takedown actually works in practice”); Seng, *supra* note 25, at 375 (noting the “paucity” of empirical studies on the DMCA’s notice and takedown system).

154. Urban et al., *supra* note 22, at 8; Seng, *supra* note 25; Seng, *supra* note 30, at 7; Urban & Quilter, *supra* note 28; Heins & Beckles, *supra* note 28. However, there are a few newer and more narrow studies. Kristofer Erickson & Martin Kretschmer, *This*

studies investigate DMCA chilling effects claims per se, though they do touch on related issues. Part of the reason for the scarce empirical work is the lack of public data and information about the notice and takedown system.<sup>155</sup> Moreover, there is little empirical research on legal and regulatory chilling effects claims more generally,<sup>156</sup> likely also attributable to the difficulty of measuring and proving chilling effects and self-censorship.<sup>157</sup>

### 1. Evidence of Chilling Effects

This empirical analysis aimed to help address this research void and now, with findings from both empirical case studies combined, a strong case in support of chilling effect claims and concerns emerges. First, survey responses on the DMCA scenario evidenced chilling effects on a range of different online activities, from online expression, to search, to sharing personally created content online, to contributing to social networks. For example, with respect to online speech and expression, 75% of respondents reported being “much less likely” (40%) or “somewhat less likely” (35%) to “speak or write about certain topics online” as a result. 81% “strongly agreed” (50%) or “somewhat agreed” (31%) they “would be more careful about what I say or discuss in certain contexts online.” 72%

---

*Video Is Unavailable*, 9 J. INTELL. PROP. INFO. TECH. & ELEC. COM. L. 75 (2018) (performing an empirical study on the factors that motivate DMCA takedown requests on YouTube); Boyden, *supra* note 30.

155. Urban et al., *supra* note 22, at 8.

156. Kendrick, *supra* note 31, at 1638.

157. *Id.* at 1637–38 (“But there are reasons to doubt the chilling effect account. A claim of a chilling effect necessarily rests upon suppositions about the deterrent effects of law. These suppositions rest in turn upon predictions about the behavior of speakers under counterfactual conditions. Meanwhile, the selection of a remedy for chilling—such as an intent requirement—rests on similar predictions about the remedy’s speech-protective effects. In short, both the detection of a problem and the imposition of a remedy involve intractable empirical difficulties.”); *see, e.g.*, Daniel J. Solove, *The First Amendment as Criminal Procedure*, 82 N.Y.U. L. REV. 112, 155 (2007) (“Determining the existence of a chilling effect is complicated by the difficulty of defining and identifying deterrence. It is hard to measure the deterrence caused by a chilling effect because it is impossible to determine with certainty what people would have said or done in the absence of the government activity. Often, the primary evidence will be a person’s own assertions that she was chilled, but merely accepting such assertions at face value would allow anyone claiming a chilling effect to establish one. At the same time, demanding empirical evidence of deterrence is impractical because it will often be impossible to produce.”).

of respondents were either much less likely (38%) or somewhat less likely (34%) to share such content online after receiving a personalized legal notice. Receiving a personal legal notice about online activities also raised privacy concerns for respondents, with a full 81% of respondents agreeing with the statement that the scenario would make them “more concerned about my privacy, taking extra steps to protect it,” with 55% strongly agreeing and 26% somewhat agreeing. These and other survey findings are consistent with chilling effects theory of DMCA notice impact, offering strong evidence that receiving a notice chills a range of presumptively legal, even socially beneficial, activities that internet users typically engage in.

Second, respondents indicated that they would not re-post or re-share content targeted by such a DMCA notice, suggesting chilling effects. 70% either were “very unlikely” (54%) or “somewhat unlikely” (16%) to re-post or re-share. It should be noted there were also resisters: 15% of respondents said receiving the notice would make them “very likely” (4%) or “somewhat likely” (11%) to re-post or re-share. Additional evidence of DMCA chilling effects comes from the finding that 66% of respondents indicated they would not challenge a DMCA notice they received *even if they believed it to be wrong or inaccurate*. These responses help account for the fact that counter-notices are rare under the DMCA.<sup>158</sup> If such a strong majority of respondents would not challenge a notice when they believe it was wrong or inaccurate, it is easy to see how most would elect not to file a counter-notice or respond in other ways in more uncertain or less clear cases. Respondents stated they would not challenge the notice because they were concerned about legal costs (81%) and avoiding additional risks (legal or otherwise) (53%) and lack of legal knowledge (34%).

These findings are corroborated by the analysis of Google Blogger and Twitter accounts targeted by DMCA notices. This case study used the offline status of content targeted by DMCA notices as a proxy for chilling effects—users, based on fear of legal repercussions, costs, or other risks, do not file counter-notices or fail to repost. As noted in Figure 1, the Blogger sample, 88% of content targeted by the DMCA notices were found to be offline, removed, disabled, or inaccessible. Only 12% remained online

---

158. Urban et al., *supra* note 22, at 95 (noting their study “and other research on this issue consistently shows that counter notices are rarely used”).

and accessible at the time of the study. In the Twitter sample, the percentage of tweets or tweeted content withheld, removed, disabled, or inaccessible, was even higher: 97%, with 3% remaining online and accessible. These numbers suggest (roughly) 90% of the Bloggers and 97% of the Twitter users failed to take steps, legal or otherwise, to get their content back online.

This suggests users have been chilled or deterred from filing counter-notices to have the content replaced or re-enabled or in the case of Blogger simply re-posting a targeted blog post. This is a kind of self-censorship as the original content or material was online *expression* and *activity*, that user had chosen to engage in and are now chilled from defending, re-enabling, or re-posting. And taking these findings with those of Urban et al. there is likely a not insignificant portion of *legal* expression targeted and rendered inaccessible in these results. In their recent DMCA empirical research, Urban et al. found legal problems with 31% DMCA takedown requests for their 2013 sample in one study (with 6.6% raising fair use issues) and 36.3% of notices legally questionable in another study (with 11.5% raising fair use issues).<sup>159</sup> If those percentages hold here, then legal expression is being chilled.

Moreover, some blogs and accounts are deleted or removed in response to a DMCA notice, which suggests an arguably even greater chilling effect as it means the user has been deterred from all future expression and activities, at least through their existing Twitter profile or Blogger account. This chill is even more substantial as deletion or removal of the user account means not just one or a few blog posts or tweets have been chilled, but every possible future post, tweet, communication or expression, has been silenced permanently. As Figure 5 showed, 13% of Google Blogs and 11% of Twitter accounts targeted by DMCA notices were deleted by the user. There may be other reasons why these Blogs and Twitter accounts were closed, but those are significant percentages and total cases (13%, n=65; 11%, n=57), suggesting at least some of these accounts were closed by users in response to receiving a DMCA notice. Where blogs may have been relocated, this can also suggest a chilling

---

159. *Id.* at 11–12.

effect—as removing a blog and re-locating elsewhere severs the blog (and blogger) from its existing online networks, peers, and audiences.<sup>160</sup>

In short, the findings in the Blogger and Twitter case studies corroborates survey responses about DMCA notice chilling effects, such as 70% of respondents either “very unlikely” (54%) or “somewhat unlikely” (16%) to re-post or re-share targeted content. Those sentiments are reflected in the high percentages of content in the Google and Twitter case study found to be offline. All of these different threads of evidence provide compelling evidence for DMCA notice chilling effects.

## 2. “Networked”/Indirect Chilling Effects

Internet users’ activity was chilled not by a legal notice that they had received, but that had been received by someone in their close personal social network in the “friend” DMCA scenario. Here, 90% of respondents either strongly agreed (64%) or somewhat agreed (26%) with the statement that the scenario would make them more “concerned” about their privacy, leading them to take steps to protect it. The scenario would also chill more robust forms of speech and expression, with 76% of respondents strongly agreeing (41%) or somewhat agreeing (35%) that they “would more careful about what [they would] say or discuss in certain contexts online” if a friend of theirs were to receive such a personal legal notice. Respondents’ online search activity was also impacted. This provides evidence of a form of *networked* chilling effect, comparable to social contagion theory, whereby chill or other impacts spread through networks through social links and connections.<sup>161</sup> There is a growing body of research that finds certain social behaviors and phenomena—including smoking, happiness, divorce, cooperative behavior, loneliness, and tastes in music or books—appears to spread through social networks via person-to-person effects.<sup>162</sup> While the mechanisms as to how these interpersonal

---

160. See works cited *supra* note 136.

161. Some social network studies have found internet users self-censor based on factors in their online networks, including audience, and that certain social effects can spread through networks. See generally Das & Kramer, *supra* note 115, at 120–27; Nicholas A. Christakis & James H. Fowler, *Social Contagion Theory: Examining Dynamic Social Networks and Human Behavior*, 32 STAT. MED. 556 (2013); Coviello et al., *supra* note 115.

162. Christakis & Fowler, *supra* note 163, at 556–57.

influences occur are still being explored—and likely vary depending on the context—a common explanation is that the spread of such behavior or actions occurs through interpersonal sharing of norms.<sup>163</sup> The findings here—where a user is chilled after being made aware, via a social media post, that a friend has received a personal legal notice (like a DMCA notice), would be consistent with this social contagion explanation.

*B. A Privacy Theory of Automated Law Chilling Effects*

There are different, at times competing, theories explaining the aforementioned impacts. Schauer's account explains chilling effects in terms of people's concerns about legal harms on receiving a personalized and personally received DMCA notices—a legal threat. However, Solove's theory is based on surveillance concerns and privacy harms and may explain other forms of chilling effects.<sup>164</sup> The DMCA notice raises concerns about being tracked, monitored, and surveilled, leading to chilling effects online.

To test these theories, the results were statistically analyzed to explore any relationship between the chilling effects found and either greater privacy concern or greater legal knowledge as reported by participants. A privacy-based explanation would likely mean any chilling effects would have a positive association with greater privacy concerns. That is, the stronger the privacy concerns among participants, the greater the chilling effects. Conversely, a legal harm theory would likely mean chilling effects should be negatively associated with greater legal knowledge (in theory, participants with more legal knowledge should be less chilled due to concerns about legal harms).

Privacy was easily the most powerful predictor of chilling effects in this study. Greater privacy concern was consistently the strongest predictor of greater impact or chill as measured across a range of contexts and activities in Table 1 and 2. Participants reporting greater privacy concerns about receiving a personally targeted legal notice were statistically also, due to receiving the notice, less likely to speak or write about certain subjects online; would be more cautious and careful about

---

163. *Id.*

164. Solove, *supra* note 86, at 488.

their online speech and online search; were less likely to share personally created content online; and less likely to engage with social network sites. In Table 3, greater privacy concerns meant participants were also less likely to re-post or re-share content targeted by the legal notice received, though here privacy was the second strongest predictor behind gender. In short, greater privacy concerns meant greater chilling effects.

These findings are noteworthy. A substantial body of legal literature has criticized the DMCA's notice and takedown system for creating "chilling effects" for people's online activities.<sup>165</sup> Today, these effects are impacting users at mass scale through automated processes. However, these concerns predominantly rely either expressly or impliedly on Schauer's legal harm theory of chilling effects.<sup>166</sup> Wendy Seltzer's widely cited critique of the DMCA offers an excellent example of this form of theorizing: she analyzes chilling effects under the DMCA as a product both of "error costs" (due to uncertainties in the law) as well as how the law's architecture imposes risks and costs on speech through forms of "intermediation."<sup>167</sup> This study certainly offers evidence that legal harms, consistent with accounts like Seltzer or Schauer, contribute to chilling effects. In Table 1, participants with greater legal knowledge were less likely to provide responses suggestive of chilling effects in relation to online speech, sharing, and social network engagement. In Table 3, participants with greater legal knowledge were more willing to challenge incorrect legal notices they had received and were more willing to re-post or re-share targeted content. Moreover, in the Blogger and Twitter studies, the vast majority of content targeted by these removal notices remained offline and inaccessible. So, legal harm theories like Schauer's are necessary to understand some of the impacts in these findings.

---

165. Carpou, *supra* note 27, at 585 ("Much literature has been devoted to ways in which automated takedown procedures necessarily result in the chilling of free speech and should, therefore, be disallowed or at least discouraged by the DMCA."). *See also* works cited *supra* note 26 (not exhaustive).

166. For the typical framing, see Seltzer, *supra* note 28, at 178, 187 (citing Melville Nimmer on the "balance" and "conflict" at the heart of copyright law); Netanel, *supra* note 88, at 381, 385 (describing uncertainty in copyright law and how it often involves balancing interests of user rights, copyright holders, and the public interest.); *See also* Wagner, *supra* note 88. *See also* works cited *supra* notes 26–27.

167. Seltzer, *supra* note 28, at 193–97.

However, given the strength of privacy concern as a predictor of greater impact across a range of activities in this study, a privacy theory of chilling effects may explain the wider spectrum of impacts and chilling effects, compared to a theory focused on legal harms alone. This is particularly so for chilling effects associated with large scale legal automation via personally received legal notices or directives like a DMCA notice. In the end, when you combine legal and privacy concerns with the fact that these DMCA notices are being sent out daily on a mass-scale by automated processes—by the million—it is easy to foresee substantial chilling effects.

### *C. Differential Impact*

#### *1. Women Disproportionately Impacted*

Another key implication from these findings is the differential impact, or chilling effects, predominantly on women. Statistical analysis of survey findings for both DMCA scenarios set out in Tables 1 and 2 showed a gender effect—female respondents were consistently more likely to be chilled across a range of online activities. That is, female participants, in both Google Blogger and Twitter DMCA scenarios, were statistically less likely to speak or write online in certain contexts, less likely to share personally created content, less likely to engage with social media, and would be more careful in their online search activities both when they were personally targeted by a notice, and when a friend was targeted. This gender effect may be a product of women being previously personally targeted with online harassment, would cause a personalized legal threat to be far more chilling and impactful.<sup>168</sup> Put another way, a prior victim of targeted abuse—which include personal threats—may also be more impacted by targeted personal legal threats.

There were other noteworthy differential impacts as well. For instance, there was an age effect when it came to caution in terms of online search in response to a DMCA notice scenario, with young participants

---

168. DATA & SOCIETY, AMANDA LENHART ET AL., *ONLINE HARASSMENT, DIGITAL ABUSE, AND CYBERSTALKING IN AMERICA* (2016). Cyberharassment laws may, however, offer salutary effects for these kinds of impacts. See generally Danielle Citron & Jonathon Penney, *When Law Frees Us to Speak*, 87 *FORDHAM L. REV.* 2317 (2019).

more impacted. Also, participants that reported more social network engagement were less likely to be chilled. But the gender effects were the most noteworthy and consistent.

## 2. *Knowledge Is Power: Mitigating Chill with Legal Information*

The findings evidenced the importance of legal knowledge. First, participants reporting greater legal knowledge (more familiarity with internet laws) were likely to be chilled from online speech, sharing, and social network engagement. Second, participants with greater legal knowledge were more willing to challenge incorrect legal notices they had received, and were more willing to re-post or re-share targeted content. Again, this might be the product of the respondent simply having a better understanding of what a targeted legal threat might mean, and therefore more willing to take those legal steps in their activities. Legal knowledge was also the sole predictor in terms of people's willingness to legally challenge a notice they believed was wrong or incorrect.

These findings suggest providing users with information as to their legal rights may be a means to mitigate chilling effects and other side effects of personally received legal notices and legal directives, like DMCA notices. This can help provide guidance to policy-makers in crafting legal and regulatory schemes with minimal chilling effects. These findings as to differential impact more generally also lay the foundation for future studies, perhaps also exploring how specific linguistic, ethnic, religious, or other insular groups are affected in terms of regulatory chilling effects.<sup>169</sup>

### D. *The Efficiency of Automated vs. Non-Automated Notices*

The impact of automated DMCA notices compared to non-automated notices were also noteworthy. Content targeted by automation was much more likely to be taken offline or made otherwise inaccessible: only 7% of Blogger content targeted by automation remained online, while 13% non-automated/individual-sent notices remained online. For Twitter, there were *no* instances where automated targeted content remained online

---

169. Jonathon W. Penney, *Internet Surveillance, Regulation, and Chilling Effects Online: A Comparative Case Study*, 6 INTERNET POL'Y REV. (2017) (discussing differential impact in relation to surveillance and other regulatory actions).

compared with 5% of non-automated. Moreover, Blogs and Twitter accounts targeted by automated DMCA notices were much more likely to be suspended, with 50% of Blogs suspended compared to only 28% for non-automated individually sent notices. On Twitter, 43% were suspended, while only 24% of non-automated targeted accounts were so. These findings were also statistically significant. Higher numbers of URLs in DMCA notices and, when recoded, automated notices (10 URLs or more) strongly predicted more serious outcomes—account suspension or removal/relocation.

There are several possible explanations here. On one hand, the results may mean that the automated enforcement of copyright through DMCA notices is doing its job—the automated notices are hitting pirates (that is, repeat offenders) and therefore more of the accounts and blogs targeted end up getting suspended. It could be that automated notices are capturing content that is more clearly illegal and unauthorized copyrighted content. On the other hand, it could be that the sheer volume of automated notices that can be sent causes platforms like Google and Twitter to act proactively to suspend accounts receiving notices, depending on their own internal policies. It appears that Twitter is more proactive about removing content targeted with a DMCA notice than Google Blogs, especially if the content is targeted by an automated notice. This is troubling because lot of legal content may very well be caught in these high volumes of automated notices and left offline than replaced, re-posted, or countered with a counter-notice.

All of this highlights the importance of transparency and accountability in legal and regulatory functions in an era of legal automation, as Perel and Elkin-Koren and others have advocated,<sup>170</sup> including transparency about algorithms used for enforcement as well as by platforms responding to enforcement. These findings strengthen that call.

#### *E. Micro-Directives and the Future of Legal Automation*

The chilling effects and differential impacts of DMCA notices have implications for other forms of legal automation, particularly what I have called *personal* and *personalized* legal automation. These forms of legal

---

170. Perel & Elkin-Koren, *supra* note 22, at 488–92.

automation are often heralded as the future of law.<sup>171</sup> Casey and Niblett's notion of "micro-directives"<sup>172</sup> offers the best and most prominent such example, which as discussed, involve highly context specific legal directives formulated by machines using specific data for a person's specific circumstances and communicated directly to them.<sup>173</sup> They will also be far-reaching—providing people with personal legal directives or guidance as people go about their daily lives, like directing a person of legal speed limits via their car dashboard.<sup>174</sup>

I also argued that DMCA notices are comparable to these more advanced forms of personal and personalized legal automation like micro-directives, albeit nowhere nearly as specific, personalized, or ubiquitous. Nevertheless, the DMCA notice offers a legal threat or directive that is neither general nor impersonal but personal and specific. As such, these notices—sent via automated processes—are comparable to personalized law, micro-directives, and other forms of context-specific and personally received algorithmic legal enforcement.

---

171. See, e.g., Casey & Niblett, *supra* note 6, at 1401; Sunstein, *supra* note 8, at 57; Verstein, *supra* note 69, at 3 ("Some papers have focused on ways in which the state could personalize its directives to citizens."); Devins et al., *supra* note 69, at 367 ("Private technology such as software apps could also provide simple directives for legal consumers to comply with the law without having to weigh the reasonableness of their actions or search for the content of specific law."); Porat & Strahilevitz, *supra* note 8, at 1442–50; Goldenfein, *supra* note 32, at 508 ("As these techniques and practices become more sophisticated and automated, greater reliance is placed on information processing such as data mining, predictive analytics and other artificial intelligence techniques, deployed at mass-scale, to detect patterns 'hidden in the data' for the purpose of flagging or identifying individuals as suspicious reasonableness of their actions or search for the content of specific law.").

172. Casey & Niblett, *supra* note 6, at 1401.

173. *Id.* at 1404 ("With microdirectives, however, the law looks quite different. The legislature merely states its goal. Machines then design the law as a vast catalog of context-specific rules to optimize that goal. From this catalog, a specific micro-directive is selected and communicated"). One important difference is that Casey and Niblett envision micro-directives to be *ex ante* behavioral directives, that is, they provide a legal directive for specific situations and it is then up to the recipient to decide how to respond. DMCA notices are *ex post facto*, to the extent that they are a response to an alleged act of copyright infringement. Nevertheless, the DMCA notice also doubles as a legal directive (or threat), for the recipient to no longer infringe the targeted copyright material.

174. *Id.* at 1406 ("Indeed, microdirectives have the potential to bring wholesale institutional changes to our entire system of laws and the way we choose to regulate behavior.").

This is an important point. Besides the earlier mentioned reasons for the distinction between impersonal and personal legal automation, another central reason is that this form of legal automation, in particular, raises significant concerns about chilling effects and impacts on privacy, freedom, and autonomy. This is easy to understand. In order to personalize automated and algorithmic legal processes then processes need as much information and data as possible on a person's circumstances in order to do so. Porat and Strahilevitz, in their leading article on personalized law, acknowledge the problem and note that privacy laws that protect such data and information may hinder legal personalization efforts.<sup>175</sup> Casey and Niblett likewise acknowledge that the "micro-directives" they herald have serious implications for privacy, autonomy, and human ethics.<sup>176</sup> In terms of privacy, micro-directives would involve state controlled machines gathering data about citizens in order to provide specific legal directives, as well as using various tracking and surveillance technologies to keep track of the "comings and goings" of citizens so to provide these directives in appropriate and timely fashion.<sup>177</sup> Similarly, as micro-directives proliferate, it reduces the free choices of individuals.<sup>178</sup>

Others raise even more substantial concerns not just based on the data collected for personalization, but on how this legal automation is personally targeted. Shay et al. focus on the *personal* aspect of this form of automated legal enforcement, arguing that its intrusion "into the public and the private spheres of citizens' lives" will "erode" trust between and state while "dehumanizing" the governing process.<sup>179</sup> They also speak of chilling effects, as this form of legal automation "degrades" responsible citizenship by modifying behavior "through fear of surveillance and reprisal rather than through a self-generated respect for the rule of law."<sup>180</sup> And these concerns are not just individual level, but also societal.<sup>181</sup> In this context, commentators like Shay et al. and Elkin-Koren

---

175. Porat & Strahilevitz, *supra* note 8, at 1468–69.

176. Casey & Niblett, *supra* note 6, at 1405, 1441–45.

177. *Id.* at 1442.

178. *Id.* at 1443.

179. Shay et al., *supra* note 1, at 39.

180. *Id.*

181. Hartzog et al., *supra* note 1, at 1765 ("[E]mployment of these technologies without careful consideration poses a distinct danger to our civil liberties and can have

and Gal<sup>182</sup> often cite Solove who, as noted, has explored how forms of tracking, monitoring, and surveillance can have broader societal impacts, that is, creating a form of regulatory “environmental pollution” that encourages chilling effects and self-censorship.<sup>183</sup>

My findings would seem to support these concerns—evidence suggesting serious chilling effects arising due to a simple legal notice delivered to internet users about a single issue—copyright. The future forms of legal automation, with highly contextual and specific legal commands and directives very likely would lead to even more substantial chilling effects, and losses of privacy, freedom, and autonomy.

#### *F. Toward a Theory of Personal Legal Automation*

If micro-directives are the future, then it would be a future involving highly personal *and* personalized automated law enforced at mass-scale. But it may also be one with far less freedom. Is there a “theory of governance” for legal automation, and those administering and implementing, to help avert this future?<sup>184</sup> Though guidance in the literature is thin,<sup>185</sup> there are two prominent possibilities. The first is Perel and Elkin-Koren’s DMCA-inspired framework for accountability in algorithmic enforcement.<sup>186</sup> The other is Hartzog et al.’s theory of inefficient legal automation.<sup>187</sup> Both, I think, offer sound ideas that can mitigate chilling effects and similar impacts of personal and personalized legal automation today, and in the future, but neither offer an entirely

---

detrimental effects on society”).

182. Elkin-Koren & Gal, *supra* note 15, at 4 n.7 (citing Solove’s work); Shay et al., *supra* note 1, at 39–40 (citing and discussing Solove’s work on chilling effects).

183. Solove, *supra* note 86, at 488; *see also* Daniel Solove, “I’ve Got Nothing to Hide” and Other Misunderstandings of Privacy, 44 SAN DIEGO L. REV. 745, 769 (2007).

184. Hartzog et al., *supra* note 1, at 1763 (“A theory of governance is critical for those who implement and administer automated law enforcement systems. Without it, systems become unmoored from ethics in the pursuit of efficiency. Failure to responsibly automate law enforcement risks creating systems that undermine law and democracy”).

185. *Id.* at 1767 (“This Article aims to remedy the dearth of guidance by developing a theory of inefficiently automated law enforcement.”).

186. Perel & Elkin-Koren, *supra* note 22, at 474 (proposing “a novel framework for analyzing accountability in algorithmic enforcement that is based on three factors: transparency, due process and public over-sight”).

187. Hartzog et al., *supra* note 1, at 1767.

sufficient solution. Instead, similar to Hartzog et al.'s theory, I argue for a new theory of governance for personal legal automation based on three principles. First, in building, administering, and developing automated legal systems, impersonal components should be first automated, before relying on more personal and personalized components. Second, as legal automation becomes more personal and personalized, then greater counter-measures and safe-guards need to be put in place to counteract its negative impacts, like chilling effects. Third, some legal processes should never be automated where doing so would be too personal and personalized and thus have disproportionate chilling effects more generally or upon certain groups.

Perel and Elkin-Koren's framework for accountability in algorithmic enforcement is based on three factors: transparency, due process and public oversight.<sup>188</sup> It aims, overall, to make decision-makers "justify their choices to those affected by these choices" and be held "answerable" for their actions and failures.<sup>189</sup> The authors provide an excellent analysis of how algorithmic copyright under the DMCA, particularly its automation and implementation on YouTube, fail on each of these accountability factors.<sup>190</sup> They also chart out the nature of the different technical, legal, and practical barriers for accountability, as well as a range of ideas and proposals for "enhancing" accountability for the DMCA, and other forms of legal automation, in terms of transparency, due process, and public oversight.<sup>191</sup> These "accountability enhancing strategies" include, among others, encouraging public participation through greater use of counter-notices, research reverse engineering algorithms for transparency, increasing "watchdog" efforts like the Electronic Frontier Foundation's "Takedown Hall of Shame", and encouraging more OSPs to engage in transparency reporting. These efforts include encouraging OSPs to submit DMCA notices to Harvard's Lumen Database (used in this empirical analysis) to allow for accountability, and encouraging lawmakers to impose mandatory disclosure or clearer standards in certain cases.<sup>192</sup>

---

188. Perel & Elkin-Koren, *supra* note 22, at 474.

189. *Id.* at 481.

190. *Id.* at 497–516.

191. *Id.* at 516–31.

192. *Id.* at 525–30.

All of these suggestions are worthwhile, especially increased transparency. But overall, these enhancements would not necessarily address the chilling effect concerns apparent from the findings in this empirical analysis. While users filing more counter-notifications may be helpful in promoting public participation in the copyright dispute process and allow for, over time, great opportunity for courts to provide guidance and clarify standards for users. But my empirical findings in this analysis suggest that greater use of counter-notices is unlikely—there were no counter-notices found relating to the 500 Google Blogs and 500 Twitter accounts analyzed. And only 34% of respondents in the survey indicated they would take steps to legally challenge a notice even when they believed it was incorrect or wrong. A full 71% were also unlikely to take actions inconsistent with the legal directive embodied in the notice, such as re-sharing or re-posting targeted content, with 55% of respondents indicating they would be “very unlikely” and another 16% “somewhat unlikely.” The suggestion for regulators to standardize measurements for uncertain aspects of copyright law and analysis—like infringement and fair use—would be welcome and could help, but it would not also obviate concerns about legal risks and other uncertainties in the legal process likely contributing to chilling effects illustrated in my findings. Nor can these standards ameliorate any surveillance or privacy related chilling effects, due to users having been targeted by a DMCA notice, being chilled from certain activities because they feel they are being tracked or monitored.

Hartzog et al.’s theory of inefficient legal automation offers another possibility.<sup>193</sup> The authors argue that in order to prevent automated legal systems from being a threat to “civil liberties” and “society” more generally, administrators and designers of such systems should abide by a “conservation” principle that retains both inefficiency and indetermination—usually integrated through human participation and free will—which the authors argue are “key” to the legal process.<sup>194</sup> That is, as a legal system is automated to render it more efficient and

---

193. Hartzog et al., *supra* note 1, at 1767.

194. *Id.*

determined, then inefficiency and indetermination should be proportionally preserved elsewhere in the system.<sup>195</sup>

After setting out a compelling case to justify this theory—and the conservation principle at its heart—the authors also offer ideas to apply the theory.<sup>196</sup> However, at the core of their theory, as it is applied in practice, is the concept of “re-allocation.” That is, the authors argue that as inefficiency and indeterminacy are removed from a legal system through automation, these features or dimensions should be re-allocated to other parts of the system.<sup>197</sup> They argue regulatory responses to increasing automation can help conserve inefficiency and indeterminacy up front (like a regulatory requirement that red light or traffic surveillance cameras not be used without an official present) or after the fact, through the appeals process.<sup>198</sup> However, the problem is that the impact of the legal automation is likely direct and immediate. And following my findings, that impact can have chilling effects that spill over into other unrelated activities negatively impacting freedom, autonomy, and privacy. If a primary concern is preventing forms of legal automation from threatening civil liberties and privacy,<sup>199</sup> and preserving overall “social welfare” in the system,<sup>200</sup> it is difficult to see how reallocation of human intervention—at least after the fact—such as through appeals, can effectively prevent the harms documented in this empirical study.<sup>201</sup>

However, Hartzog et al.’s conservation principle is elegant in its simplicity and provides a helpful means to guide automated legal system design. I also agree that a “theory of governance” is needed for legal automation—on this count, a theory specific to personal and personalized legal automation is probably warranted, one that builds on Hartzog et al.’s basic theoretical framework. A comprehensive elaboration is beyond the scope of this article, but such a theory could be based on three principles,

---

195. *Id.*

196. *Id.* at 1778–92.

197. *Id.* at 1794.

198. *Id.*

199. *Id.* at 1767.

200. *Id.*

201. In fairness to the authors, they do speak extensively to “counter-measures” and are aware of chilling effect risks, see *id.* at 1788–91, it would have been helpful or the authors to offer a few more examples of how these would work in practice.

each informed by my empirical findings in this analysis. The first is that as administrators work to automate aspects of a legal system, more impersonal components should first be automated before automating more personal components. Returning to my earlier distinction between personal and impersonal automation, it would mean first employing automation in legal functions that do not directly communicate or interact with subjects or people—like automating forms of crime or legal liability detection or data analysis—before moving onto more personal and personalized forms of automation like micro-directives. This is comparable to Hartzog et al.'s conservation principle, but it is applied to *how* administrators choose to automate a legal system, rather than simply re-allocated inefficiency and indeterminacy. The point is to minimize harms and impacts associated with more personal and personalized forms of legal automation as documented here, by directing efforts to achieve efficiencies and cost-savings elsewhere in the system.

A second related principle would be that as legal automation becomes more personal and personalized, greater safeguards should be implemented to counteract negative impacts like chilling effects. This may involve incorporating inefficiencies and indeterminacy as Hartzog et al. suggest, or incorporating features designed specifically to mitigate or counter chilling effects and similar impacts, such as human “in the loop” buffers or automated counter-processes that mitigate chill by providing information and options to people subjected to the personal and personalized legal automation.<sup>202</sup>

A third principle might be that in some instances, a form of legal automation should not be implemented at all if it would be too personal and personalized. Andrew Verstein in a forthcoming article on personalized law asks “[w]hen should the law personalize?”<sup>203</sup> To answer, he sets out a range of factors including the value of formal equality in a given context, availability of data about subjects, and whether providing

---

202. The author is presently involved with a research study, with collaborators at Princeton's Center for Information Technology Policy and the MIT Media Lab, that is testing automated processes that can mitigate chilling effects by delivering legal information to targeted users. See *Reducing side-effects of copyright bots on Twitter: Jon Penney and Merry Mou at #CSMIT2018*, YOUTUBE (Jul. 9, 2018), <https://perma.cc/YEN5-G2XP>.

203. Verstein, *supra* note 69, at 554.

legal directives make sense in a given context.<sup>204</sup> However, I would suggest another factor in light of findings here: impact on privacy, civil liberties, and potential for chilling effects. Put simply, in some cases the risks or impacts would be too great, and no “trade-off” for efficiency or cost-saving could be worth the negative impacts or side effects. In those cases, administrators or system designers should avoid legal automation altogether. There might be other principles in a theory of governance for personal legal of automation, but this is aimed to simply be a start or foundation.

## VI. CONCLUSION

Existing scholarship on legal automation, algorithmic law, and artificial intelligence is growing but presently narrowly focused,<sup>205</sup> leaving uncertainty on a range of issues,<sup>206</sup> and there is a need for more empirical research to understand and explore both the impact and effectiveness of legal automation.<sup>207</sup> This gap is especially noteworthy in light of concerns that new forms of personalized algorithmic law and automated legal enforcement pose serious risks for civil liberties and society at large.<sup>208</sup> This article set out to address this research gap with an original empirical analysis of the DMCA and its notice and takedown system, which through increasing automation at mass scale sits at the “forefront” of algorithmic law enforcement today.<sup>209</sup> The automation of DMCA notices has also long

---

204. *Id.*

205. Alarie et al., *supra* note 1, at 1 (“Legal scholars investigating artificial intelligence are preoccupied with regulation. The literature has largely focused on the need for humans to regulate the behavior of automated systems.”).

206. Hartzog, *supra* note 2, at 1 (“People making decisions related to technology law, policy, and ethics have not faced such uncertainty since the advent of the Internet.”).

207. *See generally* works cited *supra* notes 13–14. *See also* Pasquale & Cashwell, *supra* note 1, at 28 (“[T]he agenda for researchers must shift toward direct examination of law’s diverse practice areas and functions.”).

208. Hartzog et al., *supra* note 1, at 1765 (“[E]mployment of these technologies without careful consideration poses a distinct danger to our civil liberties and can have detrimental effects on society.”).

209. Perel & Elkin-Koren, *supra* note 22, at 477 (“Copyright law was at the forefront of algorithmic law enforcement beginning in the early 1990s, conferring safe harbor protection to online intermediaries who re-moved allegedly infringing content upon notice under the Notice and Takedown [N&TD] procedure designed by the Digital

been criticized for potentially having serious chilling effects, impacting people's autonomy, freedom of expression, and privacy online.<sup>210</sup> Despite this, there are few empirical studies on the DMCA and none specifically on chilling effects.

This article's empirical analysis examines chilling effects and other impacts caused by users receiving DMCA notices, synthesizing findings from two original empirical legal case studies to triangulate and explore this issue. In the process, it has offered a number of new, noteworthy, and original findings, including compelling evidence for DMCA notice chilling effects across a range of user activities—addressing long time skepticism among judges, lawyers, and scholars about such claims, first-of-its-kind evidence of differential impacts including how women are disproportionately chilled or the mitigating effect of legal knowledge or social network engagement on chilling effects, as well as insights as to the effectiveness of DMCA notices in obtaining legal compliance and the efficiency of automated DMCA notices as compared to non-automated ones. The analysis also explored the implications that these empirical findings had for the future of legal automation and laid the foundations for a theory of personal legal automation and how it should be administered and governed going forward.

Of course, this research had a number of limitations. First, recruiting survey respondents through an online crowdsourcing platform raises a concern as to bias in the participant pool; without randomized probabilistic sampling, the survey sample was not representative, and subject to concerns about self-selection and sampling bias. However, the participant pool was “relatively representative” of the U.S. adult internet user population and, additionally, measures were taken to strengthen the validity of findings, including, among other things, extensive field testing of survey questions/design, question ordering to avoid priming respondents, attention checks, and use of hypothetical scenarios to better measure behavioral responses rather than self-reported attitudes. Second, the Blogger and Twitter analysis as a case study cannot be taken as generally representative of chilling effects caused by other laws or regulatory

---

Millennium Copyright Act [DMCA].”); Elkin-Koren, *supra* note 22, at 1084.

210. Carpou, *supra* note 27, at 585 (“Much literature has been devoted to ways in which automated takedown procedures necessarily result in the chilling of free speech and should, therefore, be disallowed or at least discouraged by the DMCA.”).

schemes; nor assumed to predict the impact of other such regulatory schemes. The case study involved 500 blogs on Blogger and 500 Twitter accounts randomly selected from DMCA notices from the Lumen Database, which is in no way representative of the wide array of websites, platforms, and online service providers that receive DMCA notices every day. Still, such limits are inherent to this sort of research design, and case study nevertheless offers evidence for regulatory chilling effects and how these effects operate in a regulatory context online.

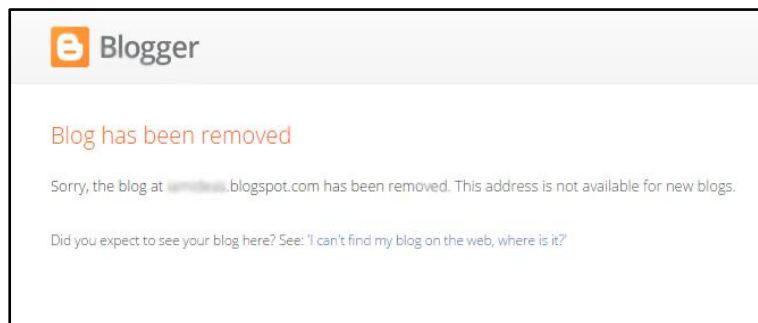
Finally, in the second case study, each blog and Twitter account was visited and the status of the content targeted by the notice was recorded as online or offline, and the status of the account itself was recorded as deleted, suspended, or still online. However, it was impossible to track down instances where Bloggers may have relocated their blogs (if deleted) or if they reposted the targeted content online in some other unknown location or site online. Nor does this coding and quantitative approach provide in depth reasons for why bloggers or Twitter users left content online or offline or removed their blogs, for example. Additional qualitative research and similar investigations would offer invaluable information and could form the basis of future projects. For example, each blogger or Twitter user that received a DMCA notice could be contacted and interviewed to understand each individual's reasons and motivations for acting the way they did.

Though far more research should be done to further document, explore, and understand chilling effects in relation to the DMCA and automated legal processes more generally, this study has aimed to lay the foundation for that work, and a better understanding of legal automation today and tomorrow.

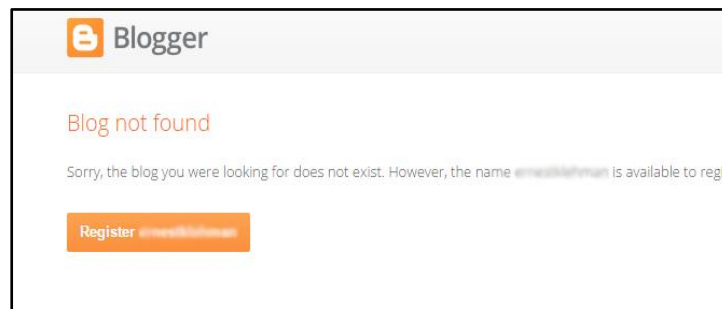
## APPENDIX A. BLOGGER SAMPLE: CODING ONLINE/OFFLINE CONTENT STATUS

*Blog suspended (Coded as 4 – Suspended)*

This return indicates that Google has suspended or permanently locked the blog targeted by the DMCA takedown notice (e.g., for Terms of Service violation). It may also indicate the blog has been deleted by the user and Google has locked the blog (preventing anyone from re-registering), which is sometimes done on request of the blog owner or due to other Terms of Service concerns (e.g., Google locks the blog to prevent spammers from re-registering it).

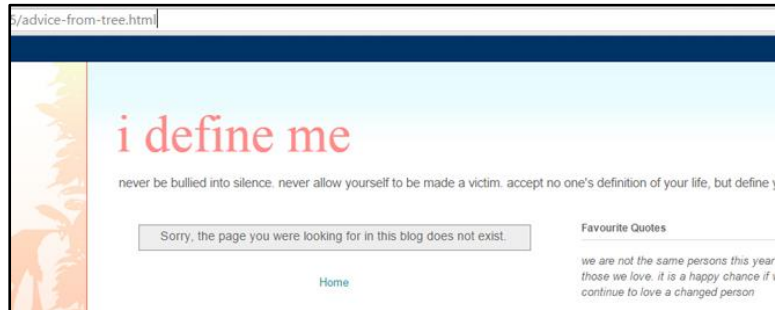
*Blog deleted/relocated by user (Coded as 3 – User Deleted / Relocated)*

This return indicates that the owner of the blog targeted by the DMCA takedown notice has taken the step to delete the blog and 90 days have passed since deletion (Google releases the name of user deleted blogs after 90 days). It is also possible that the blog has been relocated on Blogger (as Google also releases the name/URL of blogs on renaming/relocating).



*Targeted post or content offline (Coded as 2 – Offline)*

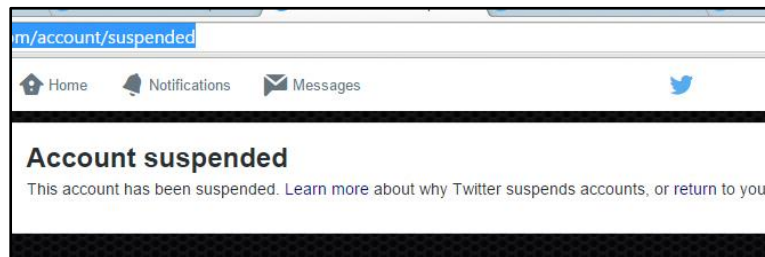
Example of a return coded as offline (code: 2). The blog post containing the content targeted by the DMCA notice remains offline.



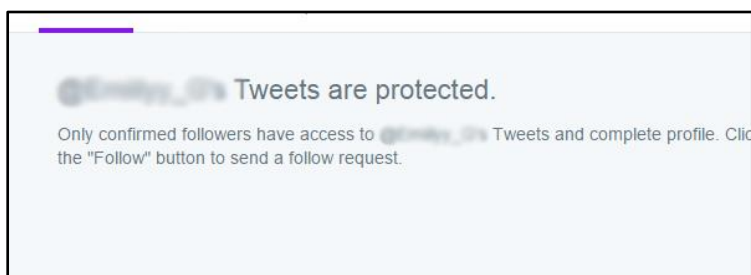
## APPENDIX B. TWITTER SAMPLE: CODING ONLINE/OFFLINE CONTENT STATUS

*Twitter account suspended (Coded as 4 – Suspended)*

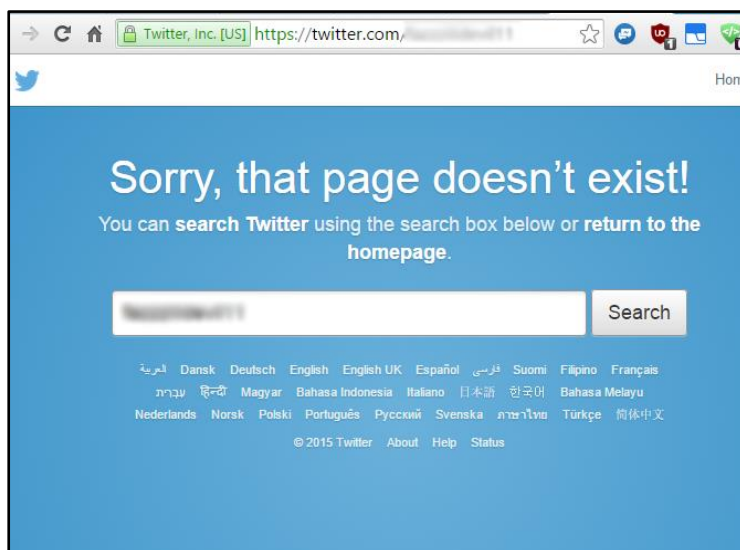
This return indicates that Twitter has suspended or locked the Twitter account targeted by the DMCA takedown notice (e.g., for Terms of Service violation).

*Twitter account deleted / protected by user (Coded as 3 – User Deleted)*

This return indicates the owner of the Twitter account targeted by the DMCA takedown notice either protected their account (as in, made their tweets private and not publicly accessible) (#1) or deleted the account (the account no longer exists) (#2)



#1



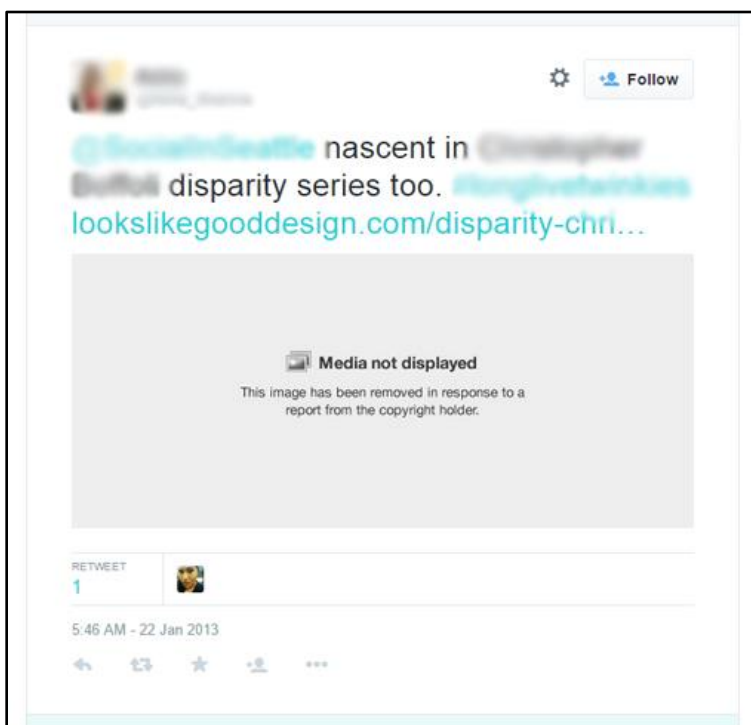
#2

*Targeted content is offline (Coded as 2 – Offline)*

Examples of returns coded as offline (code: 2). The Tweet containing the content targeted by the DMCA takedown notice remains offline (#1) or the specific media targeted in a Tweet remains offline (#2).



#1



#2